Thesis
3525

ATTITUDES TOWARDS FINNISH-ACCENTED ENGLISH

BY

IAN MORRIS-WILSON

Thesis submitted to the University of Stirling
in fulfilment of the requirements
for the degree of Doctor of Philosophy

September 1999

Attitudes towards Finnish-accented English

Ian Morris-Wilson

DECLARATION

Thesis submitted to the University of Stirling
for the degree of Doctor of Philosophy

I declare that this thesis was written and composed solely by myself, that I myself did all the work that it embodies, and that no part of it has been included in any other thesis.

Ian Morris-Wilson
September 1999
Oulu

ABSTRACT

PhD: Attitudes Towards Finnish-accented English

The thesis opens with a discussion of what attitudes are, and develops with a review of studies of attitudes towards pronunciation error, attitudes towards foreign accents and perception of foreign-accented speakers.

The empirical part of the thesis attempts to identify how native (British) and Finnish listeners of English react to and evaluate typical segmental features of mispronunciation in the English speech of Finnish men and women of various ages. Two experiments using modifications of the matched-guise technique were conducted, one to consider error evaluation and to establish a hierarchy of segmental mispronunciation, the other to examine speaker evaluation, the image of the speaker created by the mispronunciation.

Recordings of Finnish-accented English were presented to male and female listeners of various ages, and reactions collected. Statistical analyses of the results were carried out and the following general conclusions were drawn: the English labiodental lenis fricative / v / when mispronounced in the typical Finnish manner as a labiodental frictionless continuant [υ] is not tolerated by native English listeners at all, though it is highly tolerated by Finnish-speaking listeners (and Swedish-speaking Finns) themselves; the degree of mispronunciation in Finnish-accented English seriously affects listeners' estimations of the speaker's age, bad mispronunciation prompting under-estimation of age and good pronunciation over-estimation; both Finnish-speaking listeners and English-speaking listeners have almost identical clear pre-set standards about what constitutes 'good' and 'bad' pronunciation; a Finnish speaker's phonemically 'better' and 'worse' pronunciation affects the image listeners have of the speaker, status/competence traits in particular being up-graded for better pronunciation, solidarity/benevolence traits remaining broadly unaffected, and English-speaking listeners generally being more positive towards the Finnish-accented speakers than compatriot Finns.

CONTENTS

ACKNOWLEDGEMENTS ix
EXPLANATORY NOTES x
1. INTRODUCTION 1
2. GENERAL BACKGROUND TO RESEARCH INTO ATTITUDES 5
2.1. Attitudes in the social sciences and social psychology 5
2.1.1. The term 'attitude' 8
2.1.2. The measurement of attitudes 15
2.1.3. Scaling techniques 25
2.1.3.1. Thurstone scales 26
2.1.3.2. Guttman scales 27
2.1.3.3. Likert scales 28
2.1.3.4. Osgood's Semantic Differential 30
2.1.4. Problems in measuring attitudes 34
2.2. Survey methods 36
2.2.1. Research design 36
2.2.2. Questionnaires 37
2.2.3. Populations and samples 38
2.2.4. Statistical tests and procedures 43
2.2.4.1. Nominal data 44
2.2.4.2. Ordinal data 45
3. ATTITUDES TOWARDS PRONUNCIATION ERROR 47
3.1. Methodological problems in investigating pronunciation errors 48
3.2. Grading of pronunciation errors 50
3.3. Error hierarchy and error preference in foreign-accented speech 54
4. ATTITUDES TOWARDS FOREIGN ACCENTS AND PERCEPTION OF FOREIGN-ACCENTED SPEAKERS 58
4.1. Early studies 59
4.2. The matched-guise technique 61
4.3. Criticism of the matched-guise technique 68
4.4. Theorising prompted by the matched-guise technique 72
4.5. Refinements to the matched-guise technique 78
4.5.1. Characteristics of stimulus voice 79
4.5.1.1. Gender of voice 79
4.5.1.2. Age of voice 82
4.5.1.3. Accent broadness 83

PAGE

NUMBERING

AS ORIGINAL

4.5.1.4. Tempo, pitch, volume and voice quality 87
4.5.1.5. Background information, context and speech content 91
4.5.2. Informant characteristics 95
4.5.2.1. Listener's age 95
4.5.2.2. Listener's gender 96
4.5.2.3. Listener's background 97
4.6. Foreign-accented accents 98
5. AIMS OF THE PRESENT STUDY 106
5.1. Error evaluation / Alternative mispronunciations 106
5.2. Speaker evaluation 107
6. METHOD 109
6.1. Materials 109
6.1.1. Experiment 1: Materials for error evaluation/alternative mispronunciations 109
6.1.1.1. Texts 109
6.1.1.2. Tapes 111
6.1.1.3. Questionnaires 114
6.1.2. Experiment 2: Materials for speaker evaluation 114
6.1.2.1. Texts 114
6.1.2.2. Tapes 115
6.1.2.3. Questionnaires 118
6.1.3. Statistical note 123
6.2. Informants and test protocol 123
6.2.1. Convenience samples 123
6.2.2. Co-operation rates 124
6.2.3. Biographical background questionnaire 125
6.3. Listening equipment 127
7. RESULTS AND ANALYSIS – EXPERIMENT 1: ERROR EVALUATION/ALTERNATIVE MISPRONUNCIATIONS 128
7.1. Error free pronunciation v. a particular phonemic error 128
7.2. Hierarchies of tolerance towards mispronunciations 132
7.2.1. Mispronounced lenis labiodental fricative / v / 132
7.2.2. Absence of plosive aspiration 143
7.2.3. Mispronounced affricates 148
7.2.4. Mispronounced sibilants 151
7.2.5. Mispronounced fortis dental fricative / θ / 155
7.3. Influence of gender of informants 155
7.3.1. English-speaking informants 157
7.3.2. Finnish-speaking informants 160
7.4. Influence of age of informants 164
7.4.1. English-speaking informants 164
7.4.2. Finnish-speaking informants 169

8. RESULTS AND ANALYSIS – EXPERIMENT 2: SPEAKER EVALUATION 174
8.1. Estimated age of speakers 174
8.1.1. Gender of informants 186
8.1.2. Age of informants 191
8.1.3. Summary 195
8.2. Estimated excellence of the pronunciation of speakers 196
8.2.1. Finnish-speaking listeners v. English-speaking listeners 209
8.2.2. Gender of informants 215
8.2.3. Summary 219
8.3. The fourteen traits 220
8.3.1. Gender of informants 246
8.3.2. Age of informants 262
8.3.3. The 'shy' (*ujo*) trait 267
8.3.4. Summary 270
9. CONCLUSION 272
9.1. Problems associated with the MGT in relation to the present studies 273
9.2. The overall findings 275
9.3. Future research 277
9.4. Implications for the teaching and learning of pronunciation 278

APPENDIXES
Appendix A: Phonetic transcriptions of the recordings for Experiment 1 284
Appendix B: Texts of the questionnaire for Experiment 1 289
Appendix C: Phonetic transcriptions of the recordings for Experiment 2 301
Appendix D: Texts of the questionnaire for Experiment 2 304
Appendix E: Biographical background questionnaire 311

REFERENCES 326

ACKNOWLEDGEMENTS

Without the help and goodwill of literally hundreds of informants I would never have been able to carry out the empirical part of this thesis. My deepest gratitude goes to my informants in Canterbury, Edinburgh, Glasgow, Hull, Kajaani, Oulu, Stirling and Vaasa. Thank you. *Kiitos. Tack.*

I also owe a debt to all those teachers who helped me find informants, giving me valuable teaching time, arranging interview sessions, and above all cajoling students to participate. I fear I may have missed some names from the following list and hence my apologies to those whose names I may omit, but thanks to everyone, *including*: Vance Adair, Julie Allan, Richard Badger, Bethan Benwell, Amelia Bryce, Martin Davies, Judy Delin, John Drakakis, Cathy Emmott, Jolene Geer, Robert Geer, Douglas Jamieson, Candy Kallio, Heather Kannasmaa, Christian Kay, Diana Kent, Leena Kiuru, Martin Kusch, Jane Laverick, Mari-Carmen Lopez, Suzy McAnsh, Malcolm MacDonald, Susana Murcia-Bielsa, Patrick Nesbitt, Maggie Nolan, Greg Oates, Rolf Palmberg, Christine Penman, Ken Pennington, Gordon Roberts, Gerhard Schmitt, Jeremy Smith, Jackie Tasioulas, Juhani Toivanen, Rauno Varonen, Rob Veltman, Andrew Wheatcroft, Goodith White. Thanks. *Kiitos. Tack. Danke. Merci. Gracias.*

I also owe special thanks to Pentti Körkkö for his technical advice and assistance. Without the co-operation and understanding of the Department of Finnish, Saami and Logopedics, Oulu University, the recordings for my empirical studies would not have been of the high quality that they were. *Kiitos.*

Ilkka Marjomaa helped me with a few translation problems and, together with Timo Lauttamus, gave me welcome moral support. *Paljon kiitoksia.*

I also thank the technical assistants responsible for the language laboratories in the Universities of Canterbury, Hull and Stirling, the library staff of Oulu University and Stirling University, and all the computer experts (particularly Robin Lawford, Rob Marshall and Oron Yoffe in Stirling University) who helped me.

Je tiens également à remercier Madame MacDonald et feu son époux, Mac.

I gratefully acknowledge the research scholarships awarded to me by the University of Oulu. *Oulun yliopiston tukisäätiö on myöntänyt minulle tutkimusapurahaa, josta olen erittäin kiitollinen.*

Finally I recognise the guidance given by my supervisors in the Department of English Studies, Stirling University: first Martin Davies, and then, after his retirement, Judy Delin and David Punter.

Ian Morris-Wilson
September 1999. Oulu.

EXPLANATORY NOTES

The phonetic script used is that of the International Phonetic Association as adapted for English by Cruttenden (1994) unless otherwise stated and explained.

Italics are used for titles of books and journals, emphasis and the citation of lexical examples.

Boldface type is used for key ideas and terminology.

The following colour conventions are used in the various sets of bar charts:

- Excellent Pronunciation
- Good Pronunciation
- Fair Pronunciation
- Poor Pronunciation

- Finnish-speaking listeners
- English-speaking listeners

- <= 22 years of age
- >= 23 years of age

Attitudes towards Finnish-accented English

1. INTRODUCTION

O wad som Power the giftie gie us
To see oursels as others see us . . .
—Robert Burns, *To a Louse*

I have been teaching English pronunciation and phonetics to Finnish university students for some twenty years, produced teaching materials for my courses and written text books for my students. Over the years I have repeatedly been faced with the same problem: there is no empirical evidence available to show what features of mispronunciation are more acceptable and less acceptable to native (or other foreign) listeners of English, and there is nothing to present to students to explain why it is in their interest to improve their pronunciation. There is no recognised hierarchy of error acceptability that teachers can refer to and use as the basis of a priority list in their teaching, and there is no description of the drawbacks and advantages of foreign-accented speech—specifically the perception of foreign-accented speakers of English—that students can refer to in order to bolster their motivation in their struggle to improve their pronunciation.

The present study arose from my wish to try and do something, no matter how modest, to rectify this situation. If I could start the ball rolling, then someone else might feel prompted to give it another shove, and then someone else and so on, until there is a body of knowledge which would be of real practical help to the teachers and Finnish-speaking learners of English pronunciation and phonetics in Finland.

Teachers of pronunciation need to be able to make reasoned decisions as to what to teach their students, how much to teach them and in what order. They also need to be able to claim with confidence that a certain level of proficiency is the minimum standard acceptable for specific and various purposes (from general tourist comprehension and service encounters through to international business meetings or whatever). Intuition, experience and second-hand advice can go so far, but there comes a point when professional conscientious teachers need something more. We need facts.

As for foreign learners of English pronunciation, they are usually prepared to learn at least a minimal approximation to some standard accent of English, whether it be English English, American English or some other native variety. Without some basic level of proficiency learners themselves are aware of the problems that their interlocutors are having in trying to understand them and to decipher what they are saying. But again there comes a point, sometimes very

early on, sometimes at a very advanced level of expertise, when learners consciously or subconsciously wonder what the point is of trying to improve even further: why bother? What information do we have to motivate learners (both beginners and advanced) to improve? It is easy for Jones (1997:105) to claim sweepingly that 'pronunciation teaching methods should more fully address the issues of motivation and exposure by creating an awareness of the importance of pronunciation,' but what does this mean in practice? How can we motivate learners when we have nothing to offer them except our beliefs? We have no facts.

At the moment the only knowledge available to both teachers and learners as regards preference and acceptability is of questionable value. We know the frequency of occurrence of phonemes in a corpus of 'general text' (Cruttenden 1994:136–7, 196–7; Knowles 1987:223); we know the frequency of occurrence of phonemes in the words making up the *Oxford Advanced Learners' Dictionary of English* (Higgins 1996); and we have lists of all the minimal pairs found in the language (Higgins 1996). While of academic interest, these tables and lists are of little practical relevance in the classroom. Would any violin teacher claim that because the most frequently occurring note that a violinist has to play is Middle C (or whatever), then that is the note that must receive top priority in learning to play the violin? No. Would a driving instructor claim that because reversing is relatively such a small percentage of normal driving activity learners need spend only a tiny fraction of their time learning how to reverse? No. So should language teachers adopt this frequency argument for the prioritisation of the phonemes of English? No. These are not the facts we need.

Moreover, the contention that lists of minimal pairs reveal the *real* likelihood of misunderstandings resulting from phoneme inaccuracy is, in my opinion, unconvincing. Lists reveal only theoretical potential for confusion among listeners and conceal the fact that listeners are not stupid: listeners can quickly grow accustomed to an accent, can make adjustments, and can often disambiguate confusions from context. Familiarity with an accent breeds ease of comprehension. Lists also hide the fact that listeners are not machines, are not automatons, but rather emotional beings reacting to mispronunciations and accents on both conscious and unconscious levels in ways which I believe cannot be predicted from number-crunching frequency counts.

I feel, therefore, that research into foreign-accented English is needed on two fronts subsumed under the general heading of 'Attitudes of listeners towards foreign-accented English.'

First we need more information about what features of mispronunciation are tolerated by native English listeners. Are phonemic inaccuracies considered to be more serious than suprasegmental infelicities? Are inappropriate intonation patterns noticed and loathed, or are they accepted at 'ear value,' assumed to be correct and intentional, and therefore misconstrued? Are poor tempo, rhythm and stress placement considered as minor deviations not worth commenting on, or recognised as something more serious? Are there combinations of inaccuracies

which are intolerable but whose individual separated entities are quite acceptable?

Second we need to ascertain what images particular forms of mispronunciation evoke in native listeners. Just as the lisp of a native speaker, or high-pitched voice, fast delivery or loud volume suggest certain types of personality (e.g. effeminate, highly strung, nervous, brash), do particular features of foreign-accented mispronunciation also evoke certain speaker-images in the minds of listeners? German-accented English, Italian-accented English and French-accented English all provoke certain well-known national stereotypical images, but are these images evoked solely because the listener *knows* that the speaker is German or Italian or French? Or is it the case that certain features of mispronunciation, when recognised as foreign-accented English as opposed to some variety of native English, always evoke the same image?

Discoveries made in these two broad areas need, of course, to be placed within various contexts, and for my part within the Finnish context in which I work. A growing Finnish interest in socio-phonetics is evident (e.g. Valo 1994, Pihko 1997), but little is yet known of the attitudes of Finns towards different regional accents of Finnish (let alone foreign-accented Finnish) or—more importantly—how those attitudes are reflected in attitudes towards foreign languages and especially towards Finnish-accented English.

Finns are used to hearing their leading politicians, newsreaders and sports commentators speaking English on television, and frequently hear their professional colleagues, their friends and even local shop assistants speaking English in business meetings, international gatherings and encounters of various sorts. Moreover, many Finns have to speak English themselves in all sorts of situations. However, no one really knows how Finns themselves, or native English listeners, evaluate and react to the Finnish-accented English of Finns, and it is this gap in our knowledge that led to the present study.

The aim of the present study, therefore, was twofold.

The first aim was to discover what attitudes listeners have towards Finnish-accented English, to determine which phonemic mistakes are more tolerated than others, thus establishing a hierarchy of acceptability for Finnish-accented phonemic mispronunciations. This part of the study has a strong pedagogical streak and may be broadly labelled as error evaluation.

The second aim was to investigate the personality images prompted by various degrees of Finnish-accented English, thus providing evidence for the more positive or negative perception of the personality traits of Finnish speakers resulting from the degree of their phonemic inaccuracy when speaking English. This part of the study is more sociological and psychological than the first, following the example of many sociophonetic and socio-psychological studies which have examined the perception of the personality traits of regional/national accented native-speakers and which have only recently started to consider the perception of the personality traits of non-native speakers in a systematic way.

Both aims were pursued by preparing special recordings of Finnish-accented English and presenting them to both Finnish-speaking and English-speaking

native listeners. The evaluations of the two groups of listeners, and particularly any discrepancies in the evaluations and reactions of the two informant groups, would indicate where the teaching of English pronunciation in Finland could be developed and improved, and suggest where further research is needed.

My two-pronged interest in the attitudes of listeners towards Finnish-accented English can be placed within the wider context of the growing discipline referred to as 'language attitudes,' a discipline I perhaps pedantically prefer to refer to as 'attitudes towards language.'[1] The two prongs reflect two of the areas that Fasold (1984:148) suggests are perhaps the main areas of study in this field:

1) 'attitudes toward language itself [e.g. is a given language variety "rich", "poor", "beautiful", "ugly", "sweet-sounding", "harsh", etc.],' and
2) 'attitudes towards speakers of language varieties.'

However, while looking at the mispronunciation of the interlanguage and at the image of the speakers of that interlanguage, my study is both non-applied and applied. On the one hand it is a simple survey to discover who thinks what, but on the other hand it is closely connected to work which takes place in the foreign-language classroom. In one sense the study is even actuarial, for it attempts to assess—through its surveys—the 'risks' involved in speaking English in a Finnish-accented manner, and—with its analyses and conclusions—to advise learners how much time and effort to 'invest' in correcting their mispronunciation. The study arose from problems in the classroom and there is a strong pedagogical quality in its conclusions, for both teachers and students should be able to use the facts the two experiments throw up in order, respectively, to plan their pronunciation teaching more efficiently, and to improve their defective pronunciation more effectively.

[1] I have never heard of anyone referring to such phenomena as 'chocolate attitudes', 'whisky attitudes', or 'pornography attitudes'. So why 'language attitudes'?

2. GENERAL BACKGROUND TO RESEARCH INTO ATTITUDES

Tout ce que tu dis parle de toi.
—Valéry, *Mauvaises pensées et autres*

Before embarking on my study, I need first to examine the term 'attitude' in order to define what I am interested in, and then to show that within the field of social psychology there is a considerable body of literature devoted to attitude theory, measurement and change which the sociolinguist would be foolish to ignore.

The core of my study is attitudes towards (or reactions to) foreign accent, specifically segmental mispronunciation, and thus my study straddles several disciplinary borders, and care is needed to place it in its context. I shall need to consider aspects of phonetics, dialectology, sociolinguistics and social psychology, as well as techniques such as questionnaires and non-parametric statistics used in sociological surveys.

2.1. Attitudes in the social sciences and social psychology

Social scientists are broadly divided into two camps as regards the role of attitudes in their studies of the interplay of social structures and human behaviour.

Some say that the concept of attitude is irrelevant and worthless since attitudes cannot help one explain behaviour or conduct in social structural terms. Attitudes do not, they claim, work together with factors of social structure to determine conduct, but (like conduct) are rather a result of social structure. Attitudes, therefore, while perhaps interesting in themselves as reflectors of the influence of social factors cannot be used to explain, let alone predict, conduct. A knowledge of attitudes adds nothing to our descriptive and explanatory models which is not already there as a result of our knowledge of social structures. Furthermore, a change in the structure of society might provoke a change in attitudes and conduct, but a change in attitudes would have no effect on society or conduct.

Others argue either that attitudes work together with factors of social structure to produce particular patterns of conduct (i.e. a parallel or joint influence), or that attitudes, which affect and help determine conduct, are themselves affected and determined (totally or partly) by society (i.e. a sequential and mediating influence).

Triandis draws a middle line between the two camps:

> Attitudes are neither necessary nor sufficient causes of behavior. They are 'facilitative causes.' . . . Behavior is a function of (a) attitude, (b) norms, (c) habits, and (d) expectancies about reinforcement. . . . Attitudes alone do not predict behavior; attitudes together with norms and habits do. (Triandis 1971:15–6)

Both groups of social scientists are necessarily assuming that social structural factors can be identified and isolated and are definable variables. However, whereas the former (i.e. the advocates of the 'parallel/joint' model) imply that attitudes are independent of social structural influences and pressures, and are not necessarily determined by them, the latter (i.e. the advocates of the 'sequential' model) view attitudes as the potentially predictable outcome of a process, and as variables open to change. This last viewpoint leaves open the possibility that attitudes are (or can be) learned, and are hence renewable following re-learning, but some people would argue that attitudes are determined to a certain extent genetically.

Despite the obvious shortcomings of this extremely brief and simplistic presentation of the social scientist's interest in attitudes, it is clear that attitudes may be viewed as either malleable or non-malleable (i.e. resistant to change). If they are malleable, what factors in society have produced them? If they are non-malleable, what conflicts may and do arise between them and social structural factors? In either case can one use a knowledge of attitudes to predict behaviour, and if so, to what extent? Have attitudes, in other words, an explanatory value as regards conduct?

Eiser (1980:17 et seq.), a social psychologist, takes a somewhat different stand, arguing that any study of attitudes which is based on an inference of their involvement in causality (regardless of whether the focus is on attitude as the response to an attitude object, or whether attitude is the instigator of a behaviour response) is doomed to failure: it is impossible to observe people's attitudes (i.e. their feelings of like or dislike, approval or disapproval, attraction or repulsion, trust or distrust, etc.) directly, and hence the validity of the inference of their existence, let alone the inference of their participating in a causal relationship, is unsustainable:

> If statements about a person's attitudes are inferences from behaviour how can one tell if such inferences are valid? How could one observe any statistical association between overt behaviour and inner experiences which by definition are unobservable? (Eiser 1980:18)

Obviously choosing his words carefully, Eiser notes that although an attitude statement *expresses* a speaker's attitude, it is not a *description* of the speaker's attitude:

> This is what attitude statements do: they describe attitude objects, and express attitudes. (Eiser 1980:19)

He continues:

> In attempting to define attitudes in a way which accounts for the relationship between feelings and inner experiences on the one hand and observable verbal and non-verbal behaviour on the other, many social psychologists seem to have made two crucial assumptions. The first is that attitudes are distinct entities with

> an independent existence. The second is that their relationship to observable behaviour is causal. I regard both these assumptions as fallacious. The relation of attitudes to expressive behaviour seems to me to be essentially a logical one, analogous to the relation between meaning and utterance. We need to assume that words have meaning to understand verbal behaviour, but we do not need to regard the meaning of a word as something which has an independent existence, nor as a distinct entity which causes the verbal behaviour. Just as words have meaning, people have attitudes, and the concept of attitude is no less important for understanding human social behaviour than is the concept of meaning for understanding language. A person's attitude is the meaning of his expressive behaviour. (Eiser 1980:19)

These are provocative words, for they call into question the very notion of measuring attitudes and the instruments typically used to measure attitudes. We do not usually measure words, or the meaning of words; we do not usually quantify words and place them on a scale or continuum; we simply understand words, and very often we have a different understanding of words to the person next to us. Eiser's point is that we understand attitudes in essentially the same way as we understand words. Any attempt to measure attitudes with some sort of numerical index and place them on a linear scale, and equate them with more 'this' or less 'that,' is not just a gross simplification but is essentially misdirected. An index purportedly representing a person's attitude does not offer a simplified picture of some mental construct in the observed person's head, but offers a summary of *our own* understanding of this person's expressive behaviour. What is presented on the scale is not 'something' in the observed, but 'something' in the observer. In our inevitable attempts to make sense of other people, to assign structure to their behaviour and to eliminate as far as possible all wisps of random disorder, we the observers make potted summaries of their behaviour, call these summaries attitudes, and then use measurement scales to help us compare the attitudes of different people, so allowing ourselves to understand these people better, to make sense of what they say and do.

This view of attitude as being external to the observed individual still holds true when we speak of our own attitudes: in making statements which express our own attitude, we may be implying that we have some inner feelings and opinions (such statements are after all kinds of value judgements), but such implied inner feelings and experiences are not the *reference* of our statements. Our attitude statements about ourselves reveal our observation, interpretation and evaluation of ourselves seen as an object in the 'external world': how do we make sense of our environment and of what we do in it? The situation is analogous to that of talking about ourselves in the third person: 'They [understand 'we'] don't approve of drinking and driving.'

Regardless of whether we accept Eiser's interpretation of attitudes or not, attitude measurement and attitude scales are certainly possible. However, it is incumbent upon all researchers to explain precisely what their measuring tools are and what their scales represent, for both their tools and scales involve an assumption as to what attitudes are, how they are formed and how they change,

and the results of any measurements made need a theory of attitudes in order to be interpreted.

My interest in attitudes to mispronunciation must be carefully separated from the social scientist's concern (or lack of it) with attitudes. I am not worried about where a listener's attitude towards a feature of mispronunciation or foreign accent comes from, what determines it and causes it. I shall not attempt to explain or predict what social structural or other factors prompt the existence of particular attitudes towards pronunciation. I am not concerned with whether an attitude is fixed or malleable, whether it changes or not as a result of structural or whatever other changes in society, though I shall be interested in trying to discover if there are any patterns of correlation between particular attitudes to particular features of mispronunciation and, for example, age and gender (socio-economic background and foreign language experience and proficiency might also prove to be interesting points of correlation but are beyond the scope of the present study). I shall not endeavour to show a link between attitudes to mispronunciation and conduct (even though anecdotal evidence of such a link is readily available and bountiful), nor shall I argue the cases for conduct being the result of attitudes to mispronunciation or for conduct being what it is despite particular attitudes to mispronunciation. All such matters are for those studying attitudes and behaviour in one and the same person. I shall limit myself to studying the existence of attitudes in one person or group of persons (the listeners), and I shall have only a passing regard for whether another person or group of persons (the speakers) varies behaviour (pronunciation)—or could be persuaded to vary their behaviour—when faced with the attitudes (and possible behaviour and reactions) of others (listeners).

2.1.1. The term attitude

As Baker (1992) points out, the term attitude appears transparent in its simplicity for we all talk about people's attitudes, about a positive attitude, an unhelpful attitude and so on. Moreover, we have probably all heard of attitude surveys, perhaps even been interviewed ourselves, the results of such surveys being used to show the current frame of mind of a particular section of the public, or show the change that has taken place in the public. Often such surveys lead to proposals and recommendations for future courses of action (administrative, political, commercial, etc.), form the bases of future policy documents, and thus underpin actual decisions, actions and their consequences.

The term attitude is certainly commonplace, but we need to remember that attitudes cannot be observed directly: attitudes are hypothetical 'constructs.' We can no more observe someone else's attitudes directly than we can put our own attitudes down on the table for someone else (or even ourselves) to look at. (The same, of course, can be said of, for example, intelligence, motives and pain, or the wind, time and electricity.) To try and express what our attitude is towards something or someone usually results in a stream of imprecise periphrastic mumblings: we cannot find the right words, and we claim that words in any event fail to capture and express the subtlety of our attitudes. Other people, there-

fore, who wish to know what our attitude to something is receive little help from our own clumsy attempts to explain our attitude and verbalise it; they are obliged to try and divine our attitude from the regular consistent patterns they think they detect in notably our speech and our behaviour. For example, our attitude to smoking, to the ordination of women, or to drunken drivers, is a complex of factors not readily put into words. Any short sentence which attempts to express our attitude to any one of these things will be hopelessly imprecise for it will inevitably miss the richness and the nuances of what constitutes our attitude.

A definition of what an attitude is, however much we feel we need one, is difficult: a definition, by definition, is never absolute.

Thurstone's (1931:20) early definition is vague: 'Attitude is the affect for or against a psychological object.' Allport's (1935) definition, often quoted (or misquoted), is superbly sweeping in its broadness, imprecision and superficiality: an attitude is 'a mental or neural state of readiness, organised through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.'[1] This mentalist definition recognises that attitudes, being mental or neural, are not available for direct observation, and implies that they are either intervening variables (i.e. are only theoretical concepts with no 'reality') or hypothetical constructs (i.e. are actual, real but unobservable and inaccessible 'realities') which are inferred by the observer (whether that be the person whose attitudes are being observed or someone else) in order to explain the dissimilarity between different people's responses to the same stimuli in the same situation. Allport thus dismisses the positivist approach which refuses to countenance attitudes as intermediate constructs or mediators between stimuli (antecedents) and responses (consequents), and leans rather towards a mediationalist approach, a development from the pure behaviourist tradition of psychology: he claims that while past experiences (stimuli) may shape future behaviour (responses), attitudes act as a moulding and determining conduit between the experience and the behaviour, a view echoed by Cardno (1955), whose views are summarised by Shaw and Wright in the following terms:

> Attitude entails an existing predisposition to respond to social objects which, in interaction with situational and other dispositional variables, guides and directs the overt behavior of the individual. (Shaw & Wright 1967:2)

Significantly Allport seems to envisage the possibility that an attitude is as much the outcome of behaviour as the determinant of behaviour, and this leads McGuire (1968:146–7) to label this approach as either (i) class-inclusionist (i.e. mediating constructs exist on both the antecedent and the consequent sides, the stimulus or antecedents producing a covert response, which produces a covert stimulus, which produces the response or consequents), or (ii) interactionist (i.e.

---

[1] The quotation is from Allport (1935) 'Attitudes,' reputedly printed in Murchison's *Handbook of Social Psychology*. I have not yet managed to see this book. Jaspars (1978:256–7) gives the quotation as I have presented it (adding a succinct but harsh critique); Lemon (1973:8) confirms the quotation. However, McGuire (1968:142), giving a detailed exegesis of the text, has a shorter slightly different version: '. . . exerting directive or dynamic influences *on behavior*.' (My emphasis.)

several antecedents *and* the interaction of those antecedents determine a mediating attitude, and the resulting behaviour is a set of interacting and counterbalancing responses or consequents). As is immediately apparent, Allport effectively precludes attitudes from any direct investigation for it is impossible to discover what past experiences moulded someone's present behaviour or to record what present experiences are shaping their actual future behaviour. The mentalist approach to attitudes has a problem, how to devise experimental methods for attitude measurement which do not resort to questionable self-reports or to equivocal inferences from behaviour patterns. Williams concludes carefully:

> Attitude is considered as an *internal* state aroused by stimulation *of some type* and which *may* mediate the organism's subsequent response. (Williams 1974:21. My emphasis.)

Bem (1968:197–8), claiming to be 'an unreconstructed behaviorist' (i.e. someone who strictly limits observation to observable details), seizes upon this heresy of equating attitudes with 'internal mediating responses.' According to Bem, true behaviourists, and indeed all who try to measure attitudes, whatever their theoretical standpoint, are forced to the practical operational conclusion that 'an attitude *is* an individual's self-description of his affinities for and aversions to some identifiable aspect of his environment' (1968:198). Bem thus refuses to countenance intervening variables and hypothetical constructs to explain the relation between experience and behaviour: we observe our own behaviour, or that of others, and depending on what we see, we say that our (or someone else's) attitude is racist, religious, intolerant, or whatever. Attitudes are 'self-descriptions' or 'other-descriptions.'

Bem's behaviourist viewpoint implies, and indeed he explicitly states it elsewhere, that 'behavior causes attitudes' (1970:54), 'attitudes follow behavior' (1970:67). He says:

> Attitudes are likes and dislikes . . . are our affinities for and our aversions to situations, objects, persons, groups, or any other identifiable aspects of our environment, including abstract ideas and social policies. . . . Our likes and dislikes [i.e. our attitudes] have roots in our emotions, in our behaviour, and in the social influences upon us. But they also rest upon cognitive foundations. (Bem 1970:14)

However, the labels we use to describe ourselves, i.e. our attitudes, may turn into guiding mottoes which influence our subsequent behaviour; and if we learn of the attitude-labels that others use to describe us, they too may influence our later behaviour.

The strictly behaviourist approach to attitudes has been criticised for its theoretical and experimental implications: if attitudes exist only as the outcome of specific stimulus situations and are thus dependent variables, then it is impossible to explain other behaviour patterns in other situations (Agheyisi & Fishman 1970:138).

Several studies and definitions followed after Allport and Bem until Kidder and Campbell claimed—in a deceptively perfunctory and desultory manner—that

> a host of seemingly unrelated terms such as acquired drive, belief, conditioned reflex, fixation, judgment, stereotype, and valence, to mention only a few, are all functionally synonymous with the concept of attitude. All describe the residues of past experience which are the stuff of which attitudes are made. They are the underlying processes, or the behavioral manifestations of underlying processes, which are products of learning. (Kidder & Campbell 1977:370)

All definitions inevitably lead to contention: either the definition is too restrictive or too vague, too specific or too general. Sherif and Sherif (1967:109) stress the involvement of judgement processes, saying that every attitudinal reaction implies comparison, evaluation and choice among possible alternatives, i.e. judgement. McGuire (1968) tersely claims that attitudes locate objects of thought on dimensions of judgement. Lemon (1973:8–15), having cited several attempts to define what attitude is, concludes that attitude may be taken to be: *either* a form of behaviour (a so-called probabilistic definition, since attitudes are equated with the recurrence of an observable and consistent, organised and predictable behaviour pattern resulting from a particular attitudinal stimulus); *or* a deterministic psychological process (a so-called latent process definition, since attitudes are said to manifest themselves in consistent syndromes of response to given attitudinal stimuli, but are assumed to be hidden or hypothetical variables functioning within the individual to mould or determine the observable behaviour: attitudes, therefore, are hidden variables mediating between stimuli and behaviour)[2]; *or* a factor working together with social structural phenomena to determine conduct and behaviour. Ajzen (1988:4) says an attitude is 'a disposition to respond favourably or unfavourably to an object, person, institution, or event.'

In some research, particularly educational research, the definition of attitude makes a lot of the fact that attitude is both input and output in many activities. Without an initial predisposing attitude many courses of action would never be undertaken; without the particular attitude the manner in which the action is performed would be something entirely different; and the outcome of the course of action is the reinforcement of an existing attitude or the formation of a new attitude. Attitudes are thus predictive, determining and product variables.

In addition, attitudes are generally considered to form a coherent system. An individual's attitudes are moulded in particular learning situations by specific external referents and by a unique set of evaluative concepts, and thus, it is claimed, they are necessarily inter-related. The fact that the consistency and logicality of the system may not be immediately apparent to other people, or even to the individual, is neither here nor there. The attitudes which are more central to the system, those strengthened by a history of reinforcement, are more stable and resistant to change, inert; and the more peripheral the attitude, the more unstable, atrophied and volatile it is.

Some social psychologists (usually those of a mentalist rather than a behaviourist propensity) suggest in their definitions that attitude is the nebulous

---

[2] This is as presented by Jaspars (1978:260) who points to DeFleur & Westie (1963). Elms (1976:21) also points to DeFleur & Westie.

intangible product of (or reflection of) a merging of three components: cognitive, affective and conative. For example, Rosenberg and Hovland (1960:3) state that 'attitudes are predispositions to respond to some class of stimuli with certain classes of [cognitive, affective and behavioral (conative)] responses.' Triandis similarly suggests that attitudes subsume cognitive, affective and behavioral components, claiming that 'an attitude is an idea, charged with emotion which predisposes a class of actions to a particular class of social situations' (Triandis 1971:2–3). However, Shaw and Wright specifically reject this tripartite conception of attitude:

> Attitude is best viewed as a set of affective reactions toward the attitude object, *derived* from the concepts or beliefs that the individual has concerning the object, and *predisposing* the individual to behave in a certain manner toward the attitude object. . . . Our conception rejects the notion that attitudes are composed of three components. Rather, the affective reactions specified by the traditional analysis constitute the attitude. . . . (Shaw & Wright 1967:13. My emphasis.)

McGuire (1968:155–7) and Agheyisi and Fishman (1970:138–9), on the other hand, again advocate the tri-partite view of attitude based on cognition, affect and conation (or habit). The cognitive component (concerning thoughts and beliefs), the affective component (concerning emotions and feelings), and the conative component (concerned with intentions or readiness to do something), work together (somehow in a manner not usually specified) to form, so it is claimed, an attitude or attitudes. These three components, however, may not be congruent with each other. Thus openly expressed beliefs may belie covert prejudices, hopes and fears, and may also contradict actual behaviour. For example, the expression and acknowledgement of a religious faith, contrasting with deep-seated emotional responses, contrasting in turn with a lack of readiness to live according to the professed religious doctrine, may make up a self-contradictory whole called a religious attitude. Curiously, no one seems to explain how the three purported components of an attitude are inter-related, how they interact, how they are to be conflated (see Agheyisi & Fishman 1970:139). Lemon (1973:16–7) wonders (somewhat disparagingly) whether there is any empirical evidence for such a highly complex view of attitude, but it was another decade before anyone tackled this problem. For many years everyone seemed to restrict their attitude measurement to the affective component (the evaluative pro–con component), brushing the putative cognitive and conative components to one side.

However, Breckler (1984) eventually tried to test the validity of the tripartite model of attitude structure. His initial definitions are illuminating:

> Affect refers to an emotional response, a gut reaction, or sympathetic nervous activity. . . . Behavior includes overt actions, behavioral intentions, and verbal statements regarding behavior. Beliefs, knowledge structures, perceptual responses, and thought [constitute the cognitive component]. (Breckler 1984:1191)

His two carefully constructed experiments seemed to confirm that the three components said to make up attitude are indeed distinguishable, and that these three components do *not* represent one single underlying construct, viz. 'attitude': atti-

tude is not unidimensional but three-sided. Attitude researchers, he concluded, should measure all three components separately, or they should specify which of the three is their focal concern.

Lemon (1973:15) speculates about whether attitude is (as generally assumed) a totally *learned* characteristic (hence open to modification and change) or whether it has some genetic origin, a significant but all too often overlooked facet. Sherif and Sherif certainly recognise that attitudes are acquired or learned:

> Operationally, an attitude may be defined as the individual's set of categories for evaluating a stimulus domain, which he has established as he learns about that domain in interaction with other persons and which relate him to various subsets within the domain with varying degrees of positive or negative affect. (Sherif & Sherif 1967:115)

But Lemon sees that we must go further: are our attitudes predetermined (at least to some extent) by our genetic (family, racial) make-up? Definitions of attitude should necessarily address this issue.

Definitions also need to tease out the differences between the plethora of closely related words: are attitudes the same as 'beliefs' and 'values' (Elms 1976), as 'opinions,' 'habits,' 'motives' and 'traits' (Lemon 1973:23–7; Shaw & Wright 1967:4–6), as 'ideologies,' 'concepts' and 'constructs' (Baker 1992:13)? **Beliefs** would seem to be assumptions about the probability that something exists (e.g. 'I believe in God'), or that something possesses certain characteristics (e.g. 'I believe the earth is flat,' 'I think she's innocent'); they are more cognitive in nature than affective or conative; and they seem to provide us with cues for potential behaviour and action, though we do not necessarily act in accordance with them. **Opinions** and **attitudes**, although perhaps merging imperceptibly for most lay people, are often distinguished by saying that **opinions** are essentially responses, being overt verbalisable beliefs usually concerned with matters of fact, containing or involving no affective element and no desire, while attitudes are essentially response predispositions, being latent, difficult to verbalise and usually concerned with matters of taste and evaluation. Nevertheless, opinions are often taken as the open verbalised manifestation of an attitude (see Thurstone in particular); it is perhaps significant that *opinion* polls usually aim to measure *attitudes* and *behavioural intentions* even though the questions they ask are typically phrased in terms of *beliefs*! **Values** would appear to imply ultimately desirable 'worth-full' goals or end states and hence involve a tendency toward specific patterns of behaviour: for example, someone may know that family unity, true friendship or financial security are (probably) not 'true,' or realisable, but they wish they were and behave in a fashion which increases (however slightly) the chances that at least for them they become so. **Habits** (a term much loved of course by psychologists of a behaviourist bent) would seem to suggest acquired enduring patterns of behaviour (or action tendencies) which, possibly evaluative, remain constant across a wide range of possibly independent situations, while attitudes, always evaluative, involve patterns of behaviour which may appear to be totally non interrelated but which are nevertheless oriented to one particular social object or phenomenon. Personality **traits** (the betrayal of

personality traits through language forms a vast body of research closely related to my present study) are generally thought to be more stable and less malleable dispositions than attitudes, to be non-evaluative (unlike attitudes), and to be directed towards no specific target (unlike attitudes). Despite some disagreement (even among academic scholars), **motives** (or **motivation**) would seem to suggest some drive and to be goal specific, but attitudes have no drive and are object specific. **Ideology** would seem to suggest an elaborate codified set of beliefs, a rationalisation of one's behaviour, a philosophy of life, in brief 'a global attitude' (Baker 1992:15).

Baker (1992:30–33 *inter alia*), while examining the nature of attitudes to language rather than of attitudes in general, forcefully focuses attention on the theoretical and practical problems resulting from the acceptance of the uni- or multi-dimensionality of an attitude. Is an attitude towards a specific object or phenomenon or whatever a unitary construct within the conceptualisation of different people? Does everyone share the same unitary construct? Or is an attitude a construct with a discrete number of dimensions (which are possibly hierarchically arranged)? The answers to these questions have far reaching implications. Some scholars have taken an *a priori* stand, designing their research on the premise that attitudes are (or are not) one-dimensional; others have taken an *a posteriori* stand and this has led them into, for example, factor analysis in an attempt to tease out the possible multiple strands which make up an attitude.

Many scholars understandably make much of the claim that attitudes vary in quality and intensity: this variation permits investigations using a bi-polar continuum ranging from positive through neutral to negative, the valency reflecting the quality, and the extremes of the continuum indicating intensity. However, the idea is not without its problems. For example, intensity may be involuted, increasing (positively or negatively) to a certain point and thereafter decreasing (Guttman 1954, referred to by Shaw & Wright 1967:7), but how does one discover where that point is, and how does one measure the degree and strength of involution? The neutral point is also difficult to interpret: does it indicate no attitude (neutrality), ambivalence, uncertainty or indifference? (See Shaw & Wright 1967:21; Kaplan 1972; Klopfer & Madden 1980.)

Shaw and Wright's definition of attitude is the following:

> [An attitude is] a relatively enduring system of affective, evaluative reactions based upon and reflecting the evaluative concepts or beliefs which have been learned about the characteristics of a social object or class of objects. . . . It is a covert or implicit response. . . . It is a drive-producing response which elicits motives and thus gives rise to overt behavior. (Shaw & Wright 1967:10–11)

Summers gives the following succinct summary of the areas of substantial agreement concerning what an attitude is, or what attitudes are:

> First, . . . an attitude is a predisposition to respond to an object rather than the actual behavior toward such object. . . . Second . . . attitude is persistent over time. This is not to say it is immutable. . . . [Third] Attitude produces consistency in behavioral outcroppings. . . . Fourth . . . attitude has a directional quality . . . [and] has a motivational quality. . . . [i.e.] attitude connotes preference regarding

> outcomes involving the object, evaluations of the object, or positive–neutral–negative affectations for the object. (Summers 1977:2)

And McGuire says:

> There are numerous desiderata for [definitions of attitude]—testability, parsimony, heuristic provocativeness, relatedness to other theoretical constructs, generality, etc.—and it is unlikely that one choice of definition will optimize all of them. (McGuire 1968:149)

In the end we have to gauge the value of a definition according to its usefulness: does the definition clarify matters for us? does it focus our thoughts? does it provide a firm basis for our investigations? As Shaw and Wright (1967:3) point out, the definition of an 'attitude' as an evaluative reaction based upon a cognitive response, whatever its weaknesses, at least allows one to use attitudinal scales composed of statements designed to prompt various degrees of positive or negative agreement: the degree of endorsement or rejection of the statement (the affective reaction) is interpreted as signalling the positive or negative evaluation of the endorser, their attitude.

Most people studying attitudes towards accents would seem to make little effort to define what they mean or understand by an attitude. By using bi-polar adjective pairs, there is an open admission that attitudes are affective (listeners *feel* the accent suggests the speaker is such and such), yet discussions and conclusions would often seem to imply a cognitive component (listeners *think* or *believe* the speaker is such and such). Only rarely do studies acknowledge a conative component or attempt to cater for one (the listener *intends* to help the accented speaker, or is ready to do something for them).

2.1.2. The measurement of attitudes

Empirical social psychology has for some sixty years now devoted enormous effort to the study of attitudes and paid scant attention to beliefs and values. This is not surprising as an attitude construct is more amenable to investigation than a belief or value. The result is an important body of knowledge on investigatory techniques for attitudes, measuring and scaling principles, and attitude theories.

Thurstone and Chave (1929), in an early and seminal paper, warns that most people intuitively realise that attitudes are highly complex psychological constructs, and that few people suspect it is even remotely possible to measure them. Moreover, since attitudes are abstract, it is necessary to approach them indirectly and measure them indirectly, but this complication is no reason to give up before trying. (As the authors say, no one hesitates to measure the height of a table even though tables are complex objects.)

Thurstone and Chave's approach is to observe opinions in order to get at attitudes:

> The concept 'attitude' will be used here to denote the sum-total of a man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specific topic. . . . The concept 'opinion' will here mean a verbal expression of attitude. (Thurstone & Chave 1929:6–7)

So, while recognising that both attitudes and opinions are subjective and highly personal matters, the authors propose that it is the examination of opinions that

allows us to measure attitudes. Opinions are taken to reflect attitude. Inaccuracies are inevitable because people do not always reveal their true opinions (they are too polite, wishing to please the investigator; they wish to deceive; they wish to disrupt the experiment of an investigator they have taken an aversion to; etc.), but similar discrepancies between measuring instruments and phenomena to be measured are easy to find (the discrepancy between the marks on a thermometer and actual room temperature is offered as an example). The authors also stress that the rating or value eventually apportioned to a person's attitude cannot in any way be taken to be an indication of their actual or future behaviour. This limitation necessarily precludes any check on the attitude scale based on cross-comparison with actual behaviour (see concurrent validity below). They also warn against interpreting a final attitude scale in too simplistic a fashion: the scale is linear (the informant is 'more' or 'less' religious or whatever), but the opinions expressed and the attitudes deduced from the expression of those opinions remain, of course, highly complex and multi-dimensional. (The height of a table, say 2 feet 4 inches, is not affected by its three-dimensionality.)

Thurstone and Chave emphasise that one major problem of attitude studies is the specification of the attitude variable under investigation. In other words, if one is examining, for example, people's attitudes towards religion, is the attitude label 'religion' (or 'religiousness') defined clearly enough? Rather than looking at religiousness, are we in fact looking at religiosity? (See 'construct validity' below.) This is no quibble over semantics: the attitude studied determines the thrust of the questions asked when trying to elicit people's opinions. The measurement of the attitude variable has to be limited ruthlessly to the measurement of only that variable and nothing else: it should not be distorted by irrelevant details. (The diminishing tapering diameter of a table leg in no way affects the height of a table.)

In addition Thurstone and Chave insist that questions must force answers of a clearly 'more' or 'less' type. Respondents are thus not divided into two groups (the 'Ayes' and the 'Naes'), but distributed along a continuum, and this spread can be shown by a frequency outline on a graph. Assuming there are several questions trying to ascertain a person's attitude, the respondent's attitude is reflected on the attitude continuum scale not by a point but by a specific narrowish range, and the attitude is read as the mean of that range. Respondents are thus differentiated one from each other both by the mean attitude they betray and by the range of their attitude. Attitude frequency graphs can be used to compare different groups of respondents (different populations), the spread and range of attitudes indicating the heterogeneity of the different groups, but it should of course be remembered that there is no reason whatsoever to expect any so-called normal Gaussian distributions here. No group of people will ever be 'normally' distributed in their opinions about something. For example, take a random sample of English people and ask them their attitude towards the eating of horsemeat: how likely is it that the distribution of answers for and against will show a normal Gaussian distribution? This is a basic fact which many people studying attitudes towards accents have chosen to ignore.

The basic problem that all investigations into people's attitudes face is that attitudes are inaccessible and cannot be measured directly: they can only be inferred. Attitudes are abstractions: they are intangible and invisible. Cook and Selltiz (1964. Reprinted in Warren & Jahoda 1973:368) conclude that investigations are obliged to resort to five different ways of measuring attitudes:

1) measures based on self-reports of beliefs, feelings, etc.
2) measures based on observations of overt behaviour.
3) measures based on reactions to or interpretations of partially structured material.
4) measures based on performance in specifically presented objective tasks.
5) measures based on physiological reactions.

All these methods are fraught with difficulties and drawbacks, and it is indeed possible that results obtained with one method are not concordant with results obtained with another. However, few investigators can afford the luxury of conducting studies where the validity of the results is strengthened by the use of multiple measuring techniques. Most investigators use only one technique, the most commonly used one being the first one: interviewees are asked, for example, to reveal their beliefs or feelings concerning some attitudinal object (a cognitive or affective bias). Only rarely are interviewees asked to state how they would behave towards the object (a conative bias).

The questions one asks people to find out their present attitude are merely trying, in a round about way, to focus on what the attitude is. No one question can reveal a person's attitude, and one has to content oneself by saying that because a person approves, or agrees with, points X, Y and Z, then that person's attitude by deduction must in all probability (but not necessarily) be A. In other words one never actually measures the attitudes; one merely counts how many statements a person agrees with or disagrees with and so on. And there is of course no guarantee that they will agree or disagree with the same statements tomorrow! Attitudes are (to a greater or lesser extent) volatile. It is rather as if one attempted to measure the height of a table by measuring its shadow when cast on an irregularly surfaced wall by a slowly but steadily setting sun. All attitude measuring scales have to account for and handle this problem of indirect measurement, and interpretation of the results must in addition heed the possibly transient and ephemeral nature of the attitude measured. As a result all studies using self-reports to measure the intensity of an attitude, or attitudes (rather than counting up how many statements a person agrees with), require various index measurement scales, scaling devices, and analytical approaches, the choice depending on the type of questionnaire used.

Many types of index measurement rating scales can be used for measuring attitudes, but most commonly questionnaires enquiring about attitudes give data of two types: nominal and ordinal. Interval scale questionnaires on attitudes (how much more or less 'X' person 'A' is relative to person 'B') are all but impossible.

Nominal questionnaires basically seek answers of a 'yes' or 'no' variety, 'headcounts,' with mutually exclusive and exhaustive categories, allowing one merely to give a simple name to the particular group (positive or negative, for or

against) that the respondent belongs to (for example, pacifist v. militarist; pro-legalisation of cannabis v. anti-legalisation of cannabis). The results give one a simple dichotomous categorisation. Occasionally multiple category scales are used, but in any event the data produced is discrete, and comparative gradation of the more or less variety is unfeasible.

Ordinal questionnaires, on the other hand, seek answers which require a classification along a continuum such as 'best–better–OK–worse–worst,' or 'excellent–good–OK–poor–awful.' Such index measurement scales use clear labels of typically an evaluative nature. Typically no integers are used in the questions asked: the informants only see the verbal labels. Beauty contests are strictly ordinal, the judges placing contestants in an order 1/2/3/, etc. School class lists are the same (top of the class through to bottom of the class). However, the intervals between the points on an ordinal scale are never regular: the beauty queen is not twice as beautiful as the runner-up, three times as beautiful as the third prize winner, and so on. A variation of this sort of questionnaire might ask an informant to agree or disagree with a given statement and then, in the processing of the results, points might be allocated accordingly: strong support → +3, moderate support → +2, slight support → +1, slight opposition → -1, moderate opposition → -2, strong opposition → -3. (Sometimes the points awarded may be from +1 through to +6, so avoiding negative numbers.) Informers have to choose a 'category' in order to indicate the degree of their agreement or disagreement, and typically there is no zero category: they cannot indicate indifference, ambivalence or uncertainty. Such scales are commonly used to show, somewhat crudely, the intensity of the attitude measured, but it is obvious that the 'scores' are not strictly available to arithmetic manipulation.

Sometimes an index measurement scale may be designed so as to associate verbal categorisations with a range of points on the scale rather than with just one point (Figure 2.1).

0 – 20	21 – 40	41 – 60	61 – 80	81 – 100
awful	poor	OK	good	excellent

Figure 2.1

A variation on this type of scale might associate the labels with particular points on the scale rather than with a particular range of points, thus allowing people to choose points 'between' the labels.

Most people have come across index measurement scales like these in the form of public opinion polls: specially chosen samples of the public are asked to indicate their attitude towards something by, for example, choosing a response category from among 'strongly approve,' 'approve,' 'disapprove,' 'strongly disapprove.'

It should be carefully noted that unlike interval and ratio measurement procedures these index measurement scales are internally inconsistent. The intervals between the points on the scale are not equal, and there is no top or bottom absolute value. Moreover, the fact that someone classifies one feature of, for

example, mispronunciation (or clothing, or political opinion) as 'awful' in no way binds them to mark another feature of mispronunciation (or clothing, or political opinion) as 'awful,' 'OK', or whatever: there is no cross-correspondence between the objects towards which people are expressing their attitudes. The measuring tool does not have equal units, does not have a standard measuring scale and cannot be used to predict how people will react to another object. To minimise possible deficiencies of internal consistency Kline (1993:10) suggests (somewhat laconically) that the minimum acceptable sample size is 100: any irregularities will probably then balance themselves out, though of course reliability should still be checked by using, for example, coefficient alpha.

An index measurement scale is also problematic when there is an uneven number of points and a central point, a so-called **zero point**. How do people interpret the central point? Are peoples' reactions to the central point consistent? For example, informants scoring an overall 'neutral' score on a Thurstone-type scale do not necessarily consider themselves (indeed are unlikely to consider themselves) as being entirely neutral towards the attitude object. It is possible that the apparently 'neutral' attitude is the result of an equal number of strong negative and strong positive reactions to the attitude object. Alternatively, the central point on a scale may be chosen because the extreme points either side are too 'intense,' but that is no guarantee that the central point is totally devoid of positive or negative value. In other words, the investigator cannot simply interpret the central point as 'neutral,' i.e. as indicating 'no opinion' or 'indifference.' The point may also on occasion indicate 'ambivalence' or 'uncertainty.' (See Kaplan 1972, Klopfer & Madden 1980.)

The tool cannot even be exploited in the confident knowledge that different informants will use it in the same way: while one person may willingly use the whole range of points on the scale, another may not like using the point 'awful,' feeling it is far too negative, and so they always employ 'poor' as the bottom point of their scale. This semantic bias leads to **the error of leniency** (or its converse, for demanding interviewees who cannot bring themselves to use 'excellent,' **the error of severity**). On the other hand, it is also widely recognised that some people dislike, or disagree with, something more violently than they like, or agree with, something: thus the scale becomes warped at the end which expresses the view or opinion antithetical or contrary to that of the interviewee. A further complicating detail is the fact that some people avoid both extremes on a scale, i.e. they shun choices 1 and 5 on a five-point scale, and 1 and 7 on a seven-point scale. This results in the so-called **error of central tendency** (or 'end-point' effect), and if noticed early enough may tip the scales in favour of the use of the seven-point scale rather than the five-point scale (or an even-numbered scale with no neutral mid point).

A similar complication is **the halo effect,**[3] where respondents classify something according to a global impression they already have of the attitude object

[3] N.B. Baker (1992:19) has a different understanding of what the halo effect is: he states that it is the conscious or unconscious tendency of people to give socially desirable answers and to put themselves in the best light. However, Lemon (1973:92) clearly sees the halo effect as a

(cf. the 'love is blind' and 'teacher's favourite' syndrome), giving a standard (e.g. acquiescent or evasive) response and transferring a general rating from one question to the next without actually considering the real content and intent of each individual question or the object to be evaluated (Nisbett & Wilson 1977; Cronbach 1946,1950; Bass 1955; Cook & Selltiz 1964). The resultant systematic errors in the measurement of attitudes are sometimes said to be attributable to 'response sets': **habits** are brought into play in situations which are simplified by viewing them (the situations) in terms of a few standard **frames** (see Esser 1993[4]).

Contrasting with the halo effect witnessed in those who do not think about what they are doing, some people think too much, indulging in doublethink: they provide the answers they think the interviewer expects or the answers which give them a more positive social image. This constitutes a **social desirability bias** (Edwards 1953, 1957b; Taylor 1961; Cook & Selltiz 1964; Diab 1967:141). This double-think process (called **'impression management'** by Fabrigar & Krosnick [1996:46] and **'subject expectancy'** by Brown [1988:33]) is similar to Helson's (1947) idea of adjustment in adaptation level, though the former produces only a temporary distortion in confessed attitude and the latter produces a more lasting adjustment. This problem of possible distortion in responses resulting from the presence of the investigator is also sometimes referred to as **'the observer's paradox'** (Labov 1972:209–10; Wardaugh 1998:18, 146).

A further complicating factor is the **Hawthorne effect**: subjects may be so chuffed (or annoyed) at being chosen for inclusion in an attitude study that their attitudes are distorted by their very pleasure and pride (or annoyance). Some people even argue that attitude questionnaires may provoke a **reactivity effect,** the questionnaire itself prompting the formation of previously non-existent attitudes, or at least the strengthening of attitudes which were merely latent or dormant. Thus where respondents are simply more willing to evaluate something more positively when it agrees with their pre-existing views the term **bias assimilation** is used, but where respondents' attitudes are prompted to become more extreme in the direction of an existing point of view this is called **attitude polarization** (Lord et al. 1979; Miller et al. 1993).

Compare all these problems associated with index measurement scales with the scale on a ruler: six inches long is six inches long whether you measure first the length of a cucumber, then that of a banana or whatever. The distance (or interval) between the inches remains constant across measurements, for you and for everyone else who uses the ruler: no one is going to say that they don't like the point marked twelve inches (or even thirteen inches!) and refuse to use it; no one is going to have a prejudice in favour of using twelve inches more frequently

---

tendency towards self-consistency, giving similar ratings to issues which are seen or thought to be related, although closer attention and examination would reveal them to be not as closely related as at first presumed.

[4] Esser (1993:293) claims 'it is impossible to completely eliminate errors in attitude measurement,' systematic errors attributable to response sets being particularly serious.

than another distance. The measuring scale on a ruler is consistent: it has an interval measuring scale.

Notice that the consistency just mentioned refers to the scale of the measuring instrument, *not to the interviewee*. Consider, for example, people asked their attitude towards Tony Blair. They might answer 'strongly approve,' and then in the next question concerning their attitude towards the Labour Party they might answer 'disapprove.' Without knowing the recent actions of Tony Blair, recent political events and so on, it is certainly possible that some people who simplistically and erroneously equate Tony Blair with the Labour Party will consider these two attitudes to be contradictory but there is nothing inconsistent in them at all: it is the measuring scale that is inconsistent. Interviewees may think Tony Blair to be sweet and cute because of the fresh-faced cherubic school-boy image that (in their opinion) he presents. But an index measuring scale cannot take this sort of external idiosyncratic feature into account. It is precisely because of this that most attitude studies try and ascertain someone's attitude to something by taking the average of three or four or more related questions (see Thurstone/ Likert scales below). It may seem a fairly crude attempt to avoid, or at least reduce, the effects of idiosyncratic bias, but it is comparable with the care that is generally taken in population sampling (or even question sampling) to avoid or reduce possible bias there.

It should be obvious from the above several paragraphs that responses given on these types of measuring scale are not necessarily a true reflection of a person's attitude. This caveat is often referred to as **'The Literal Interpretation Fallacy.'** Basically, a response may be distorted by linguistic, social, cultural, personal and ultimately idiosyncratic factors. Thus, the responses gained on these index measurement scales are meaningless in and of themselves (unlike the responses you would obtain after giving a ruler to twenty people and asking them to measure the length of a cucumber). The responses have to be interpreted, and neither the measurement technique nor the measuring scale have any point or use at all unless somehow or other we can interpret the results and give them meaning.

The reliability and validity of attitude index measuring scales are clearly a problem for anyone working with them. How can we be sure that the results are repeatable, i.e. reliable? How can we be sure that we are measuring what we want to measure, and that the results we have obtained with different people on one particular attitude scale are indeed valid, do reflect a difference in attitude and not something else? How can we be sure that the error creeping into the data for whatever reason (a faulty measuring instrument, the tiredness of subjects, an unsatisfactory test environment) is not jeopardising the integrity of the results? (See *inter alia* Shaw & Wright 1967:16–20; Hatch & Farhady 1982, Ch. 17.)

Reliability remains a major problem for all attitude studies, but there are four basic ways of tackling it.

The so-called **test-retest** method, where test-retest scores are correlated, can ascertain the stability and consistency of a test over time, but it is not always practicable. For example, respondents may dislike answering long lists of

questions twice or indeed be unavailable for a second test. In addition answers to the questions on the retest are coloured by the experience of the first test, and attitudes may have changed by the time the second test is administered: in other words the retest is not testing the same thing, and the results are necessarily incomparable.

The second way is the **parallel (or equivalent) forms** method, two versions of the questionnaire being given to the same people, possibly at the same time, and the results being correlated. But again there is a problem for studies of attitudes: how does one ensure that the second version of the questionnaire is equivalent to the first and focuses on the same 'universe of content,' i.e. spans and embraces exactly (and only) the same range of attitude dimensions? It's impossible. The parallel forms test ends up being more a test of equivalence than a test of reliability.

The third way is the one most investigators of attitudes resort to: the **split-half** method. The questions in the test battery are divided into two matched halves and the results correlated to check for reliability.[5] In effect, though, the split-half technique is merely a test of internal consistency, of homogeneity, not really a test of reliability.

The fourth approach focuses on **inter-rater reliability**: do raters understand the task at hand in the same way? have they understood the questions in the same way? do they come from the same population? If there is wild discrepancy between raters, i.e. if they fail independently to give more or less the same 'score,' is it the result of unreliability in the variable being tested (in the test itself), in the instructions, in the informants or what?

The issue of reliability has important implications on both questionnaire design and the scope of a study (Moser & Kalton 1971:355). First, questionnaires are necessarily limited in length because respondents will refuse, or be unwilling, or be unavailable, to answer long lists of questions. Second, to reduce the effects of bias on any one answer (and hence increase reliability) several questions are needed to focus on and pinpoint any particular attitude. The consequence is that the number of attitudes charted is restricted in any one survey in order to achieve high reliability. If one wanted to chart a whole range of different attitudes and still aimed for high reliability, one would need an extremely and unacceptably lengthy questionnaire.

Validity is also a major problem for attitude studies.

For a study to be valid *internally*, we have to ensure beyond any reasonable doubt that our independent variable is responsible for any systematic variations we may detect in the dependent variable(s). And for a study to be valid *externally*, we must be confident (again beyond any reasonable doubt) that our results and the conclusions we draw can justifiably be generalised to other comparable

---

[5] Moser & Kalton (1971:354) refer to the Spearman-Brown formula and the coefficient alpha, pointing to McKennell (1970). Shaw & Wright (1967:16–7) refer to the Pearson *r* coefficient for the test-retest method, and to the Spearman-Brown prophecy formula and the Kuder-Richardson formula 20 (K-R 20) for the split-half method. Kline (1993:39–40) also refers to Kuder-Richardson KR20 for testing the reliability of dichotomous data.

populations and contexts. These two goals of internal and external validity are essential for all empirical studies, though 'how much' validity is achieved in attitude studies is open to question.

Take **content validity**: if we have a set of three or four questions trying to pinpoint someone's attitude to something, do all of the questions in fact focus on the same attitude, and, perhaps more importantly, do they evenly capture or encompass the three or four major facets of an attitude? If some important feature or characteristic of an attitude is missed, then the attitude measurement will be distorted. Part of the problem here, of course, is the problem of defining the attitude under investigation, and deciding whether an attitude is uni- or multi-dimensional. The assessment of content validity unfortunately remains almost entirely subjective.

Some scholars insist that in addition to embracing the whole attitude, the questions we ask should embrace nothing but the attitude. In other words, the questions must make sense: the answers one obtains provide us with a variable value, but such a value is useless unless it makes sense to interpret it as an indicator of the attitude we are trying to measure. Additionally the questions should make the *most* sense, no other possible questions being more plausible prompts for providing the values we are searching for. All this is a matter of **logical validity**.

Test validity is sometimes taken to refer to the problem of ascertaining whether informants actually answer the question one thinks one is asking (and answer it honestly). If informants misunderstand the questions asked, then the test is hardly valid. The wording of the question may be obscure or ambiguous, or else the informant, believing the whole test to be a waste of time, is bloody-minded, refusing to understand, failing to concentrate, giving pat mechanical answers.

Concurrent validity is occasionally employed in some studies: results are compared with actually observed behaviour, but there is always a danger of a circularity here: already observed behaviour colours the forming of the questionnaire, and the results are interpreted when one knows what they should reveal.

More frequently researchers may turn to **predictive validity,** contrasting the results of their attitude questionnaires with subsequent behaviour. Political pollsters typically seek to increase their confidence in the validity of their measuring instrument by noting that their results have 'predicted' subsequent voting behaviour, but such a blatantly *post hoc* check is unsatisfactory: temporal succession is not evidence of a causal relation. Moreover, *post hoc* behaviour may be inaccessible, as is the case with attitudes towards mispronunciation: it is extremely difficult to observe present or subsequent behaviour which reflects a dislike of, for example, the confusion of sibilants.

It is perhaps surprising that Summers (1977:7) contends that 'the ultimate aim of attitude research is better understanding *and prediction of overt behavior*' (my emphasis). While researchers probably all seek more understanding, I don't believe anyone can expect attitudes towards accented speech to manifest themselves readily in predictable overt behaviour. Certainly if the present experiments

were to find that listeners react violently against the lack of aspiration in Finnish-accented English, that result in no way predicts present or future unfriendly behaviour, rejection, ostracism, or whatever in the listener. Admittedly Summers continues to report on studies which show that

> verbally expressed or self-reported attitudes do not correspond perfectly with overt behavior toward the attitudinal object . . . In some studies the correspondence is slight. (Summers 1977:7)

Note that 'the ultimate aim' espoused by Summers was not shared by Osgood and his colleagues, who tersely warn against overdrawing the goals of attitude scales, stressing that attitude scores will never allow us to predict actual behaviour in real-life situations. Attitudes, they conclude, only indicate disposition towards action or behaviour, and disposition

> accounts for only part of the intervening state which mediates between situations and behaviors. . . . While it is true that different attitudes imply different behaviors toward the objects signified, at least in some contexts, it is not true that the same attitude automatically implies the same behaviors. (Osgood et al. 1957:198–99)

Attempts to confirm validity by cross-checking with other *non*-index measurement techniques is extremely difficult, if not impossible. On the other hand, cross-checking employing more index measurement techniques, while often possible, serves no function: similar stimuli (measuring instruments) may produce similar responses (expressions of the presumed attitude), but unless dissimilar stimuli (dissimilar measuring instruments) can provoke the same responses (evidence of the same attitude), then nothing has been proved or disproved.

One serious danger to the validity of attitude research is the fact that the very research itself distorts the object of its enquiry, i.e. attitudes. The person administering the attitude test or questionnaire may affect the respondent, and the perceived purpose of the research may affect the response (Diab 1967:141; Cook & Selltiz 1964). Such potential sources of error and contamination (free-floating uncontrolled variables) must be recognised, though usually little can be done to ensure that they are totally eliminated.

One final type of validity which is often referred to in the literature is **construct validity**.

Since attitudes are constructs, i.e. hypotheses or inferences proposed to explain sets of observations, they are abstractions which cannot be measured directly. Hence measurers of attitudes have to validate their measuring instruments to ensure that they truly respond systematically to the attitudes they are supposed to be measuring. In other words, does the measuring instrument, the test, really measure the unobservable underlying theoretical/psychological construct (e.g. aptitude for languages, or intelligence, or, as here, an attitude) which the investigator has proposed? (See Luppescu & Day 1990.) Essentially construct validity is a check on the uni-dimensionality of the construct proposed: is the construct delimited in sufficiently clear, precise and narrow terms? The investigator first hypothesises as to the types and degrees of association between

the construct and other variables (such as other aspects of a person's personality), and then the extent to which the hypotheses and expectations are confirmed provides an indication of the test's validity. Unfortunately such a check is more a check on the hypothesising and theorising than on the test itself. For example, a differential group experiment could administer a test to two groups, one clearly with the construct under investigation, the other clearly without it. Alternatively an intervention experiment could be performed on the same group of informants: the test is administered, and then re-administered to the same group of informants after they have been 'treated' (taught, trained, whatever) in a way thought to have a direct influence on the construct under investigation. If the two test results are identical, then the construct validity of the test is obviously zero. (See Lemon 1973:40–44 and Brown 1988:103. For a report on the invalidity of a measuring instrument, see Luppescu & Day 1990.)

Thus the inter-dependent validity and reliability of studies of attitudes remain a serious problem. Moser and Kalton conclude their discussion of the issue with the following:

> It should be noted that the reliability and validity of a scale—unlike those of a foot-rule—are always specific to a particular population, time and purpose, not invariant characteristics. In any given study the researcher has to decide what degree of unreliability and invalidity he will regard as acceptable. (Moser & Kalton 1971:357)

Shaw and Wright point out that while the assessment of an individual's attitude is open to question, the assessment of attitudes held by groups of individuals is less problematic. As they express it:

> Errors of measurement may be assumed to be randomly distributed about the mean, so that with a sizeable number of subjects in each group the obtained mean attitude score approximates the true mean of the population. Consequently, an attitude scale that measures individual attitude imperfectly may yield a reliable and valid measure of the mean attitude held by members of a group. (Shaw & Wright 1967:565)

A crucial remark that Shaw and Wright make is:

> Ideally . . . scales should be reliable, valid and unidimensional, with equal units and a zero point. (Shaw & Wright 1967:561)

Unfortunately ideal scales are impossible in the cold harsh world of practical empirical research.

2.1.3. Scaling techniques

Four scaling techniques are commonly used in attitude studies: Thurstone, Likert and Guttman scales, and Osgood's semantic differential.[6] All share the characteristic of starting first with the careful formulation and wording of possible questions, progressing (ideally) by means of a pilot study with special pre-test judges to a selection of the most appropriate questions for the actual scale to be employed in the test, and finally administration of the actual test. The aim is to

---

[6] Other scaling techniques exist, of course. For example, graded dichotomies, scalogram analysis, scale discrimination, the unfolding technique and unfolded partial rank orders. See Shaw & Wright 1967:22–9.

include all and only those questions which have true relevance to the attitudes to be measured, and, most importantly, to cover the whole range of possible positions people may occupy on an attitude continuum, so showing not only that a person is or is not, for example, a racist or very religious, but also where that person is in relation to others on the same attitude continuum. The scale in effect serves as a measure of the strength or weakness of an attitude.

2.1.3.1. Thurstone scales

Thurstone scales typically focus on people's reactions to written statements so they are perhaps not immediately pertinent to the present socio-phonetic study. However, the attempt to construct an interval scale of measurement with '**equal-appearing intervals**' (Thurstone 1928; Thurstone & Chave 1929; Webb 1955) has had wide repercussions on all studies of attitudes. Basically a set of fifty or more judges are used in a pre-test run to place a set of around one hundred cards with written statements on them about the topic under investigation into a limited set of piles according to their message content. From these piles 20–30 statements are chosen to form the final test battery to be used with the real informants. The respondents either accept or reject the opinions expressed on the cards, or register non-committal (neutral), an individual's score being the mean score of all the items with which they agree.

Thurstone scales have been widely criticised for various inadequacies, particularly for the elaborate co-operation required of the pre-test judges, for the number of judges used, for the absence or presence of a middle point on the scale, and for the fact that the supposedly objective evaluations of the pre-test judges may not reflect the range of opinions, beliefs or whatever of the eventual respondents (they may not form part of the same population). Thurstone acknowledged the latter problem himself (Thurstone & Chave 1929:92). However, a body of evidence has accumulated which suggests that as long as pre-test judges are neither ego involved nor extreme (leading to a wider attitude range than the item range: see e.g. Upshaw 1962), then the assumption that the values that they award are indeed impartial and universal is safe (Triandis 1971:41).

Some critics are not so charitable. Eiser (1980:78), for example, warns that while Thurstone scales may be used 'as a measure of *relative* differences along a specified attitude continuum,' it has been 'completely established' that the judges' ratings are affected by their own attitudes; Kline (1993:75) quite simply refuses for practical and theoretical reasons to countenance the use of Thurstone scales; and Eiser and van der Pligt (1984) invoke accentuation theory and place an emphasis on context:

> Judgments are relative, and depend on context. . . . Even if we accept the assumption of an underlying continuum of favorability–unfavorability, we cannot know the absolute position of any item or person on that continuum. Indeed, we have to say that there are no absolute positions, only relative differences between different items or people, depending on the context. This invites the inference that we cannot measure attitudes absolutely, but only comparatively and in context. If we accept this, can attitudes ever have any meaning, except within a context? (Eiser & van der Pligt 1984:60).

While not denying the undoubted importance of context, it is extremely difficult to cater for context in attitude research design. Nevertheless, serious attempts have been made to do so in some studies of attitudes towards language (see below, Ch. 4).

The main lasting influence of Thurstone scales has perhaps been the misconception that equal-appearing interval scales permit the conversion of scale labels into numbers, the re-classification of nominal/ordinal scales as interval scales, and the subsequent use of parametric statistics. This regrettable influence is to be found in most studies of attitudes towards language.

2.1.3.2. Guttman scales

The goal of a Guttman scale is to establish a regular, cumulative and interlocking pattern of statements which forms a hierarchic set where agreement with one statement necessarily entails agreement with all statements 'below' it. First investigators propose a tentative pool of questions of the Yes/No (Agree/Disagree) dichotomous type, hoping the answers will betray a respondent's attitude to the particular item under scrutiny. A pilot test is then carried out and a scalogram analysis performed on the results in order to establish the ultimate battery of questions for the actual attitude study.

This procedure is perhaps best understood by reference to a simple arithmetical example: if children can calculate the sum (29 x 34) - 13, then they can certainly calculate the answer to 63 + 31, and also of course to 2 + 2. Barring slips and errors on the part of respondents, there is an order of difficulty for these three sums (though admittedly not an interval order), and a matrix of right answers plotted against problems tried gives a straightforward triangle:

✓ ✓ ✓
✓ ✓
✓

The same sort of interlocking cumulative order is aimed at in the formation of a Guttman scale for attitudes. If a question does not lead to answers which fit into the hierarchical pattern, then that question obviously does not focus on the same universe of content (i.e. the same attitude) as the other questions in the set: the set is not yet unidimensional. This uni-dimensionality does not (at least in theory) depend on the personal subjective views or opinions of the pre-test trial judges: it depends on the inevitable hierarchical pattern that the answers form regardless of who is answering the questions (again with the proviso that the respondent makes no 'errors' because of a lapse in attention or whatever). Once the badly worded, badly focused or irrelevant questions have been weeded out (i.e. all those questions which lead to a scatter of responses), a final selection of questions is made and the test given to the authentic population (or a sample thereof) under investigation.

Inevitably the final test may not give the ideal triangular pattern of responses expected, and the investigator then has to decide whether the measuring instrument is at fault, i.e. the set of questions, or whether the respondents do not form a representative unbiased sample of the intended population, or something else.

Sometimes investigators merely content themselves with a quasi-triangular pattern which looks as if it more or less fits the expected symmetrical shape. However, the computation of a coefficient of reproducibility is an easy matter (Shaw & Wright 1967:25) and a Rep of .90 or better is usually set as a minimum guarantee for the uni-dimensionality and scalability of the items making up the instrument.

Be that as it may, it is significant that in theory at least, both the questions in the set and the respondents in the final test are scaled. The sum total of a respondent's Yes's (or Agreements) reveals the intensity of the person's attitude, placing the respondent somewhere on the attitude continuum of, for example, Strong Agreement–Strong Disagreement: a score of 14 shows that Person P agrees with everything that Person Q with a score of 13 agrees with but goes one step (and only one step) further. The difference in attitude between Persons P and Q is specifically the acceptance or rejection of question 14.

It all seems very neat and attractive, but the uni-dimensionality of the set of questions is the key issue. If the triangular pattern of responses fails to appear, then the choice of questions is wrong. If, on the other hand, the triangular pattern is spot on, this may be because the questions are too widely focused and consequently of no interest. Kline (1993:72–5) clearly has no sympathy for Guttman scales.

Guttman scales are rarely if ever used in studies of attitudes towards foreign accents, even though teachers of L2 pronunciation are well aware that there is a limited hierarchy for the acquisition of particular L2 articulations (e.g. Finnish speakers use a dental [t̪] in Finnish, and unless they correct this to an alveolar [t] when speaking English they will not be able to produce a nasally released plosive as in [ˈbʌtˡn̩]; similarly the acquisition of the alveolar and palato-alveolar fortis sibilants seems to be a prerequisite for the production of [tʃ] and [ts]). Guttman-like scales could be useful to discover whether there is an order of like/dislike for various features of mispronunciation in foreign accents. It's possible that if someone dislikes absence of aspiration, they *necessarily* (or at least very probably) also dislike, for example, confusion of affricates, *and* sibilants, *and* labiodental fricatives. If such an order were found (for a particular population at a particular point in time), the implication would be (as for all such scales) that 'there has been sufficient uniformity of experience for the population of individuals so that the attributes mean essentially the same thing to the different individuals' (Guttman 1944: reprinted in Summers 1977:185). However, it could perhaps be argued that the uniformity is not a result of the socio-cultural homogeneity of the particular population, but the result of some universal linguistic precept beyond both an individual's and a society's control.

2.1.3.3. Likert scales

Thurstone scales are incredibly difficult to construct, and consequently many people resort to Likert scales. But Likert scales are incredibly difficult to interpret. Swings and roundabouts.

Briefly, as with Thurstone scales, respondents are asked to indicate their degree of agreement or disagreement with a set of statements on 5-point scales, and points are awarded accordingly (Likert 1932; Busch 1993; Turner 1993). The sum (or the average of the sum) of the scores for *all* the categories is taken as an indication of the person's attitude, especially the intensity of that attitude (hence the name '**a summated rating scale**'). The scale thus assumes that *each* item has a monotonic trace line, that the trace line for the sum of *all* the items is monotonic, and that the scale is sensitive to, and measures nothing but, the attribute in question (Kline 1993:75–6.)

Notice that there is no ranked order of questions as determined by a quasi-impartial group of pre-test judges, that each question provides in itself a rating scale, and that the sum of all the responses forms the Likert rating scale. Thus the responses to the separate questions of a Likert scale are treated as related and comparable interval scales even though they are in fact totally unrelated ordinal/ranked scales, and even though from a mathematical point of view they simply cannot be pooled and summed: the '2' awarded on one item is a name and has no mathematical relation to the '3' (again a name) awarded on another item. Despite this extremely serious flaw in the theory of Likert scales, the validity and reliability of Likert scales are generally thought to be ensured by carefully screening the items, by testing for the intercorrelation of items (i.e. for the presence of a mutual common factor, the attitude or whatever), and by checking the scale on a sample drawn from the target population. As Loewenthal says:

> Generally, rating scales may be treated as interval scales, unless you suspect the contrary. Likert (1932) suggested, when he developed his method of scale construction involving rating, that the intervals in such rating scales should be equal-appearing. (Loewenthal 1996:94)

Knowledge of a respondent's total score does not of course allow one to deduce how they performed on any one particular question: the responses elicited for a particular question, unlike those elicited on Guttman scales, are not 'reproducible.' Nevertheless, each question is as potentially interesting as the whole scale itself.

It is also noteworthy that each question is considered as being of approximately equal 'attitude value': the questions themselves do not form any hierarchical order or scale. In other words, any subset of the total battery of questions should theoretically yield the same attitude score as the total or any other subset: thus it is not the questions that are scaled, but the individuals responding to the questions.

The major drawback of a Likert scale is the utter impossibility of knowing how similar or dissimilar are the attitudes of two people with identical scores on the attitude scale. The final single score is inadequate and unrealistic as a representation of a person's attitude unless reference is also made to the person's strength of feeling, and, concomitantly, to their latitude of acceptance, rejection and non-commitment (see Diab 1967:141–3, Sherif & Sherif 1967:116–21). For example, two people may find themselves with the same attitude score and therefore find themselves in apparent agreement in their attitude towards the attitude

object under appraisal, and yet they may strongly disagree in their tolerance and/or non-committal towards the object and in their degree of involvement in it.

Another drawback is that the score obtained by one person on a Likert scale can only be interpreted by relating it to the distribution of scores on the *same* scale obtained by (many) *other* people drawn from the *same* given population. One person's score on its own means nothing, for there is no point of reference, no anchor point, no universal standard.

Despite the problems and drawbacks, most studies of attitudes towards language and accents use Likert-type scales, generally considering each question or category separately, and treating the scales as being interval. However, it is not always obvious that care has been taken in choosing the category labels, the number of response positions offered on the scale, and the inclusion or exclusion of a central neutral point, and in considering the strength and universality of responses obtained.

2.1.3.4. Osgood's Semantic Differential

Osgood's semantic differential, developed by Osgood and his colleagues, is closely connected with Likert scales, and is based on the theoretical premise that attitudes are multi-dimensional (Osgood et al. 1957). Informants have to indicate their attitude towards something or someone by reference to a pair of reciprocally antagonistic bipolar semantic adjectives placed at the ends of a 5 or 7-point scale, the mid-position being taken to be neutral or indifference (Figure 2.2). Depending on the experiment, the mid-position may of course be omitted, so forcing respondents to choose one way or the other.

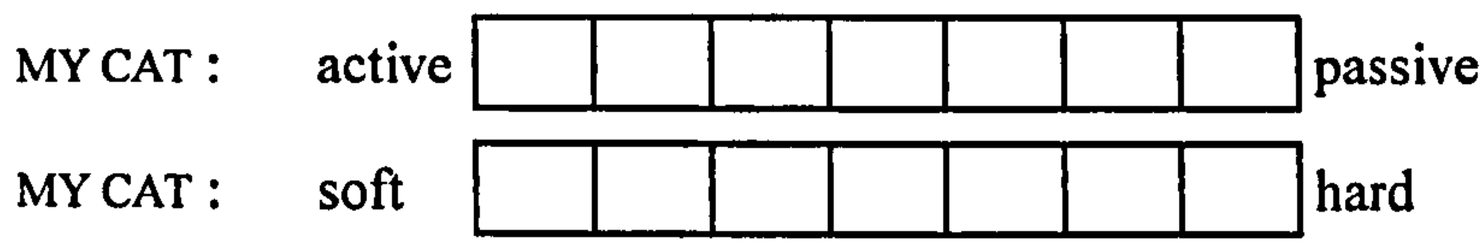


Figure 2.2

Each scale for each adjective pair can be seen as a straight 'cut' going through a 'semantic pincushion,' and the meaning given to a concept by someone is accordingly located on the line of the cut. The complexity and subtlety of the possible meaning of a concept is quickly realised when one contemplates the semantic pincushion formed by just three adjective pairs: using a five-point scale for each pair, then 125 discrete positions are established in the semantic space. A seven-point scale correspondingly yields 343 discrete positions. Ten adjective pairs with a seven-point scale would produce 282,475,249 positions!

Osgood's scale reveals how people think things differ: the 'differential.' Various factor analytic studies have shown that regardless of both the cultural and language background of the evaluator (Osgood 1964, 1965), all pairs of adjectives used in such tests collapse into three main 'factors' which broadly reflect the semantic dimensions of **evaluation** (good–bad, nice–awful, etc.), **potency** (strong–weak, big–little, etc.) and **activity** (active–passive, fast–slow, young–old, etc.)—the so-called 'EPA dimensions.' Thus, if we know how someone rates something on these three factors, then we can be reasonably confident

as to how they rate the same thing on different bipolar semantic scales. Typically, therefore, studies using this technique take three or four bipolar adjective pairs to assess each of these three factors. By awarding 1 to 5 (or 1 to 7) points for the responses to each pair of adjectives in each set, and then by averaging the scores for each set, the position of the thing measured can be represented by a model drawn in a three-dimensional space (see Figure 2.3).[7] Notice that this technique runs counter to all ideas of an ordinal scale: the scale is not an interval scale and yet the analysis treats it as such, the justification being the oft-quoted principle of 'equal-appearing intervals' which is basic to Thurstone scales.

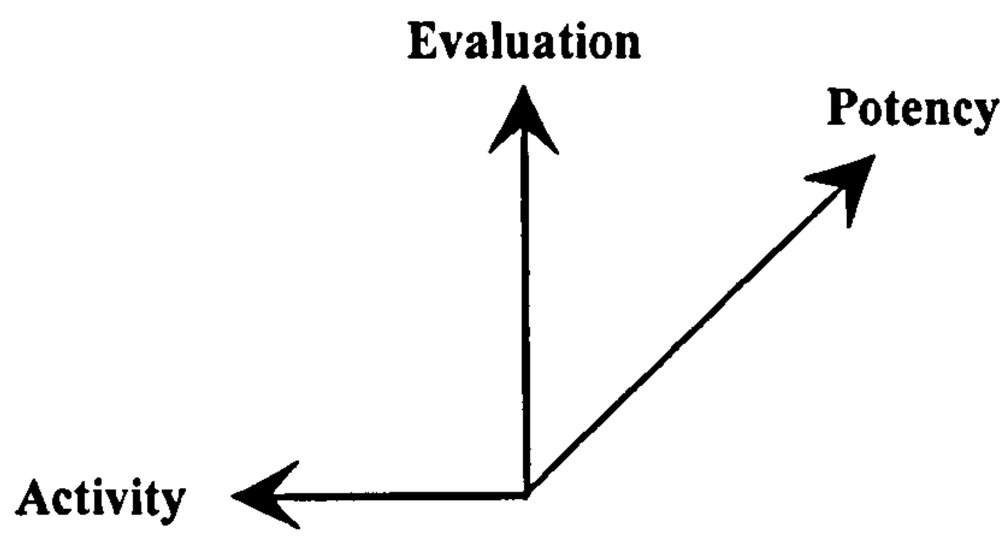


Figure 2.3 Model of the EPA dimensions

Different objects and concepts yield different three-dimensional models, and consequently, by comparing the models, one can see the difference in the attitude of the evaluator towards the objects and concepts. If the experiment is repeated at a later date, one can even evaluate changes in attitude by the changes observed in the location of the various points in the three-dimensional space. The technique is useful for either individuals or groups of people, the values for each pair of adjectives being averaged for groups of people.

The EPA dimensions are thought to account regularly for at least two-thirds of common variance in the evaluation of any particular concept, and are said to be easily distinguishable across a wide variety of languages, societies and cultures. However, it is possible that some concepts, objects or whatever being evaluated are better described by only two of these factors, or by some other factor or factors (for example, 'understandability': see Nunnally 1961). Naturally, if more than three factors are used, the resulting multi-dimensional model becomes much more difficult to visualise and interpret.

One fundamental problem in constructing a semantic differential, and in determining the factors which are pertinent in any particular experiment, is the fact that adjective pairs that are apparently irrelevant to the concept being judged may turn out, unexpectedly, to be significant. Attitudes are highly complex mental constructs and it is not always possible to foresee which adjective pairs will throw up the consistent systematic variations which betray relevance. Consequently, most standard textbooks seem to advise experimenters to keep to the fifty adjective pairs suggested by Osgood, and to his three EPA factors, unless

[7] See diagrams in Dawes (1972:100–1). Also Osgood's (1965) diagram presented in Fishbein (1967:109).

common sense and pre-test trials show the need to increase or alter the adjective pairs and to start hypothesising about the existence of another factor. One possible diagnostic pre-test trial would be to ask judges to describe in their own words the particular concept or whatever to be evaluated, and then to choose or construct adjective pairs for the final test scale based on the most frequently used adjectives in these descriptions. Notice that a high proportion of midpoint ratings on an adjective pair in the final test would seem to indicate the irrelevance of the pair to the concept being measured, and hence the pair should probably have been discarded during the test construction stage: nothing would have been lost.

Bem (1967) suggests that the fact that attitudes reduce to three broad factors (even across different cultures) is

> consistent with the view that an individual is unable to make more than a small number of independent discriminations among stimuli that have never been publicly available to a socializing community. (Bem 1967:185)

People's attitudes follow the broad patterns set down and inculcated by the society in which they live, and even when describing their *own* attitudes, people use the cues potentially available to an outside observer and commentator.

> Self-descriptive attitude statements [are] based on the individual's observations of his own overt behaviour and the external stimulus conditions under which it occurs. (Bem 1967:185)

Bem sees us as the product of the society in which we live, with little autonomy to form our attitudes in our own terms, a depressing view perhaps, and one which contemporary existentialists would probably not have subscribed to. However, it is a view which may have importance for cross-language studies of attitudes towards accent: to what extent are attitudes towards foreign-accented speech universal? to what extent are they moulded by the L1 of the listener? to what extent are they shaped by the stereotype of the speaker's identifiable and/or assumed nationality? More specifically, are attitudes towards Finnish-accented English always the same regardless of the L1 of the listener?

Although the tripartite EPA dimensions are a direct result of factor analyses, and not some arbitrary concoction pulled out of thin air, they have been criticised for not corresponding to the dimensions of normal spontaneous involuntary reactions (as opposed to conscious mental constructions) in person perception and speaker evaluation processes (e.g. Brown, Strong & Rencher 1973:33, Brown 1980:296–7). In normal communication situations, categorisation is considered to be *bi*-partite in nature, viz. intellectual (competence, strength, status, success, intelligence, decisiveness, endeavour, etc.) and social (friendliness, helpfulness, honesty, warmth, etc.). (See Bierhof 1989:14; quoted by Valo 1994:55.) Typically strong positive values are achieved on only one of these two dimensions, not on both, suggesting that a competent respected prototype (someone one can respect) is offset by a pleasant liked prototype (someone one likes).

In studies of attitudes towards accents, the bi-partite pattern receives support from Brown's competence v. benevolence scale (Brown 1980; Brown et al. 1985), and from Street's competence v. social attractiveness scale (Street et al. 1983; Street 1984, 1985; Street et al. 1984). Many studies of the perception of

speakers based on accent in fact place their adjective scales for personality traits in two or more groups: status, competence, solidarity, benevolence, etc.

Osgood's Semantic Differential is condemned by supporters of Kelly's (1955) **repertory grids** (see Fransella & Bannister 1977:112) for it imposes a set of pre-determined dimensions on the evaluators rather than seeking to discover the dimensions that they (the evaluators) would readily use when confronted with the attitude object. To fill in a grid Kelly's original triadic elicitation method required interviewees to consider a whole series of three-element sets (admittedly chosen by the interviewer) and to place together two out of the three elements in each set, thus simultaneously acknowledging similarities and dissimilarities within each set. Then—and this is the important bit where non-predetermined differentiating constructs come into play—interviewees had to label the characteristic which united the pair and also name the characteristic of the remaining lone element. Originally designed in the context of psychotherapy with the aim of allowing people to make sense of their own life problems, grids, according to Fransella and Bannister (1977:112), offer us the chance to recognise 'a person's own superordinate axes as well as his particular judgements.' The lesson for investigators of attitudes towards accents to learn here is to hold pre-test experiments in which people are asked to describe accented speakers, and to construct bi-polar adjective pair scales only from those adjectives most regularly used (Giles & Bourhis 1973).

There is still one important question to be asked concerning the statistical tests and procedures which are appropriate for use in conjunction with Osgood's Semantic Differential (and indeed with Likert-type scales): can parametric statistical operations be carried out on the data obtained?

In order to perform parametric statistical tests, one assumes that the scores of the dependent variable(s) are measured on an interval scale, and that the distribution of scores is normal. Now, the bi-polar adjective scales of a semantic differential are certainly *ordinal*, but it is a brave person who claims that these scales are rated by respondents using some sort of *interval* scale. Moreover, it is rash to assume without question that the same fundamental interval scale is used for different adjective pairs. In other words, one cannot legitimately combine or sum scores from different adjective pairs; nor can one legitimately sum or average the scores from several adjective pairs to form a factor total. Nevertheless, while the scale values themselves are essentially ordinal, it is of course the *means* of the values across a population (whether they are legitimately procured or not) that are analysed, and those means *are* interval values, and thus legitimately available to parametric operations. At least this is what some sociolinguists argue (e.g. Shaw & Wright 1967:564–5; Fasold 1984:177).

Most studies (see Lemon 1973:105; Heise 1977:241–2) would seem (perhaps surprisingly) to indicate that worries about the interval or non-interval status of the scale are not confirmed by experimental evidence: it is suggested either that departures from a strict interval scale would seem to be so small as to be unimportant (i.e. we should forget about the problem), or that differential rating or multiple regression formulas could be used (i.e. if we really are worried, we can

do something). And yet the doubt lingers—at least in my mind—that while respondents may intuitively scale each adjective pair on a subjective interval scale, and hence indicate the direction of their response as concerns that particular adjective pair, there is nothing to confirm that the extremity or range of responses is the same for different adjective pairs, or that different respondents use the same scales, extremities and ranges for the same adjective pair, variation probably reflecting not just age, gender and social (or socio-economic) class but also unpredictable idiosyncratic features.

There is therefore a problem of comparability of responses (Diab 1967). Does the neutral (centre) position indicate 'equally good or bad' ('strong or weak,' etc.), or does it signify lack of feeling and absence of attitude? Might the ticking of the neutral (centre) position denote the recognition of the most desirable (politically correct) middle-of-the-road stand and the masking of a covert conscious or unconscious attitude, or is the informant saying that for them the scale is totally inappropriate for the stimulus to be evaluated? Does the ticking of one extreme end of an adjective pair scale indicate the same strength of attitude for all respondents in addition to the same relative position of attitude? Do identical cumulative totals indicate or guarantee identical attitudes?

Despite certain well-founded reservations concerning both the scaling of a semantic differential and the parametric statistical procedures generally employed, this attitude measuring technique is widely used and respected (see Lemon 1973:109–10; Heise 1977:246–51). Its validity and reliability is not generally held to be any more questionable than other attitude measuring techniques. Most studies of attitudes towards language and accents use a combination of Likert scales and semantic differentials, broadly labelling the scales as 'Likert-type scales.'

2.1.4. Problems in measuring attitudes

All of the measuring techniques presented in the previous section are strong in measuring direction and intensity of attitudes, but they all fail to tap other aspects of attitudes, in particular specificity, multiplicity, interconnectedness, centrality, and consistency (Shaw & Wright 1967:567–8). This failing of the scales and techniques used for measuring attitudes has still not been addressed and remedied. Attention is paid to reliability, validity, uni-dimensionality, equal units and a zero point, but to little more (*ibid.* Pp. 20–1, 561). One is obliged to ask (heretically?) whether the measurement of attitudes is really possible, some investigators—including myself—saying yes, and others—who refuse to tackle the problems—saying no. Certainly a number of highly critical and sceptical studies have appeared over the years.

Shaw and Wright (1967)—a relatively early study—present an important selection and critique of attitude measuring instruments. Their general conclusion is not very reassuring: researchers do not use the best scales available, fail to construct their own scales using the best methods, and conspicuously neglect to test the reliability and validity of their own scales.

Saris (1993)—a more recent study—is even more scathing. Although using Fishbein and Ajzen's (1975) definition of attitude as his starting point ('Most investigators would probably agree that attitude can be described as a learned predisposition to respond in a consistently favourable or unfavourable manner with respect to a given object'), he considers the alleged 'agreement' questionable, claiming that 'it is . . . doubtful whether there is still hope for attitude measurement' (Saris 1993:190). In particular he wonders whether attitudes can be subjected to factor analysis and whether attitudes can be analysed as latent factors: the idea that attitudes are predispositions is widely held, and factors can be drawn to explain correlations between various responses, but this does not prove 'the possible existence of general attitudes,' nor 'the necessity of attitudes' (*ibid.* P. 192).

Moreover, because of the imprecision concerning what constitutes an attitude or attitude object, Saris dismisses the idea that attitudes can be measured by summing the scores of a number of different items assumed (erroneously, in his opinion) to belong to the same attitude domain, or by summing the evaluations of different aspects of the same attitude object. Reformulations of questions about an attitude object are useless as the new questions will inevitably have a different shade of meaning.

Saris even rejects the basic presupposition of most research, viz. that informants express their opinions in the same way when they answer questions: the response function (the relationship between opinion and response) is not the same for all respondents. The problem is that respondents have no objective standard to set their response against: while people can probably gauge lengths, distances, weights and temperatures with a reasonable degree of inter-judge consistency, this is not true for attitude objects. One respondent may have and use a wide range of responses, while another respondent may have an 'all or nothing' type of response set with a bunching of responses at one end or other of the scale. The latter is the result of a failure (on the part of the interviewer's instructions) to establish a clear connection between the interviewee's subjective scale and the questionnaire's response scale. The problem of variation in response functions needs to be addressed in the data collection stage, not later.

Saris also points out that respondents are not equally sensitive to the same attitude object at different positions on the object scale. For example, a mild case of environmental pollution may provoke an extreme reaction from a 'green' activist but no reaction from someone else. With a major case of pollution the green's reaction is equally strong, but now the 'non-green' person's attention is caught and they too now show a reaction. The problem (according to Saris) is to distinguish between the absence of a clear connection between response and scale, and the absence of any response at all.

Saris's conclusion is: 'We very much doubt whether the ongoing attitude research will produce meaningful results.' (Saris 1993:202)

One final stumbling block to the study and measurement of attitudes—whatever scale is used—is experimenter effect: how can we be sure that we (the investigators) are not producing and influencing the attitudes we observe in our

subjects? We need to be aware of the biosocial (e.g. age, gender and race), psychosocial (e.g. mood, status, warmth) and situational (e.g. experience, venue and previous acquaintance with subjects) effects that we ourselves produce (Ambady & Rosenthal 1995).

A healthy scepticism is perhaps called for in reading all attitude research papers. And naturally care and attention needs to be given to the construction and execution of our own experiments, followed by humility and hesitation in the evaluation and interpretation of our data.

2.2. Survey methods

If sociolinguistic surveys are to be of any value and not degenerate into anecdotal pieces of information of questionable interest, then certain procedures and assumptions must be accepted and adhered to. In particular the objectives of the survey, the research design, the population of interest, the sampling procedure, the questionnaire design, the practical data collecting techniques, the data and statistical analyses, all need to be carefully considered. All too often sociolinguists would seem to ignore these basic points.

2.2.1. Research design

A research design is the basic plan or strategy of a research project, the logic behind it, an explanation of the sampling procedures, descriptions of the measurement instruments, of the variables to be measured, of the comparisons to be drawn (Oppenheim 1992:6). In particular the design presents clear research questions, and, equally important, the methods for answering those questions in a way which allows valid inferences and conclusions to be drawn. This is basic.

However, a problem for attitude research would seem to be that of keeping in mind the distinction between a descriptive survey and an analytic survey.

A descriptive purely fact-finding enumerative census-type survey is perhaps the easiest type of survey to perform, the fundamental (risky) premise being that respondents' attitudes are non-ephemeral and non-fickle, and the aim being first to 'count' the attitudes of a presumed representative sample and then to draw inferences concerning the population as a whole. The results would give proportions of attitudes for a particular sample and, if sampling errors have been avoided, then generalisations for the population. Associations between attitudes and certain features (age, gender, etc.) could also be pin-pointed, but naturally such associations would offer no proof of causality: causal relationships between one variable and another could not be established. A descriptive study attempts to answer the question 'How many?' and not to tackle the question 'Why?'

An analytic study of attitudes is much more difficult to carry out, some might say impossible, attempting to find the causes for the attitudes observed. A multicausal model, with potentially highly complex causal pathways (or causal careers), is required: the causes may be independent or interrelated, and if interrelated, the causes may act simultaneously, or successively (perhaps in a particular defined sequence), the effects of the causes thus being singular or cumulative. It might be easy to demonstrate associations or correlations between various variables (as in a descriptive study), but the attribution of causes for attitudes

would be rash in the extreme. As Oppenheim (1992:18) points out, our 'notions of causality . . . imply a linear, additive progression, a kind of geological or "layered" causality in which bounded one-way influences are brought to bear on people at successive times causing them to change.' But such a model is not appropriate for attitudes where 'spiral reinforcement' processes are probably the rule (and probably without exception): influencing formative factors for attitudes are probably 'incremental but not additive, influential but also interactive.' We cannot as yet reveal the multitudinous myriad causes for an attitude, nor predict an attitude.

Studies of attitudes are thus forced towards descriptive research designs, though they may of course try to highlight patterns of association between particular attitudes and the age, gender, social class, etc. of the informants.

According to Oppenheim (1992:26–9) a factorial design and multivariate analyses (e.g. regression analysis, analysis of covariance and analysis of variance [though analysis of variance is chiefly applicable to quantitative variables, not categorical variables], factor analysis [to try and discover the main underlying determinants for the data], cluster analysis and discriminant function analysis [both to develop a typology, a way of classifying subjects into a few distinctive sub-groups]) would 'help to disaggregate the variance (the fluctuations or differences) in the dependent variable according to the relative importance of the independent variables which we have entered into the design.' Unfortunately, multivariate techniques generally exploit statistical procedures using interval-type variables, and such variables are rare in studies of attitudes. The more appropriate general loglinear analysis, used to study association patterns among categorical variables, is rarely used (an exception being Rahkonen & Juurakko 1998).

The temptation for sociolinguists studying attitudes would appear to be to start a study with a basically descriptive design and yielding nominal or perhaps ordinal data, and then blithely to slip into an analytic framework, abusing inappropriate statistical procedures and drawing unstable, even groundless, conclusions. Generally speaking, the safest conclusions one can draw concern distributions (or proportions) of attitudes (e.g. for various accents or for features of mispronunciation) found in a particular limited sample, indications of the significance (statistical probability) of these distributions, and patterns of association for the attitudes. Naturally, one could offer no proof of causality.

The two experiments presented in this present study are descriptive.

2.2.2. Questionnaires

The writing of questionnaires is a far more complicated task than most people would seem to realise. The wording of the questions (the avoidance of bias and ambiguity, of double-barrelled and double-negative phrasing, of hypothetical situations, of presuppositions, of emotionally loaded vocabulary), the types of question (closed or open, alternatives or multiple choice, rank ordering, etc.), the format, layout and length of the questionnaire, the order of questions, the sequence ('funnelling') within groups of questions, directions for possible

'skipping' of questions ('filters'), coding of answers, the demands made on the informants (in terms of e.g. memory and time), the instructions, everything has to be considered extremely carefully. Much has been written on this subject; much, I suspect, has been left unread by sociolinguists. The usual approach to the writing of questionnaires, according to Luppescu and Day (1990:131), remains the 'shot-gun approach: 'shoot at a wide array of targets and hope that you hit something.' People fail to validate the measuring instrument they design, to check that the instrument measures what they intend it to measure.

The writing of questionnaires is an extremely important topic underpinning the present study. However, I do not propose to deal with it here at all.[8] The questionnaires used in the two experiments of this present study are described in Chapter 6 (6.1.1.3 and 6.1.2.3) and presented in Appendices A–D.

2.2.3. Populations and samples

Every survey needs to be directed at one particular group of people, i.e. **a population**. However, it is unusual to be able to collect data from all members of that population: such a census is usually impossible. As a result **a sample** has to be taken from the designated population, and this sample needs to be picked in such a way that it is representative (or as representative as possible), for otherwise no laws of probability can be applied, no statistical inferences can be drawn and hence no results can be generalised to the whole population. An unbiased sample is essential for the validity of a survey's results (a fact generally understood by most people when they consider political opinion polls, but often ignored by linguists when it comes to linguistic surveys, particularly those linguists investigating attitudes towards accents).

Samples are typically of two kinds: probability and non-probability samples.

Probability samples (often misleadingly called random samples: they are far from being haphazard, accidental and slipshod) ensure that everyone in the population has an equal chance of being included in the sample. This is done by using either simple random sampling or stratified random sampling. The former method (simple random sampling) is self-explanatory, the cliché technique of drawing names from a hat not in fact being very random and the preferred technique nowadays being computer electronic sampling. The latter method (stratified random sampling) involves 'blocking,' classifying the target population according to certain pertinent strata (e.g. age, gender, mother tongue) and then randomly—but proportionally—sampling each of the strata. The aim is to ensure that no important part of the population is omitted. The essential feature of both methods is the element of equal chance in the selection of respondents: the target population must be clearly defined; all its members must be known and listed; and all and only those people who are selected from the list by chance in some systematic fashion are included. Without chance, the tools of statistics—basically the science of generalising from samples to populations—are unusable, and

---

[8] Essential reading for the writing of questionnaires is Oppenheim (1992). See also Kerlinger (1973 Chs. 24 & 28); Weisberg & Bowen (1977 Ch. 3); Youngman (1984); Alreck & Settle (1995).

hence the results are purely anecdotal. Few studies of attitudes would seem to use a probability sample of either kind (though Hultfors 1986 and 1987 comes close: see below P. 40. Cf. the voice quality survey of Esling 1978). Just prior to tearing Labov (1966), Shuy et al. (1968) and Wolfram (1969) to shreds, Davis (1986) says the following, though I believe her comments could be extended to the wider sociolinguistic field:

> No major dialectological study to date has used [a random sample] in the strictest sense of the word. (Davis 1986:43)

The corollary to the requirement that a probability sample be taken from among the informant population is of course the requirement that a probability sample be used for the data under investigation. This is a point forcefully driven home by Clark (1973) in a biting condemnation of the disdainful disregard of linguists for the essential statistical link between randomness and generality:

> If the investigator is to treat language as a random effect, then he must draw a sample at random from the language population he wishes to generalise to. (Clark 1973:350)

As far as I can tell, no sociolinguistic study has ever worried about this matter at all. In the present study I myself make no attempt first to collect a body of Finnish-accented speech and mispronunciations, and then to select a probability sample from it for my empirical experiments.

Non-probability samples are typically 'quota samples': the population is said to have certain characteristics, and candidates for inclusion as respondents are accepted on the sole basis of having the required characteristics. The stringent requirement of random selection of respondents is waived and respondents are 'picked' on a more or less 'first come, first served' basis. Hence the likelihood of any one person being included in the sample is completely unknown: it is not '1 over the total population size.' Therefore, sampling error (i.e. the margin of error), which 'reflects the discrepancy between the results one obtains from a particular sample and the results one would have obtained from the entire population' (Schwarz & Wellens 1996:644), cannot be calculated, and hence, crucially, this seriously limits the extent to which one may generalise the results from the sample to the total population. Many political opinion polls and market research surveys use this basic technique of non-probability sampling but generally forget to point out this grave limitation.

Some so-called surveys—and many sociolinguistic surveys fall into this category—fail to define the population under study and leave the problem of sampling unresolved. Such studies use **convenience samples**, i.e. **opportunity samples**, accepting as respondents anyone at all who happens to be around in the right place at the right time. Who or what these informants represent is never elucidated. Nor can it be. Their representativeness is a complete mystery, a pure figment. Such studies can tell us little of significance about any specified population (not that the population is specified anyway) unless the sample size is extremely large, i.e. is in the thousands (see below).

However, convenience samples are useful in some experimental research (such as medical research) where the presence or absence of some control factor

is being tested on two comparable groups: as long as the assignment of participants to the two groups is random, and as long as the data consists of interval data, then some interesting—but strictly limited—conclusions may be drawn. The two groups are *assumed* to be representative of populations with a normal distribution and are compared with each other with, for example, a between-samples test or t-test or with one-way analysis of variance. Most sociolinguistic studies are carried out within this framework. Or rather the claim is that they fit within this framework.

Where a survey is obliged to use a convenience sample (as indeed I am), and tests not two but only one single variable (e.g. a mispronounced phoneme, as in my Experiment 1), then the test is a test of distribution, **a goodness-of-fit test**: does the observed distribution of responses confirm our expectations, e.g. 50/50? In such cases, it is essential that there be no destructive bias or injurious system in the 'convenient' selection of informants, no systematic exclusion from the informant pool. Unless scrupulous attention is paid to this point, then results must be dismissed as anecdotal.

One sociolinguist who does take care with sampling is Hultfors (1986, 1987). On the one hand concerned both with randomness and with representativeness, and on the other hand wanting to include as many different types of people as possible, to make his informant groups as heterogeneous as possible, and thus (somewhat ambitiously) to have as representative a sample of the total English population as possible, he decided to exploit a modified form of stratified random sampling. He found informants from: comprehensive schools, adult education centres, day centres, athletic associations, lunch clubs and residential homes for the elderly:

> Each age group of informants makes up a predetermined proportion of the entire corpus, . . . [and] each of the institutions referred to, located in the research area, had an equal chance of being chosen, independently of the other choices that were being made. (Hultfors 1986:209)

This is truly impressive. Few sociolinguists would seem to aspire to such a sampling procedure—probably because of the horrendous difficulties they would encounter in trying to locate *all* the clubs, schools, associations, homes, etc. in a specific area, in obtaining in advance their willingness in principle to participate, and *then* in choosing one or two by chance, and so on.

Three basic warnings about sampling are regularly given by books on statistics:

1) Most statistical methods depend on the independence and lack of bias of a simple random sample, for otherwise the results must be 'modified' (using computer algorithms and highly sophisticated techniques comprehensible to a limited few).
2) No matter how results are modified, without a randomised design there can be no dependable statistical analysis, i.e. random sampling remains the *sine qua non* for the statistical accuracy of a survey.
3) A convenience sample, using the first *n* population units that come along, is especially prone to distortion and unreliability.

Unfortunately, few sociolinguistic/sociophonetic studies seem to heed these warnings or to address the problems. While I myself may recognise the dangers, there is little I can do to minimise them apart from aiming for the highest number of informants possible. The theoretical justification for my statistical analyses, for the tests I use, for the generalisation of the interpretation of my results, hangs on the unverifiable representativeness of my convenience samples. However, we do not live in an ideal world: practical constraints prevent us from adhering to theoretical niceties. Nevertheless, some statisticians view the statistical methods of linguists 'as just about as scientifically-grounded as alchemy' (Davis 1986: 47).

The drawbacks of convenience size are offset to a certain extent by sample size, for large sample size can (1) reduce sample bias and sampling error, and (2) improve **statistical power**, the probability of wrongly rejecting or accepting the null hypothesis (see Cohen 1992, and Hatch & Lazaraton 1991:239). The first point is recognised by everyone, but the interplay of all the factors involved here is still poorly understood. As Cohen explains:

> Statistical power analysis exploits the relationships among the four variables involved in statistical inference: sample size *(N)*, significance criterion *(α)*, population effect size *(ES)*, and statistical power. For any statistical model, these relationships are such that each is a function of the other three. . . . For research planning, . . . it is most useful to determine the *N* necessary to have a specified power for given α and *ES*. . . . The statistical power of a significance test is the long-term probability, given the population *ES*, *α* and *N*, of rejecting H_0. (Cohen 1992:156)

In effect Cohen is arguing that for a minimum level of statistical power one has to have a minimum sample size. Thus for a power specification of 0.80, and where a medium effect size is expected, the minimum sample sizes for basic research design are given in Table 2.1. As Cohen says, somewhat mordantly, few studies seem to have taken heed so far of these facts.

Table 2.1 Minimum sample sizes *(N)* for basic research design

	$\alpha < .05$	$\alpha < .01$
mean difference	64	95
correlation	85	125
χ^2 with 1 d.f.	87	130
Anova with 3 groups	52	76 per group
Anova with 4 groups	45	63 per group

(From Cohen 1992:158. And reported by Loewenthal 1996:35–6)

Munn and Drever (1990:15) address the same issue for non-parametric statistics and dichotomous data: if one uses the 95% confidence level of significance ($p < .05$) with dichotomous data, then the estimated error margins for specific sample sizes are seen in Table 2.2, P. 42. In other words if, for example, 42% of informants in a sample of 250 said 'Yes' rather than 'No,' there is an estimated error of ± 6.0%: the chances of results lying outside the already wide range of

36–48 are 1-in-20. To narrow the confidence range, many more informants would be needed, and few people have access to hundreds of informants. Many readers of Munn and Drever's book may perhaps wonder whether the use of the word 'small-scale' in its title (*Using Questionnaires in Small-Scale Research*) is entirely appropriate, or whether its use is ironic.

Table 2.2 Estimated error margins for specific sample sizes

Sample Size	95% confidence range
100	± 10.0%
250	± 6.0%
1000	± 3.0%
5000	± 1.4%

Suzanne Romaine is one of few sociolinguists who openly worry about sample size. She criticises Labov's (1966) claim that 'the number of informants and tokens needed to produce results representative of the linguistic behavior of a speech community is much smaller than that required for other forms of social behavior' (Romaine 1982:107). On what grounds does he claim this? She further criticises both Labov (1966) and Trudgill (1974b) because

> no prior thought had been given to the degree of precision (defined statistically in terms of standard error) which would be required for the results. . . . Many sociolinguists . . . have not considered the question of the optimal size sufficient to reveal the sociolinguistic structure of a speech community, but have accepted Labov's assumptions about the relative homogeneity of linguistic behavior. (Romaine 1982:108)

Random sampling also comes under criticism when applied to sociolinguistics:

> The terms 'random' and 'representative' are being used in sociolinguistic research without the technical precision attached to them in statistics by the social sciences. (Romaine 1980a:169)

> Provided the method of sampling is independent of the characteristic which is being sampled (which is the essential principle underlying the theory of random sampling), it does not matter how regular or systematic the procedure is by which the individual samples are selected. . . . [Yet] it has been said that linguistic expression is too much a matter of conscious choice for it to be amenable to pure chance consideration. (Romaine 1982:112)

The crucial point is that not all levels in the organisation of linguistic material are governed in the same way and to the same degree by laws of chance. Some characteristics are determined by inviolable linguistic constraints, some governed by both chance and choice. And the place of a given type of linguistic form on this cline determines the size of sample needed: language-conditioned factors require a smaller sample to reflect the total (statistical) population, while style-conditioned factors need a larger sample.

Sample size *is* a problem, but if the research design is robust then its importance diminishes. If the population under investigation is clearly defined, and if the sample chosen is representative, then the amount of data to be collected

becomes a matter of common sense. Once an acceptable level of sampling error is fixed, then the homogeneity or otherwise of data being collected will prompt either an end or a continuation to sampling. However, the decision to continue collecting more data must be tempered by the level of inconvenience acceptable to the researcher and the informants. It must also be remembered that most statistical tests are problematic with sample sizes less than 30, correlations being notably unstable for samples smaller than 50 to 100 (see Fitz-Gibbon & Morris 1987:116–7).

The population and the sample sizes were a major problem in the two experiments I report in the present study, but one feature of the research design of my second experiment makes a large number of informants inevitable there and distinguishes this study in an important way from most previous research in the field. Nearly all voice studies using the matched-guise technique present the two versions of a voice which are being contrasted and examined to the *same* evaluators, hoping that listeners do not cotton on to the fact that they are listening to the same voice in different guises. Various techniques and devices are naturally used to minimise the likelihood of listeners seeing through this deception (typically the versions are separated by other voices and are not presented consecutively), though how successful the artifice is is open to question. For my first experiment, contrasting mispronunciations, it is obvious that the same speaker is presenting the two articulated versions of the sentence: there is a straight contrast and no attempt is made to fool the listener. However, in my second experiment, reactions to various degrees of accentedness, I present the two versions being evaluated to two *different* samples of informants: no deception is involved, and my analysis of reactions hinges not on the reactions of the same people to different versions of the same voice, but on the reactions of two different sets of people which each hear a different version—and only one version—of the same voice. In other words, my second study is analogous to a medical study which involves two patient groups each undergoing a different treatment, and my analysis is one of how similar or how different the reactions of the two groups are. Inescapably this sort of 'between-subjects' design requires many more informants than a 'within-subject' design, and this in turn might explain why most sane people eschew this sort of research design.

2.2.4. Statistical tests and procedures

Statistical tests are regularly grouped into two exclusive categories: parametric and non-parametric. While people still argue about the necessity of employing only those parametric or non-parametric statistical tests devised for particular types of data and no other,[9] as a theoretical purist I feel more comfortable claiming that my data is amenable to only non-parametric tests and statistics since:

[9] A good summary of reasons for choosing nonparametric procedures instead of parametric ones is given by Hatch & Lazaraton (1991:237–9) though they are (I feel) on the liberal side in presenting the assumptions underlying nonparametric tests. Cf. the presentation of non-parametric tests in Meddis (1975) and in Bryman & Cramer (1997:117–8). The standard textbook for non-parametric statistics is Siegel and Castellan (1988).

1) the informant samples were not strictly random and were only assumed to be so;
2) the non-dichotomous scales were not anchored externally, the units of the scales being subjectively interpreted;
3) the measuring scale for the dependent variable(s) was not interval but nominal/ordinal (i.e. the data was frequency data);
4) X-bar and *s.d.* were inappropriate measures of central tendency and dispersion;
5) the total population could not confidently be claimed in advance to have a normal distribution.

Many sociolinguistic studies about attitudes towards language and accents (possibly most such studies) fail to acknowledge some or all of these points and proceed with parametric tests.

For example, many studies (the vast majority) assign numbers to the points on their Likert-type scales, and then treat the scales as interval (or even ratio) scales and use the mean as the statistic to indicate central tendency. The main assumption being made here is that the distances between any two numbers on the scale are always (and for everyone) of a known given size. This is clearly not the case with scales measuring attitudes, where the poles are marked 'strongly agree–strongly disagree,' 'excellent–awful,' or whatever, and where one person may have very strong opinions about the subject being investigated and another person may be totally indifferent towards the same subject. There is no between-subject consistency (sometimes not even within-subject consistency) as to what the points on the scale represent in terms of quantity or size. Moreover, attitudes do *not* have normal distributions. Consider any extremist view of politics, religion or sexual mores and this fact, I contend, is obvious. We are obliged to assume (until proved otherwise) that all attitudes, whether extremist or moderate, be uniform and non-normal in their distribution, that the attitudes expressed towards phenomenon P by an informant sample S will be evenly spread across a 10-point scale, 10% on the first point, 10% on the second point, etc. We cannot expect a bell-shaped curve for the distribution.

My two experiments were chosen with two types of data in mind: nominal and ordinal. Occasionally there is some overlap of the type of data encountered since some analyses require combinations of the two types.

2.2.4.1. Nominal data

The usual test employed to examine the frequency of cases found with a dichotomous nominal variable (as in my Experiment 1) is the Binomial test. The Chi-Square test (Pearson Chi-Square), a 'goodness-of-fit' test (does the *observed* number of objects falling into each category differ from the *expected* number based on the null hypothesis?), is also suitable for dichotomous nominal data and cross-tabulation/contingency tables.[10] Since the Binomial test can be used for small samples whereas Chi Square cannot, I shall use the Binomial test where appropriate but the Chi-Square test more generally, in preference to the Binomial

[10] See *inter alia* Welcowitz et al. (1971, Ch. 16) and Meddis (1975:57).

test, since it allows me to progress more readily into crosstabulations and the use of Cramér's *V*.

It should be noted that the Chi-Square test reveals nothing about the strength or form of the association between the variables. The higher the Chi-Square value, the more confident one can be that the data from another sample would give the same result, but this is not the same as saying that the association between the variables is stronger. This point is underlined by the fact that the test is sensitive to the value of *N* (in addition to the degrees of freedom): the higher the value of *N*, the higher the Chi-Square value. Thus it is difficult to compare results obtained from different samples of different sizes.

For examining the relationships between variables and for measuring the strength of the relationships in contingency tables, I shall use Cramér's *V*[11] since the value of *V* is reasonably transparent. When Cramér's $V = 0.00$, the distribution of responses for the contrasted variables are identical: there is no association. When the variables are completely dependent upon each other (e.g. if all women were to love bananas and all men to hate them), Cramér's $V = 1.00$: there is a perfect correlation. If all women loved bananas and men were evenly divided 50/50 in loving and hating them, then Cramér's *V* would be .50. Put simply, the higher the value of Cramér's *V*, the stronger the relationship between the two variables. Sample size naturally determines the level of significance for a particular value of Cramér's *V*, and the significance level is the value given by the Chi-Square crosstabulation. The limitations imposed on Cramér's *V* by sample size clearly upsets some social scientists (e.g. Bohrnstedt and Knoke, 1988:309, dismiss Cramér's *V* and prefer Lambda).

The test I shall use to indicate the proportional reduction in error for predicting a dependent response when knowing the values of another independent variable (e.g. gender and age) is Lambda (λ).[12] Since the lambda statistic is an asymmetric measure of association for non-orderable discrete measures (i.e. one must be able to recognise which variable is independent, and which dependent, because the result varies depending on which variable is being predicted), and since it is based on prediction from modes (i.e. the prediction of the mode of the dependent variable), it is well suited to the present study.

2.2.4.2. Ordinal data

The first question of my Experiment 2 (estimated age of speakers) yields what I shall take to be nominal/ordinal data, not interval data. This is because the nature of the measuring tool presented to informants is visibly nominal/ordinal, and especially because the perception of age (as opposed to statements of real age) is an equivocal 'more-or-less' estimation rather than a precise absolute measurement. Hence I shall use primarily mode and median (in preference to mean) as indicators of central tendency, and exploit the Mann-Whitney *U* test and the

---

[11] Siegel & Castellan (1988:226–32) refer to Cramér's *V* as 'the Cramér Coefficient *C*.'

[12] See Norušis & SPSS (1993:213, 223); Bohrnstedt & Knoke (1988:307–9); Hatch & Lazaraton (1991:511–12). Reynolds (1977:34–41) presents a daunting mathematical explanation of lambda.

Median test (rather than parametric tests) for the data analyses. However, despite my reservations about the suitability of the mean, I shall both present it and use it.

The second question (estimated excellence of pronunciation) clearly yields ordinal data and I shall exploit the Chi-Square, Mann-Whitney *U*, Median and Kolmogorov-Smirnov tests for the analyses.

The third question (the fourteen traits) uses 7-point ordinal scales, but this measuring instrument poses both a theoretical and a practical complication. Theoretically the central point is *not* part of the scale and does *not* fit the sequence, for it may variously signal 'undecided,' or 'unwilling to commit myself,' or 'the scale (the adjective pair) is inappropriate, and hence it is impossible for me to decide' (see end of section 2.1.3.). For practical purposes the middle point could be removed, the 'abstainers' being asked to indicate their abstention in a box to the side of a 6-point scale. However, despite the problem of interpretation, I shall retain 7-point scales, and for the analyses of the rank-order data I shall use the Chi-Square test, the Median test and the Mann-Whitney *U* test.

The weakness with the Chi-Square test here is that since it is a simple goodness-of-fit test, it fails to exploit information concerning rank ordering. Similarly the Median test, concerned only with the location of a score above or below the combined median, also fails to exploit information about rank ordering. The Mann-Whitney *U* test considers rank values and is therefore a relatively more powerful test. Its weakness lies in tied ranks, though a correction procedure is available.

An alternative to the Mann-Whitney *U* test would have been the two-sample Kolmogorov-Smirnov test, a test which compares the observed cumulative frequency distribution with that of a uniform distribution. In practice the test focuses on the largest of the observed deviations from the expected population distribution, and hence is sensitive to any type of difference in the two distributions (e.g. median, dispersion, skewness). However, the test is sensitive to the value of *N*: Siegel and Castellan (1988:145–7)[13] suggest that 25 is the critical figure; Hatch and Lazaraton (1991:282) and Bryman and Cramer (1997: 133) state that it is 40. More problematic than minimum sample size is the fact that the Kolmogorov-Smirnov test expects the variable being examined to have a continuous distribution, and this is not necessarily the case with attitudes. Hence my reluctance to use this test.

[13] The original edition of this book—Siegel (1956:129)—mentions 40 as the critical minimum figure per group.

3. ATTITUDES TOWARDS PRONUNCIATION ERROR

> They spell it Vinci and pronounce it Vinchy; foreigners always spell better than they pronounce.
>
> —Mark Twain, *The Innocents Abroad*

Speakers of foreign languages are regularly perceived to speak them with a non-native accent and only rarely are they mistaken for native speakers. Foreign accent is recognised by several features, particularly the number, type and severity of segmental misarticulations, and such non-segmental features as abnormal stress placement, poor rhythm, inappropriate intonation patterns, slow tempo and pausing (Varonis & Gass 1982, Bond & Fokes 1985). Perhaps because of this wealth of vocal signals betraying non-native ability, a foreign accent is noticeably persistent and irremediable, even people who start learning an L2 at a very young age (under 9) and who move to an L2-speaking country still retaining some degree of foreign accent (Flege 1984). However, learners improve noticeably in their tacit knowledge of what an L2 sentence should sound like, the aural perceptual ability outstripping and surpassing the motor-mechanic productive skill. Features of foreign accent are of course often accompanied by mistakes in grammar and lexis, but pronunciation (excluding intonation) and hesitation together have been found to distract both native and non-native listeners most from a message (Fayer & Krasinski 1987), and phonetic mistakes in and of themselves may be sufficient to provoke misunderstanding and negative reactions in listeners. An important goal for foreign learners and their teachers, therefore, is damage control, the treatment and mitigation of those mispronunciation features which signal foreignness most clearly and cause most distraction in listeners.

It has been hypothesised that the perception of foreign accent rests upon the poor fit between a listener's phonetic category 'prototypes' and a speaker's production of the phone aimed at, i.e. upon the deviation between the expected phone and the heard phone (Flege 1984, 1988). Listeners assign a category space to a phone and a tolerance region (TR) to that space, the TR varying as a function of such factors as social context, conversation context (Carranza & Ryan 1975), the listener's attitude (Brennan & Brennan 1981a and 1981b), and the listener's and/or speaker's social background (Ryan & Sebastian 1980). When the production of a phone is considered to fall outside the TR, then the greater the distance of the realisation from the borders of the region, the greater the degree of perceived foreign accent.

Native listeners are acutely sensitive to any trace of a foreign accent, even to subcategorical phonetic differences: detection is easy, regardless of the origin of the foreign accent, though recognition of the particular accent is not, depending on the severity of the accent, the listener's experience, etc. (Flege 1984). Curiously the removal of pauses from test recordings presented for accent judgement has no appreciable effect on foreign accent scores (Flege 1988): silence does not indicate foreignness. This would seem to imply that degrees of perceived accentedness depend on hierarchies of phonemic errors and non-segmental errors with an inter-dependence between the two levels.

Both the concept of foreign accent and the concept of pronunciation error necessarily entail the idea of a standard or a native pronunciation against which utterances can be judged. However, no such standard exists: it is a linguistic chimera. Regional and national variations, social differences and stylistic shifts, all lead to a plethora of potential standards. Nevertheless, paradoxically, linguistically naive native listeners are probably more than willing to claim categorically that Person A's pronunciation is 'foreign,' 'foreign-accented' or 'wrong,' and Person B's is not.

3.1. Methodological problems in investigating pronunciation errors

The investigation of reactions to mispronunciation and the evaluation of mispronunciation gravity is fraught with methodological difficulties. The practical problems are enormous. Opinions and attitudes are always hard to get at, but the obstacles to overcome are of a subtly different order to those when, for example, investigating people's beliefs ('Are you in favour of capital punishment for convicted child-abuse offenders?'), or people's preferences in whisky ('Take a sip of this Glen Morangie and a sip of this Isle of Jura: which do you prefer?')

One major methodological complication is the authenticity of the material to be evaluated, the stimuli. Some people rail violently against the use of any non-authentic material, stressing, quite rightly, that you would never offer someone an ersatz whisky in order to find out whether they like the real McCoy. However, others are quick to point out that authentic native-speaker discourse is full of false-starts, interruptions, hesitation fillers, slips of the tongue, slurring, ungrammaticality, etc. and thus has limited value for research into acceptability or error gravity. Moreover, the error tokens in recordings of natural spontaneous speech are haphazard, and some anticipated errors may be absent altogether, and thus it is impossible to correlate specific errors with specific reactions and evaluations. The alternative is to use artificially constructed material, but then the results are problematic in that there is no obvious connection between reactions to such sanitised material and to the allegedly disorderly jumble of so-called normal communication.

While some research may have gone too far in the construction of artificial test material, perhaps more tolerance and understanding is needed than that shown by those who argue vehemently against the use of 'simulated rather than authentic learner discourse' (Rifkin & Roberts 1995:516). There is little point in insisting that researchers should use 'authentic sample[s] of learner language,'

that 'we must design studies that come as close as possible to using authentic language . . . actual learner language' (*ibid.* Pp. 517–8). The criticism of listening tests using specially recorded texts clearly overshoots the mark: 'although script recitation can . . . capture authenticity of accent . . . the recitation of isolated utterances or paragraph-length texts is clearly not natural.' This is nonsense. While the reciting of a text is not something we all do every day, we certainly all spend a lot of our everyday lives listening to (and reacting to, digesting, and inwardly trying to understand) recited texts. What are, for example, news bulletins, public address announcements, vast numbers of radio and TV programmes, political speeches, radio and TV advertisements, and particularly films and plays, if not recited texts? And which (if any) academics can honestly claim that they have never had to sit through (dare I say 'given'?) the recitation of a written text in a lecture theatre? The reading aloud of a text *is* natural (though perhaps not common in that relatively few people do read aloud), and the listening to recited texts is even *more* natural (and it is a *very* common and usual activity, too).

What is perhaps unusual is the forcing of listeners to evaluate the speakers they have just heard, though even here I would hesitate to say that such forced evaluation is always unnatural: leaving a cinema or theatre, surely it is natural to turn to one's companion and enquire, 'Well, what did you think of that?' (The focus of this question is not specifically on the speech of the film, I concede that, but it certainly embraces the speech.) However, having just heard a railway platform announcement, I readily admit that it would be unnatural to ask one's companion the same question ('What did you think of that?'), though one might say, 'Well, I didn't make out a word of that. Did *you*?'

The qualms expressed by Rifkin and Roberts (1995) are not in fact supported by other writers. According to work conducted by Munro and Derwing (1994), material spoken spontaneously and the same material read by the same speaker from a written transcript are perceived as being the same as regards degree of foreign accent, even though there may be noticeable differences between the two versions (see also Flege 1984:704). At the same time as the authenticity or otherwise of stimulus material provokes discussion and disagreement, there is a second and separate issue here: the way in which informants' reactions are collected and measured, the stimulus context, the experiment environment. This is yet another methodological challenge. The problem is that the very act of evaluation of a single phrase or short passage of doubtful semantic significance, spoken by an unseen stranger and heard outside of any communicative context, is totally unrealistic, completely artificial (see *inter alia* Davies 1983:309). It is indeed probable that informants have no ready-made and explicitly or clearly formulated attitude about the accent/voice they are asked to evaluate: the test itself prompts the creation of a previously unformulated attitude and it is well known that such spontaneous 'reactive' attitudes are extremely unstable.

Moreover, the evaluation is a subjective (or judgement) test as opposed to an objective (or operation) test (see Quirk and Svartvik's distinction: 1966:18–21). While the filling in of a judgement questionnaire as a conscious response to a linguistic stimulus is not a usual activity in everyday conversations, the perform-

ing of some operation as a response to a linguistic stimulus is common, more lifelike, even normal, but unfortunately the correlation between the error gravity and the provoked informant behaviour is indirect and therefore not necessarily valid or reliable. Both approaches pose inordinate problems, but the vast majority of studies opt for the first, for subjective direct evaluations. (See Johansson 1978 for studies showing close correspondences between the results of subjective and objective measures.)

This problem of stimulus context remains unsolved and this fact should prompt us all to interpret error gravity results with caution. Nevertheless, once again, I feel Rifkin and Roberts overstep the mark:

> Without clear contexts within which to evaluate language output, we cannot claim with certainty the source of trouble in any given utterance, and therefore cannot make serious claims as to the relative gravity of one type of error over another. . . . [We need] naturalistic stimuli, well situated in a larger communicative context. (Rifkin & Roberts 1995:519–20)

The semantic content of sentences and passages chosen for evaluation raises surprisingly few passions: perhaps we are all too blasé nowadays, having been anaesthetised by Chomsky's now notorious (written) sentence 'Colourless green ideas sleep furiously.' To say that one sentence is natural and meaningful, and another unnatural and meaningless, is an arbitrary value judgement, yet some studies still resort to isolated words and word lists and self-consciously struggle to justify the practice. For experiments contrasting different pronunciations, enormous time, patience and optimism is required to cull the phonetic contrasts desired from real news bulletins, plays, filmscripts or whatever, i.e. from texts which have been recited in 'normal conditions.' As a result some authors—and I group myself with them—have no hesitation in using, as their stimulus material, texts which have been especially constructed and recited, merely ensuring that the text is broadly neutral—neither humorous nor scurrilous, neither political nor religious, neither childish nor demagogic. On the other hand, other authors still argue that no text can ever be truly neutral (Ryan & Carranza 1977; Giles & Coupland 1991:54–5). What is needed is a whole series of experiments using different texts with different semantic contents to corroborate or contradict findings.

From the methodological point of view, the stimulus voice itself also deserves careful attention. Two people's accents may be recognised as being the same (e.g. Lancashire, Australian or Italian-accented English), and yet—as we all know—there are still clear differences between the two voices. Certain differences can be held as control variables relatively easily—gender, voice quality, volume, tempo, and pitch range and variation (see my review in the next chapter)—but others, such as the precise phonetic segmental pronunciation, are less amenable to control. Studies of mispronunciation must stipulate precisely what errors they are investigating.

3.2. Grading of pronunciation errors

Objectively one might expect pronunciation errors to be graded and ranked for seriousness according to the degree of deviance and to the interference they

provoke in communication, but evaluating and quantifying the deviance and disruptiveness of mispronunciation is not easy.

For example, the frequency and regularity of occurrence of a particular phonemic error may often determine its perceived seriousness: one token may not be noticed, but a more consistent (though not necessarily exclusive) repetition may become a hallmark. However, whether such a signature tune is irritating or pleasing to a listener's ear is not easy to predict. A frequent mistake may be expected, and may pass practically unnoticed, the listener having become accustomed and insensitive to it. On the other hand a frequent mistake may be a conspicuous positive mistake, a plus in the acceptability stakes (as was the case with Maurice Chevalier's accent). However, a regular mistake may be a negative mistake, a source of extreme irritation, a turn-off for the listener. And here the distinction drawn by Corder (1967:166–7) between an error and a mistake should perhaps be recalled: the consistent error (the error of competence) is likely to be judged more severely than the occasional mistake (the slip of performance). Cartier (1987) draws attention to the same point, using the terms (somewhat confusingly) 'articulation error' and 'pronunciation error' respectively. Unfortunately consistency in pronunciation (or mispronunciation) cannot be guaranteed, and it is difficult to accommodate this basic fact in a practical system of grading errors (Samarin & Kalmar 1979:184). Teachers may recognise and allow for variation (improvement and attrition) in the pronunciation performance of learners, and may explain it by linguistic (in-)security, tiredness, communicative situation or whatever, but naive non-trained non-linguist listeners probably assume that the performance of a given speaker at a given moment in a given context is indicative of a general ability at all times in all situations. Few people, I suspect, appreciate the fact that non-native speakers have a continuum of accents, stronger to weaker, though it may be of a relatively narrow range and for the most part beyond their control.

Frequency, regularity and variation, however, are tempered by perceived degree of error or deviance. Naive listeners and expert phoneticians regularly agree on degree of accentedness even though their method for deciding are entirely different (Brennan et al. 1975; Brennan & Brennan 1981a and 1981b; Anderson-Hsieh & Koehler 1988). At its mildest the deviance may consist of distortion, sub-phonemic errors giving rise to a mild foreign accent: the speech is distorted but is easily understandable, is acceptable. At its worst, deviance may constitute substitution, phonemic errors constituting a thick heavy foreign accent: the speech is so distorted that it is incomprehensible, is unacceptable. Occasionally deviance may even be manifested by omission. However, the evaluation of the degree of deviance is not easy: there is no obvious cline here. Is the omission of voicing more serious than wrong place or wrong manner of articulation? Is a vowel error more serious than a consonant error? Is the deviant articulation quite simply a non-native deviation, or is it also found among some native (and perhaps native non-standard) accents and if so, how widespread/well-known is it? (See Johansson 1978:92 ff.)

The discriminatory load (or functional load) of a phoneme is also important. It may be so high that only a few (if any) tokens of inaccuracy (whether minor or major, however that is judged) are tolerated, the disruption and incomprehension caused by the deviance being so enormous that the listener gives up the struggle to understand and the speaker is labelled as hopelessly and unacceptably foreign-accented. Such would undoubtedly be the case, for example, in a conversation on oenology where 'vine' is consistently mispronounced as 'wine.' The mistake would be the death knell of the conversation. A phonemic mistake is—at least theoretically—a more serious mistake than an allophonic mistake, though whether there is a clear division between the two is perhaps tendentious, and it is indeed doubtful whether the variability of communicative situations makes such a division tenable. The touchstone is probably not so much phonemic v. allophonic status as discriminatory load.

One kind of allophonic error may be referred to as infringement of a general pronunciation rule, such as the compulsory lenis or fortis production of the plural 's' marker in English. A speaker may be able to produce fortis and lenis alveolar sibilants (*sip* v. *zip*, *bus* v. *buzz*), but consistently fails to use lenis 's' in such words as *dogs*, *tables* and *chairs*.[1] One listener, perhaps a linguistically naive listener, may rate the mistake as merely quaint; another listener, perhaps a language teacher, may rate it as terrible (certainly some guide-lines given to language teachers recommend more draconian penalties for such errors than for others, probably because it is easy to recognise them and classify them [see Johansson 1973:102–3]). The discriminatory load here is low, and the disruption to comprehension minimal, but the irritation may be overwhelming.

Comprehension and irritation do not always sit comfortably with each other, and concern has been expressed that many error gravity studies confuse them (Ludwig 1982; Gynan 1985; Rifkin & Roberts 1995). The fact is that the foreign accent of some speakers may be extremely irritating (though the level of irritation is difficult to quantify) yet the message conveyed may be perfectly understandable (Fayer & Krasinski 1987). Alternatively, speakers may be completely unintelligible (and here the degree of unintelligibility is easy to assess: it's simply zero) even though the impenetrable accent is not exactly irritating. However, note that intelligibility is not the same as comprehensibility: how much have listeners understood of what has been said, and can they transcribe the words accurately (= intelligibility)? how do listeners perceive the intelligibility of the speech, and how do they rate it (= comprehensibility)? (Munro & Derwing 1995; Derwing & Munro 1997).

Combinations or co-occurrence of particular pronunciation mistakes may also be significant in determining perceived error gravity, though little research has been conducted into this (see Brennan & Brennan 1981a). There is no obvious reason to assume that all features of mispronunciation impinge equally on the ear of the listener. Why should the cumulative effect of, for example, the mispronunciation of the RP palato-alveolar lenis affricate / dʒ / as an alveolar [dz] or

[1] This is of course a hypothetical example. Whether such a speaker exists, and whether such an interlanguage pronunciation pattern exists I don't know.

postalveolar [dz] in combination with the mispronunciation of the lenis labio-dental fricative / v / as the non-fricative continuant [ʋ] be the same in degree of gravity as the same affricate mispronunciation in combination with the mispronunciation of the palato-alveolar fortis sibilant / ʃ / as a postalveolar sibilant [s]?

The relative difficulty of an articulation has sometimes been put forward as a valid factor in assessing the seriousness of mispronunciation (Dretzke & Martin 1975, Dretzke 1987), but it is difficult to see how one can quantify such a factor in a satisfactory way. What scale can be used to grade difficulty, what scale to assess seriousness? Although some clarification of this point might be welcome in order to 'reward' the hard work of learners, the degree of effort which has gone into mastering an articulatory difficulty, and the amount of skill which is involved in producing a correct articulation, have no bearing whatsoever on how the average listener (as opposed to learners themselves and knowledgeable teachers) values the accent.

A further factor which will affect perceived seriousness of mispronunciation is the *listener*'s knowledge of the language being spoken, their familiarity with the *speaker* (or the 'type' of speaker) and their predisposition to hearing or expecting a mistake. Who, quite simply, is the listener grading the mispronunciation? Listener characteristics are by no means trivial: native listener v. foreign learner (Smith & Bisazza 1982); listener's mother tongue and dialect (Davies 1983:308); listener's age and gender; listener's educational, socio-cultural and socio-economic background; listener's language skills; listener's degree of exposure to (the) non-native accented speech (Wingstedt & Schulman 1984; Hultfors 1987) and to native accents (even to native defective speech); listener's familiarity with (accommodation to, diminished sensitivity to) the particular accent/mispronunciation features ('I've grown accustomed to your voice'). It is sometimes claimed that people who are for ever listening to many different accents of English (whether they be regional accents of English or foreign accents) will notice non-standard or defective pronunciation less readily than others. Some listeners may indeed not recognise or understand the concept of mispronounced English, feeling that all accents and pronunciations are equally acceptable. Note that students and teachers of languages, linguistics and logopedics do not always agree with lay untrained naive listeners as to what an error is, let alone how serious it is (see *inter alia* Perrin 1954; Quirk & Svartvik 1966: 32–4; Davies 1983).

Moreover, a listener's mood may be decisive in the evaluation of an error: with no willingness to understand, an error, whether big or small, will always be a serious error, unacceptable, incomprehensible. Perception of errors and comprehensibility, therefore, is not only a function of phonetic deviance, but also of the listener's transitory emotional state, or their general attitude towards foreigners and foreign speech (Anderson-Hsieh & Koehler 1988). The perceived seriousness of an error is strongly influenced by any ill-will brought by the listener to the conversation, and by any ill-will (generalised negative affect: Ryan

1983) prompted by the extra effort, frustration and uncertainty resulting from trying to understand the mispronounced English.

In addition the communication setting (e.g. business meeting v. chance encounter in a pub) and the speaker's role (e.g. speaker on an international airport public address system v. lumberjack felling trees in the back of beyond; highly advanced proficient professional foreign-language speaker v. beginner) may also seriously influence the perceived seriousness of pronunciation errors.[2]

3.3. Error hierarchy and error preference in foreign-accented speech

While error analysis has received considerable attention over the years, contrastive error evaluation has perhaps enjoyed less scrutiny. The situation has not changed much since Carl James and Pitt Corder wrote the following over twenty-five years ago:

> A linguistic approach to error analysis should at least attempt to explicate and rationalise the teacher's subjective evaluations. . . . At the present time nothing is known about the relative gravity, from a native speaker's point of view, of *errors of pronunciation*, of grammar, or of lexis. (James 1972:76. My stress.)
>
> We do not yet have . . . any hard facts about attitudes to, or expectations of, foreigners' linguistic behaviour in our own or anyone else's culture. (Corder 1973:282)

Evaluation of error gravity and people's reactions to different types of errors are nevertheless important. Without any understanding of the seriousness of pronunciation mistakes and of people's differing reactions to pronunciation mistakes, foreign language learners and teachers of pronunciation are in danger of wasting time on irrelevant minor details and/or downplaying particulars with enormous and often incalculable impact and repercussions.

In addition, little is known of people's reactions to and attitudes towards variously foreign-accented speech, though a considerable body of data is slowly building up on those towards regional-accented native English.

Attitudes towards foreign-accented speech and error gravity go hand in hand (Eisenstein 1983), and this interface is the starting point for this present study of attitudes to Finnish-accented English. However, a comprehensive search of the literature shows that there is a dearth of material and even interest in this area.

One study of the gravity of errors in *written* English claims that a scale of perceived error can be established:

> a) native judges stigmatise errors in tense and concord most, while non-native judges stigmatise errors in case and lexis most, and
>
> b) the error-types considered most serious overall were, in descending order: transformations, tense, concord, case, negation, articles, order. (James 1977)

Moreover, native judges are generally more lenient than non-native judges, individual assessors tend to be consistent in their judgements, and native judges tend to use a narrower mark range than non-native judges. An extensive study of

---

[2] Note that there is some evidence to suggest that a speaker's degree of accent is *not* affected by setting in that a learner's attention to pronunciation does not necessarily improve performance and authenticity/accuracy (Flege 1984:704).

'Swedish-contaminated' written English produced similar results (Hultfors 1986).

However, few comprehensive studies of features of mispronunciation in foreign-accented speech have been conducted. One such study has been undertaken with 17-year-old native informants from north-east England listening to German-accented English (Dretzke 1987): see Table 3.1 below. However, there is no discussion as to whether *all* native English listeners should be expected to react in the same way, or whether other regional and social-accented native listeners of the target language (let alone listeners of other mother tongues) would react differently. In one study of American-English-accented Spanish, informants were asked to evaluate the pronunciation of vowels, diphthongs, 'p, t, k,' and 'b, d, g,' and 'r' on separate 7-point scales, but no meaningful hierarchy of perceived degree of error is presented (Gynan 1985). In another study with Chinese-accented English, it was found that at fast talking rates prosodic deviance affects comprehension more adversely than segmental deviance, though the absence of pauses to mark thought groups, and poor word and sentence

Table 3.1 Perceived seriousness of mispronunciation in German-accented English

Grade	Mistake
Serious	1 [p] for [b̥]
	2 [k] for [g̊]
	3 [ŋ] for [ŋg]
	4 [s] for [θ]
	5 Steep rising intonation for falling or level intonation
	6 Syllable-timed for stress-timed rhythm
	7 [ʁ] for [ɹ]
	8 [ʊ] for [w]
Medium	9 [ŋk] for [ŋ]
	10 Wrong wordstress
	11 [ʃ] for [ʒ]
	12 [ɛ] for [æ]
	13 [l] for [ɫ]
	14 [f] for [v̥]
	15 [t] for [d̥]
Minor	16 [z] for [ð]
	17 Jagged instead of smooth intonation
	18 [tʃ] for [dʒ]
	19 Strong Form for Weak Form
	20 [oː] for [əʊ]
	21 [s] for [z̥]
	22 [ʔ] + vowel for pause/silence + stressed vowel
	23 [ʊ] for [v]

(Adapted from Dretzke 1987)

stress, also played a significant part in decreasing comprehensibility (Anderson-Hsieh & Koehler 1988).

Few other comparable foreign accent studies have been made of segmental mispronunciation, suprasegmental mistakes, and combinations of the two. We simply do not know what gravity native and non-native informants—when listening to various foreign-accented languages—assign to errors in, for example, plosives, sibilants, fricatives and approximants, and to the inappropriate use of falling and rising pitch contours, what combinations of pronunciation mistakes they consider to be less (or more) foreign-accented, and what mistakes they find to be irritating or (possibly) pleasing.

Those investigations of mispronunciation in non-native speech which have been undertaken seem to have followed the same paradigm more or less as most sociophonetic research into native speaker accents of English, viz. what image of the speaker's personality does the foreign-accented pronunciation produce. Few studies (apart from Wells 1995) seem to be interested in preference or error hierarchy: which mispronunciation is better? which mispronunciation do you prefer? Intelligibility is not the issue here, though of course informants may be swayed in their choice by suspicions of *un*-intelligibility or of potential (if not real) confusion (e.g. *pat* v. *bat*). Preference for one version over another cannot be taken to indicate the better comprehension of that version (a higher comprehensibility rating), nor to indicate irritation at the rejected version. Preference may even be interpreted by some listeners as acceptability (see Santos 1987), or as conformity (see Johansson 1973:102), i.e. as 'closeness to native-speaker pronunciation,' the target that the speaker is presumably expected to be aiming at. However, preference for the 'better' version cannot be analysed further than what it is: a preference. Both versions may be intelligible, and/or irritating, and/or acceptable, and/or 'native-speaker-like,' but only one version is preferable.

It is worth noting that preference may sometimes be an unconsciously motivated choice. In studies of dichotomous choices, all the informants listen to the same tape and hear the same differences between the alternatives offered, though some informants, for some of the pairs, are totally unaware of having heard any difference at all, perhaps even swearing blind that the alternatives are identical. In such cases the informants choose one or the other without being able to explain why or how they choose one rather than the other. Nevertheless, their choice is motivated, albeit unconsciously, by what they have (unconsciously) heard. Such a situation is not random choice, not 'arbitrary polarisation' (Quirk & Svartvik 1966:13). It is for this reason that it is perhaps better to use a 2-point scale 'Preferred–Disfavoured' rather than a 3-point scale 'Preferred–No Preference–Disliked.' One's results may reveal a group of informants' overall choices splitting 50/50 between two accents or pronunciations A and B, and one can say that there is no preference within the group as a whole one way or the other. Nevertheless, each individual's personally expressed preference is not random, and there may be underlying explanations for the split to be found elsewhere (e.g. gender of speaker, gender of informant, age of informant, etc.).

Any attempt to achieve some sort of objective assessment of foreign-accented English and to produce an error hierarchy depends upon surveys carried out with naive native listeners. No extensive study (as far as I am aware) has ever been carried out of listeners' evaluations of Finnish-accented English speech. Certainly no systematic studies have yet been undertaken where speakers of various different languages are asked to evaluate Finnish-accented English. Such a study might shed interesting light on the matter of whether there are universal patterns of preference.

4. ATTITUDES TOWARDS FOREIGN ACCENTS AND PERCEPTION OF FOREIGN-ACCENTED SPEAKERS

> The man Hari fell foul of probably spoke English with a Midlands accent, and resented the fact that an Indian spoke it like a managing director.
>
> —Paul Scott, *The Jewel in the Crown*

Research into attitudes towards accents of a language essentially recognises that people listen to spoken language for more than just its content or referential information: people also attend—both consciously and unconsciously—to its vocal 'signals' for clues as to the personal and social characteristics of the speaker. On the basis of the voice alone, people deduce such things as a speaker's gender and age, emotional status (e.g. happy or angry), job competence (e.g. successful or unambitious), benevolence traits (e.g. friendly or dishonest), social attractiveness (e.g. acceptable as a member of the family, or someone to be shunned) and regional, national and ethnic origin. The perceptions of a speaker so formed have an enormous impact on social interaction and communicative behaviours, and have been the focus of a considerable volume of L1 and bilingual research for some seventy to eighty years now. Relatively little attention has been paid to the perception of foreign-accented FL speakers, but the results of all this research so far is of considerable importance to teachers trying to improve the pronunciation of foreign language learners, for such teachers need to be concerned about the model they present and about the image that learners give to their interlocutors with their foreign accent. An awareness of how people react to both native and foreign accents is essential, and teachers and learners need to know how different combinations of accent features and different degrees of accentedness affect listeners' perceptions of speakers.

The first studies in the field of accent perception appeared in the late 1920s and early 1930s, a period which coincides more or less with the advent of the recorded voice, the telephone and radio broadcasting. People could now explore how listeners form impressions of the personality and emotional state of a speaker without the person being physically present in front of them, how they react not to the whole bundle of interrelated and interacting features made up of the speaker's personal appearance, clothes, hairstyle, gestures, behaviour and so on, of the ideas and opinions the speaker expressed, and of the voice used to express them, but to the voice alone. Clearly a person's voice carries decodeable signals which are left intact even when the speaker is 'absent,' the quality and richness of the signals nevertheless depending on the quality of the instruments

used to record and/or transmit and/or reproduce the voice. At first the focus of attention was on attitudes towards *native* accents of a language, but attention also turned towards *foreign* accents of a language.

In 1927 Edward Sapir claimed that although everyone intuitively recognises that speech is 'an index of personal expression' (1927:892), it is extremely difficult to examine the various facets of someone's personality as revealed by speech. Speech is a complex consisting (he believed) of five levels of behaviour: the voice as such (voice quality), speech dynamics (intonation, rhythm, relative continuity and speed), pronunciation (at the segmental level), vocabulary, and style. But during impression formation the expressive values presented on one level may be contradicted by those on another level. Hence speech is a highly variable phenomenon and open to infinite individual variations. Moreover, it is limited by the linguistic norms of the society in which it is used, differences between languages in these linguistic norms necessarily rendering the signalling and identification of personality in a foreign language extremely difficult.

Concerning pronunciation on the segmental level, it is significant that Sapir points out that it would be wrong to draw conclusions about people's personalities on the basis of their foreign accent when speaking a foreign language: 'It would manifestly be wrong to make inferences of a personal nature from such mispronunciations' (1927:901). However, the implication is that native listeners *do* make such inferences, consciously or unconsciously, and the listener, Sapir implies, forms two impressionistic pictures of the speaker: the first concerns the national stereotype (the speaker produces such and such phonemic mistakes, is thus e.g. French, and therefore is such and such a sort of person, conforming to certain patterns of behaviour, etc.); and the second concerns the speaker's own individual personality. Sapir recognises the difficulty of combining speech and behavioural studies, but concludes that it is essential to analyse speech in more detail because 'intuitively we attach an enormous importance to the voice and to the speech behavior that is carried by the voice' (1927:905).

For the next forty years or so, Sapir's influence is clear: studies concentrated mainly on speaker personality in native and bilingual speakers and avoided foreign accents.

4.1. Early studies

The first few pioneering experiments exploited the revolutionary invention of the day: the radio. For example, one investigation was conducted into 'Radio-Personality' with the active support of the newly formed BBC, and with the participation of 4000 radio listeners (Pear 1931).[1] Although the study is weak in analysis and presentation, gender, age and occupation would appear to have been well recognised. A similar German-language study conducted with the collaboration of Radio Wien and with 2700 listeners answering a questionnaire published in the Radio Wien magazine produced similar results (Herzog 1933).[2]

---

[1] This study is extremely difficult to get hold of but is referred to in many studies. I have not seen it myself.

[2] This text is also difficult to obtain, and I have not seen it myself.

In another experiment to test people's judgements of personality the broadcast voice was used together with a live 'face-less' voice presented from behind a curtain (Allport & Cantril 1934). All evaluations of the voices were checked against the real 'objective' values, and significant correlations were found between the real and rated values of age, extroversion–introversion, ascendance–submission, dominant values and summary sketches. Most importantly the strong patterns of consistency that were found suggested stereotyping and the halo effect were at work.

Stereotyping was also recognised as a powerful factor in a fourth experiment evaluating the psychological traits of aggressiveness and nervousness when listening to a voice, though this time the result was ascertained tackling the subject from a non-holistic angle: specific vocal cues were used rather than an overall vocal impression (Stagner 1936). Preconceived stereotypes for voices were also found to play a role in a whole series of closely related experiments testing people's ability to recognise and judge specific characteristics and traits (e.g. intelligence, leadership, tiredness) from the voice as transmitted over a public address system (Fay & Middleton 1939–44). Since many judgements proved to be unreliable, the authors deduced the probable role of preconceived stereotype voices in the rating process.

Several other studies appeared at this time, examining diverse aspects of the voice in relation to various personality traits and emotions (e.g. Fairbanks & Hoaglin 1941; Moses 1941a, 1941b, 1942; Sanford 1942a, 1942b; Duncan 1945; Ruesch & Prestwood 1949).

However, one text published just after the War became a classic: Asch (1946). This work does not in fact focus on how impressions of personality are produced and affected by the voice, but it does make important observations concerning the use of adjective polar pairs in impression and attitude formation. In addition it opposes two models of how impressions are formed: an algebraic elemental model where each trait is separate and forms its own impression, and are added together to form a total picture; and a holistic configural model where an overall impression is formed, with each trait being perceived through, and in relation to, every other trait. The results of the study highlight the halo effect and the role of age, gender, experience, time (contemporaneity), culture and nationality, the last two again indirectly warning of the problems inherent in between-language/ cross-language impression formation.

The 1950s and 1960s produced more important work in the perception of personality through the voice. For example, Firth (1950) broached the whole area of personality and language in society, while Starkweather (1956a, 1956b, 1961, 1964) concentrated on content-free speech and the audible elements of para-language as the information channel for speaker personality, as also did Davitz and Davitz (1959a, 1959b, 1961), Kramer (1963, 1964), Austin (1965), Mehrabian and Ferris (1967), Abercrombie (1968) and Mehrabian and Williams (1969). Other authors focused on voice quality (Diehl et al. 1959; Hargreaves & Starkweather 1964; Markel et al. 1964; Hargreaves et al. 1965; Laver 1968; Markel 1969), tone of voice (Pittenger 1957) or some other vocal feature such as speak-

ing rate and pitch variety (Mallory & Miller 1958; Addington 1968; Markel 1965). It was recognised, however, that the content of speech, the gender of the speaker and even features of the listener (e.g. intelligence, bilingualism and even prejudices concerning stereotypes) exert a powerful influence on speaker perception (Anisfeld & Lambert 1964; Markel & Roblin 1965; Addington 1968; Gardner & Taylor 1968).

These initial studies should be seen in the context of a wide range of developments in theories about attitudes and attitude formation, in the techniques available for measuring attitudes, and in social psychology (e.g. Guttman 1944, 1947; Heider 1946; McNemar 1946; Edwards & Kilpatrick 1948; Hovland & Sherif 1952; Cardno 1955; Webb 1955; Edwards 1957a; Festinger 1957; Helson 1959; Rosenberg et al. 1960; Sherif & Hovland 1961; Upshaw 1962; Tajfel 1962, 1969; Cook & Selltiz 1964; Osgood 1964, 1965; Diab 1967; Fishbein 1967; Shaw & Wright 1967; Bem 1967, 1968, 1970; Sherif & Sherif 1967; Greenwald 1968; Greenwald et al. 1968; McGuire 1968; Zajonc 1968; Kerlinger 1973; Kiesler et al. 1969).

4.2. The matched-guise technique

The two simplest methods of accessing and logging people's attitudes towards accented speech were thought to be the content analysis of written documents (the indirect approach), and interviews with informants and the use of questionnaires (the direct approach, e.g. Taylor 1973, Cohen 1974). However, an important step was made by Lambert and his colleagues in Montreal in the 1960s and 1970s: they invented, used and perfected a so-called 'matched-guise technique' (MGT). The method was simple but revolutionary: the investigators elicited the reactions of respondents to recordings of dialectally different versions of the same voice. Thus, a bilingual or bidialectal (or multilingual/multidialectal) speaker made two (or more) recordings of the same text using two (or more) native accents of a language (or languages). The recordings were then presented to listeners for assessment in such a way that the listeners were unaware that they were listening to the same voice.

The first pioneer study to use the MGT (Lambert et al. 1960) ignored Sapir's earlier warnings concerning the influence of the conventions of different languages on the perception of the personality of native speakers in different languages, and hence perhaps belittled the dangers of cross-language distortions. However, it established a basic methodology which is still adhered to today. Four bilingual (French/English) speakers recorded a text in French and in an English translation equivalent of it. French and English Canadians (all monolinguals, though one may assume they knew the 'other' language) then rated the recorded voices on 6-point scales from 'very little' to 'very much' for fourteen personality traits, and also answered various open questions about their language use, language preferences and attitudes towards English and French Canadians.

The results surprised everyone: both English and French background informants favoured the English guises; and the French informants rated the French guises generally less favourably than the English informants (own group rejec-

tion). The English informants upgraded the English guises for looks, height, intelligence, dependability, kindness, ambition and character, but the French informants *also* upgraded the English guises for intelligence, dependability, likeability and character. Curiously no correlations (and no in-group/out-group biases) were found between the ratings of the guises and the attitudes towards English and French Canadians as expressed in the answers to the other questions. The conclusion was that the evaluational reactions to the voice guises reflected community-wide stereotypes shared by both English-speaking and French-speaking Canadians: representatives of the minority linguistic group (the French Canadians) are downgraded by *all* members of the community and are considered as being relatively 'second rate' people (the data is reviewed and re-examined in Lambert 1967).

Although clearly placing their study in the bi-cultural diglossic community of Montreal with its historically and socially significant French–English schism, the authors do not reveal whether they think similar results would be found with bi-dialectal (or multidialectal) studies in monocultural monolingual French and English communities. The study, being both multi-language and cross-language, is a special case, and it is by no means certain that the factors at work are the same as in monocultural monolingual societies. Few people seemed to acknowledge the fact then, and few people would appear to acknowledge it even now.

The MGT of this first seminal study by Lambert et al. (1960) was soon attacked for artificially limiting information about the speaker and forcing attention onto the voice alone (Tajfel 1962:38–42): 'situation' is completely absent, and judgements about the speaker are necessarily based on stereotypes that the voice brings to mind. The ability of the producers of the vocal guises to maintain the same character and personality, speech rate and paralinguistic features across guises was also doubted (Tajfel 1962). Moreover, if listeners hear a guise in terms of a particular stereotype, who is to say that the speaker did not *produce* the guise with the same stereotype in mind?[3] Is the guise credible and authentic, or is it just a caricature?

A second study restricted itself to two accents of one language, investigating listeners' reactions to 'pure' and Jewish-accented English (Anisfeld et al. 1962). The same MGT, almost identical procedures and the same fourteen traits were adopted. In brief, both Jewish and gentile respondents rated flawless English guises more favourably for height, good looks and leadership than Jewish-accented English. The gentile respondents perceived no traits at all as more favourable in the Jewish guises, whereas the Jewish respondents did (for sense of humour, entertainingness and kindness). The Jewish guises were downgraded by everyone for height, good looks and leadership, and also devalued for self-confidence and either sociability (by the Jewish informants) or dependability (by the gentile informants). The results showed that Jewishness was not the factor prompting down-grading on certain traits but rather foreign/non-standard accent. Confirming the findings of Lambert et al. (1960), the lack of a relationship

[3] The point is repeated by Smith (1985:89).

between the ratings and the personal variables suggested that community-wide stereotypes about people with accents are indeed operative.

A third study, Anisfeld and Lambert (1964), was again a multi-language (French and English) study but used 10-year-old Canadian informants who were either monolingual (French) or bilingual (French/English). Basically four 10-year-old bilingual girls recorded a passage from Little Red Riding Hood, once in English and once in French (though one might legitimately wonder whether 10-year-olds can hold personality, tempo and paralinguistic features across guises), and the guises were evaluated on fifteen bipolar traits presented on 5-point scales. The monolingual informants upgraded the personalities of the French guises on all traits (in-group support) except height (but in fact French Canadians *are* shorter than English Canadians), but the bilingual informants saw few differences between the two guises. Curiously the bilinguals revealed a slightly more favourable attitude to Parisian French (the accent of an out-group) than to Canadian French.

The results contrast sharply with the self-downgrading and self-hatred results obtained with adults in Lambert et al. (1960), suggesting that the pre-pubescent children are still at a stage of development which involves rejection of out-groups, i.e. speakers of English: the invidious attitudes, prejudices and language stereotypes of the wider Canadian society outside the family circle have not yet impinged on the consciousness of the children. It is possible to view the rejection of the English speakers as simple xenophobia towards people speaking foreign languages, but the monolingual children are in a special position in the bi-cultural diglossic community of Montreal: for them English is not just some strange incomprehensible *foreign* language; they almost certainly recognise English when they hear it since it is their society's *second* language. Unfortunately this point is only mentioned in passing and not developed.

In a fourth study, Lambert et al. (1965), the evaluation of personality characteristics from matched-guise recordings of voices was extended to Arabic and two dialects of Hebrew (Yemenite and Ashkenazic), the setting being Israel where both language and dialect variations have important social and political implications. Jewish and Arab students in their mid-teens rated the guises for twenty traits on 6-point scales labelled from 'very much' to 'very little'. Four yes/no questions concerning social distance were also asked. The Jewish respondents rated the Yemenite Hebrew guise less favourably than the Ashkenazic Hebrew guise for intelligence, ambition, wealth, cleanliness, prestige, potential neighbour, honesty and reliability. However, the Yemenite guise was thought to have a greater sense of humour. The Jewish respondents also had a clear stereotype view of the Arabic speakers, downgrading them for humour, friendliness and honesty, and finding them less desirable as friends or marriage partners, and upgrading them (puzzlingly) for wealth, although this does not tally with the economic realities of Israel. The Arab respondents, in a mutually antagonistic manner, downgraded the Hebrew guises for intelligence, self-confidence, good-heartedness, friendliness, honesty, and desirability for marriage. Overall it was clear that the respondents were drawing on social and/or cultural stereotypes in

their attempts to gauge the speakers' personalities. Once again the fact that this was a bilingual study involving first and second languages, where the informants knew the two languages they heard were the two languages used in their country, is not stressed, and there is no mention of the fact that evaluation of a speaker speaking *unidentifiable* Arabic or Hebrew (i.e. an unknown *foreign* language) might produce entirely different results.

A fifth study, Lambert et al. (1966), took the preceding matched-guise studies of language, dialect and accent variation one step further. Three moderator variables were used: various age groups; monolinguals (French) and bilinguals (French/English); and two socio-economic backgrounds. The voices for rating were French-Canadian guises v. English-Canadian guises. Fifteen traits were presented with bi-polar adjectives on 5-point rating scales. The results suggested that listeners' social class is more important in determining attitudes towards the guises than their age or 'bilinguality' (*sic*). The upper middle-class informants of all ages (from 12 to 16/17 years) strongly favoured the English-Canadian guises rather than the French-Canadian guises (in-group downgrading), the bilingual informants being even more extreme than the monolinguals in favouring English-Canadian guises. The girls from lower class or lower middle-class families generally did not retain their less marked upgrading of one or the other guise beyond the age of 14. While the influence of age and social class on reactions is stressed, the wider implications of the fact that this study involves the guises of a society's recognisable first and second languages are again not highlighted.

A full exposition of the MGT and a concise summary of the initial studies are found in Lambert (1967). The author carefully places the work in the socio-psychological context of classroom foreign language learning/teaching, conscious language shifting and bilingualism. He clearly recognises the fact that certain features of speech style interact with characteristics of the *speaker* (gender, age, ethnicity, social-class background) to prompt reactions in listeners, and the equally important fact that the impression formed in the listener is further dependent on the *listener's* age, gender, ethnicity, socio-economic background, speech style and bilinguality. As a tool for measuring group biases in evaluative reactions, Lambert claims the technique is valuable and reliable, but he fails to disentangle the reactions of people living within a bicultural diglossic community towards the two languages in their community from the reactions of monolinguals living in a monocultural monolingual community towards unidentifiable foreign languages and accents.

Unfortunately Lambert's example and methodology concerning the MGT was not always followed with care. For example, Webster and Kramer (1968) re-examined evaluational reactions to French-Canadian accented English and American-accented English, but the method used, the informant sample and the analysis of the results is badly presented and obscure, seriously undermining the integrity of the technique.

Three matched-guise studies, modelled largely on the original study of Lambert et al. (1960) and the review of Lambert (1967), were soon conducted in Britain.

The first, Strongman and Woosley (1967), yielded inconclusive results. Two speakers (presumably male) produced two guises each, London and Yorkshire, and the recordings were presented to two groups of college-student listeners, one from the south of England, the other from the north, who rated each voice on 5-point scales for 18 bi-polar personality trait adjectives. Intergroup ratings revealed nothing of significance, i.e. both northern and southern respondents (with a common cultural background, and with no feeling of belonging to a majority or a minority group) shared the same stereotyped view more or less of each guise. For example, everyone agreed that Yorkshire speakers are more honest and reliable and less self-confident than Londoners. However, intragroup values showed that northerners judged Yorkshire speakers to be more industrious, and southerners judged them to be more serious, than London speakers. Northerners also judged Yorkshire speakers to be more generous, good-natured and kind-hearted than London speakers whom they rated to be more mean, irritable and hard. No marked favouring of one accent or the other was detected.

The second study, Cheyne (1970), compared Scottish and English regional accents (presumably educated Scottish and RP, but this is not stated explicitly). Four students (two men, two women) produced two versions each of a text, and so did four other speakers using their own voices (male/female, Scottish/English). A wide variety of listeners then rated the voices on 6-point scales across twenty-one traits. Generally the English guises and accents were upgraded for traits concerned with status, though Scottish male listeners favoured Scottish voices (ingroup support) by upgrading them on personality traits (friendliness, sense of humour, generosity, good-heartedness, nervousness, likeability). Clear stereotype reactions appeared, the stereotype dependent on the nationality of the listener and the perceived nationality of the voice. Interestingly the stereotype images provoked by a voice were not necessarily the same as those prompted by other stimuli, for the Scottish voices were not considered to betray the trait of meanness.

The third study, Giles (1970), investigated the attitudes of school children towards various accents of English, significantly adopting two contrasting techniques: matched-guise vocal stimuli, and written conceptual stimuli (i.e. accent labels). The study was at one and the same time a study of both native and foreign accents, since the guises to be rated involved one speaker producing—incredibly (and some people would claim impossibly)—*seven* regional or social British accents (RP, Affected RP, Birmingham, Cockney, Somerset, South Welsh, Yorkshire), plus *three* native but non-British accents (Indian, Irish, North American), and *three* non-native accents (Italian, German, French). For the non-aural evaluation, lists of sixteen accent labels were prepared, the correct thirteen accent labels minus Affected RP, but with Scottish, West Indian, Liverpool and 'an accent identical to your own' added. Both sets of stimuli were rated on Likert-type 7-point scales across three dimensions which mirror Osgood's EPA dimensions: (a) Aesthetic (How pleasant or unpleasant does the accent sound?); (b) Status (How much prestige or status is associated with speaking this accent?); (c) Communicative (How intelligible is the accent and how comfortable do you

feel interacting with the speaker of the accent?). The respondents were classified according to a 2 x 2 x 2 x 2 matrix of age (12-year-olds and 17-year-olds), gender, regional background (South West England and South Wales) and social class (working-class and middle-class parents, using job categories established by the Registrar-General's 'Classification of Occupations,' 1966; no more details given in the article's list of references. See OPCS 1980 and 1986).

The results, not surprisingly, show that the 17-year-olds are better at recognising accents than the 12-year-olds, and also suggest that the younger listeners have not yet assimilated the conventional social evaluations: 'the 17-year-old subject perceives significantly more prestige value in speaking RP than the younger child' (Giles 1970:219). A trend towards accent loyalty was clear, and it would seem that both sets of youngsters are unaware of the real broadness of their own accent, believing it to be less marked than it actually is. Gender differences were minor, restricted mainly to perception of French-accented and American-accented speech. Social class differences were also minor, consisting mainly of a strong tendency among the working-class sample to be more loyal towards their accent than the middle-class sample were to theirs.

Giles (1970) was followed by an unpublished (and impossible to get hold of) doctoral dissertation (Giles 1971a). Briefly described in Giles and Powesland (1975:32), it appears that the thesis extended the study of 1970 by introducing a third group of informants (21-year-olds), increased the variety of their regional backgrounds, and exploited a 'verbal-guise' (Cooper 1975) rather than a matched-guise technique, i.e. thirteen *different* speakers. While one may suspect the 'authenticity' of the thirteen accent guises produced by the one speaker in the 1970 MGT study, the use of different speakers in the thesis presumably resulted in the presence of uncontrolled intervening variables. The research designs of both these two studies are far from being robust.

The MGT was quickly and widely recognised as a useful and innovative new tool for investigating reactions towards accents, but several studies which did *not* use the technique continued to appear in the 1960s and 1970s. They are generally of poor design and of questionable validity.

For example, one study (Markel & Roblin 1965), which had speech content rather than accent as its focus, had a male student record three *different* passages of varying content. By adopting Osgood's semantic differential, it was found that the content of what a person says and the gender of the listener-judges had *no* significant effect on the Potency and Activity dimensions, but they *did* have an effect on the Evaluative dimension. The authors also separated voice set (features resulting from non-controllable physiological and physical peculiarities such as gender, age, health and location) from voice quality (qualities to some extent controllable, such as pitch range, articulation control and tempo, and appearing in a particular speech event), and suggested that whereas a general attitude towards a speaker may be determined by content and congruency of that content with voice set, specific impressions of a specific speaker are determined by that speaker's voice qualities. (Markel had in fact already established that inter-rater

and test–retest reliabilities for ratings of voice pitch, volume and tempo are remarkably high: Markel 1965.)

The semantic differential was also used in another study contrasting two different regional American accents, New York v. Buffalo (Markel et al. 1967). The broad conclusion was that regional dialect (*sic*) elicits stereotypical evaluations of speaker personality, but the research design is weak: six Buffalo speakers set against six New York speakers; no control for type and frequency of accent features; and no control for other 'linguistic and paralinguistic features' such as pitch range, tempo and volume. The validity of the experiment is in doubt.

The problem with these two Markel studies is obvious: the number of uncontrolled variables influencing the responses of the informants. However, this same weakness characterises another study which dropped the MGT and boldly resorted to separate voices representative of six specific dialects of American English: two white and four Negro (Tucker & Lambert 1969). It was found that respondents almost unanimously rated Network (national newscaster, white) English as the most favourable dialect, and that both northern white and southern Negro students placed Educated Negro Southern English in second place, but Southern white students placed Educated White Southern in second place (in-group preference). The dialect rated most unfavourably by both northern and southern white college students was 'Mississippi Peer' (southern Negro), and by southern Negro students was Educated White Southern (out-group rejection). While no evidence of gender differentiation was found, no reference is made to the possible interference of such variables as voice set, voice quality, style and tempo. There is no true discussion of methodological problems and principles.

Another five studies also outside the matched-guise framework concerned themselves with specific attitudes towards language and accent in schools: Frender et al. 1970; Williams 1970; Giles 1971d; Williams et al. 1971; Seligman et al. 1972. The disturbing results—again flawed by certain weaknesses in the research designs and methodologies—can be summarised broadly as finding that school teachers are prejudiced against non-standard accents, and unjustifiably upgrade standard-accented speakers for intelligence, skills and achievement.[4] People at the time were presumably prompted to draw the conclusion that if teachers have distinct (and pernicious) reactions to accents, then other groups of people whom one might expect to be less on their guard than teachers would yield equally startling results. In other words, these five studies prompted people to be more careful about defining their informant population: who are the informants? whose attitudes are being investigated? and where (in what situation, environment, setting) are these attitudes being expressed?

The introduction of the MGT undoubtedly boosted attempts to investigate attitudes towards accents, for the technique goes 'some way towards eliminating the effects of the more idiosyncratic features of speech such as rate, loudness, timbre, pitch and so forth' (Giles & Bourhis 1976:294). However, the technique did invite criticism.

---

[4] The implications are reviewed in Bradac & Giles (1991).

4.3. Criticism of the matched-guise technique

The evaluation of studies of attitudes towards accents prompted a polemic for and against the MGT, a whole series of discussions about theories, models and methodology, and a flow of valuable investigations quickly following one after the other.

Doubts were soon raised about the ability of speakers in MGT studies to hold variables (such as tempo, volume, pitch and voice quality) across guises, thus questioning the very *raison d'être* of the whole approach (Tajfel 1962). Disquiet about this very point continued to be expressed for many years, the thirteen guises in Giles (1970) certainly straining credibility, despite the author's reassuring comments as to their basic 'authenticity.' Occasionally authors clearly state that their test guises have been pre-tested for authenticity, but few admit to the fact that their particular guises are (or may be) caricatures. One exception is Delia (1972:288) where two of three guises were recognised as 'contrived' by native speakers of the dialect. Disbelief was also expressed that informants fail to recognise that the voices they hear are produced by one and the same speaker (Tajfel 1962), and it is indeed noteworthy that to avoid this potential detection of speaker duplication no one to my knowledge has ever advocated the presentation of the different guises for evaluation to different samples of people randomly selected from the same population. No pharmaceutical company would test two competing products on the same patients, so why *mutatis mutandis* do sociophoneticians?

However, the main attack at the beginning of the 1970s was aimed at the artificiality not of test stimuli, the guises, but of the test situation (e.g. Agheyisi & Fishman 1970; Lee 1971; Robinson 1972; Tajfel 1972). Informants are asked to make conscious value-judgements about the personality not of one *speaker* but of a whole series of disembodied *voices*, a task that no one ever does (it was claimed) in real life. Moreover, the methodology is flawed (allegedly) since attitudes towards a speaker are constantly changing and evolving under the barrage of different signals being received and processed during a communication event. In particular the accent of a speaker is fluid and unstable, for discourse shifts and style shifts are a normal and inevitable part of conversation (a point repeated in Coupland 1985). No one ever speaks with a regular unchanging accent, not even in a monologue situation (such as reading the news on radio or preaching from a pulpit). Hence, one particular accent, whether it be regional, stylistic or situational, is a continuum which speakers are forever sliding along, to and fro, mild–broad, formal–informal, depending on a wealth of interruptions, unpredictable events and emotional swings. Thus the attitudes of listeners towards speakers when hearing only necessarily short recordings of the voices in artificial anaemic situations is utterly unrealistic: real-life situations are volatile; and real-life accents are not static. The repetition of the same short message also inevitably distorts evaluations and reactions: a listener's attention is pushed unnaturally away from the message content and its pragmatic significance towards the vocal variation. Worse still, a bias is engendered by the previously-established evaluative set (Giles & Bourhis 1976).

The apparent disregard for the fact that communication is a composite of verbal and non-verbal elements vexed some people, and so, apparently to redress the balance of purely vocal matched-guise experiments, tests were conducted on the interaction of verbal signals with visual signals (such as head-nods, facial expression, gesture, posture and eye-movements) as conveyed on video recordings (Argyle et al. 1970; Williams et al. 1971). However, concern continued to be expressed about the 'unmentioned complications' of non-verbal behaviour, social context, and the experiment situation:

> When an Englishman switches to weak French and candlelight, his dining companion will be surprised if the behaviours manifested are identical to those he displays speaking English while drinking tea with the vicar's wife. . . . Personality *per se* cannot change with language switching, but the traits likely to be manifested do. (Robinson 1972: 98–100)

Meanwhile, Tajfel (1972) famously claimed that since life does not take place in a vacuum, one should not conduct experiments in a vacuum. And Lee (1971), in a biting attack upon the artificiality of person perception experiments using tape-recorded voices, concluded that:

> Interpretation of the research in dialect perception has ranged beyond the limits of its data. . . . Improvements must be made in elicitation and presentation techniques, the design of instruments and the statistical treatment of data. (Lee 1971:410, 417)

Giles and Bourhis (1973) rose to the bait and wrote a succinct carefully reasoned rebuttal. They took care to focus on Lee's (1971) four main concerns:

> 1) repetitive, content-controlled speech stimuli,
> 2) the lack of a social context for these stimuli,
> 3) the selection of the dependent measures, and
> 4) the reliability of these dependent measures. (Giles & Bourhis 1973:337)

Basically the thrust of the paper urges Lee to re-read the literature and not to be blindly dismissive. However, one point concerning dependent measures (i.e. the adjective labels that are chosen) stands out: the authors claim that although a research paper may not state the fact explicitly, pilot studies are generally held to determine which adjective labels are potentially salient for the particular study in hand: 'traits are not chosen arbitrarily, but are those that subjects most frequently find pertinent to the evaluative situation' (Giles & Bourhis 1973:340). Whether we can (or should) trust Giles and Bourhis on this point or not, the fact remains that a trait may inadvertently be omitted from the list even though it is pertinent: the choice of adjective pairs may be defective, and hence the measuring instrument may be intrinsically unable to measure and/or discover what it is supposed to, may be less revealing than it could be.

The authors also stress a further theoretical point: the adjectives (like 'considerate') and terms (like 'faith in God') used in matched-guise studies should not be understood to mean that the voice *per se* sounds 'considerate' or whatever. One cannot answer the question: 'What does faith in God sound like?' The voice merely provides the listener with clues as to the speaker's social group membership, and the clues activate stereotype imaging. The voice does not somehow magically reflect the trait itself. Curiously the authors do not seem to consider the

possibility that a voice may be so bland, and prompt so few cues as to social group membership, that stereotyping occurs only with the broadest of brush-strokes if at all.

Giles returns to a defence of the MGT with a meticulously designed study in Giles et al. (1975). The matched-guise procedure is used not with tape-recordings, but with a *live* bidialectal speaker talking with an RP or a Birmingham accent face-to-face with two separate groups of 17–18-year-old listeners. The stimulus speaker requested the listeners to perform a written exercise, and the length of responses was taken as an indicator of compliance and willingness to provide information for a 'standard' and 'non-standard' accented speaker.[5] The speaker then left the room and the listeners, with no prior evaluative set, were unexpectedly required to evaluate the now absent speaker's personality on 9-point rating scales for eighteen bi-polar adjective pairs. Whatever one's qualms about the uncontrolled variables inevitably introduced by the live performances of the speaker, the results show that significantly more information was provided when the request was made in RP than in the Birmingham guise, but that the RP guise was upgraded only (surprisingly) for the trait of intelligence. The discussion stresses the need for more scrupulous attention to be attached to the design of attitude studies and for more emphasis to be placed on 'the role of setting, topic, type of audience and so forth' (P. 79).

A further ingenious attempt to respond to criticism of the MGT (and which also seems to acknowledge the conative component of an attitude) exploited the natural setting of a theatre in Cardiff: listeners were unaware that they were participating in an experiment (and thus had no prior evaluative set), and the response to a stimulus voice (with no other cues at all) was measured in terms of co-operative altruistic behaviour (Bourhis & Giles 1976). The stimulus consisted of recordings of a bilingual Welsh/English male speaker making an announcement over the theatre's public-address system at the beginning of an interval in four different guises: RP, broad South Welsh-accented English, mild South Welsh-accented English, and standard Welsh. The informants were the audiences on various different evenings, thought to be either Anglo-Welsh (those attending films in English) or bilingual Welsh (those attending a play staged in Welsh). The dependent variable was the number of questionnaire forms completed by the audience in response to the announcement, the questionnaire being an authentic audience appreciation survey and preference enquiry for future programmes compiled in collaboration with the theatre management: this was no 'experiment in a vacuum' (P. 16).

The results showed that co-operative behaviour was a function of speech style. The co-operation of the Anglo-Welsh audience was about 25% when the plea was in RP or mild Welsh-accented English, but only 8% when it was in broad Welsh-accented English. The response of the bilingual Welsh audience showed an almost steady cline starting from practically non-co-operation when the

---

[5] The terms 'standard' and 'non-standard' are the terms used by Giles and his colleagues though it is perhaps worth noting that their use of the word non-standard for the Birmingham guise does not seem to suggest *sub*-standard. Knowles (1987:3) stresses this distinction.

request was in RP (2.5%), rising through broad and then mild Welsh-accented English, to more willing co-operation when it was in Welsh (26%). The response of the Anglo-Welsh suggests, according to the authors, that there is probably a spin-off between the formality of the setting and the perceived formality of accent, the broad Welsh accent being heard as more informal and therefore less appropriate for the task in hand than the milder counterpart. As for the bilingual Welsh listeners, presumably their Welsh language is such a salient dimension of their identity that they virtually ignored the RP-accented plea: the accent of the linguistic (and national) out-group carries little persuasive power. The authors concluded by saying:

> The importance of the status or the ethnicity of a speaker as symbolized by his speech style may be attenuated or accentuated depending on the purpose, setting, and topic of a particular interpersonal or intergroup exchange. . . . Studies using the matched-guise technique must now be carried out in a *range* of more dynamic, interactive settings. (Bourhis & Giles 1976:16)

An almost identical authentic audience appreciation survey has been performed in Denmark, in a five-screen cinema complex, with four different Danish accents (Kristiansen & Giles 1992). The results revealed that the Standard Danish accent (as opposed to mild Zealand, broad Zealand and Copenhagen accents) was overall the most successful in eliciting co-operative behaviour. However, the audience labelled as the 'intelligentsia' tolerated all accents, while young people still at school practically failed to respond to all accents except Standard Danish. The authors admit that the research design fails to exclude many potentially confounding variables, but stress that the dependent variable is a behavioural response incorporated within a natural setting.

One major criticism of the MGT which latches onto the dual problem of measuring instruments and of the transience of attitudes long recognised by socio-psychologists is the fact that attitudes are not by nature stable. They are variable. They are not quantifiable by any common measuring instrument or standard but are forced into the straightjacket of a scale designed by experimenters. It has been claimed, therefore, that since attitudes are not viscous mental entities existing in the minds of people but are stances continuously being constructed and reconstructed in discourse, it would be more fruitful to investigate attitudes towards language with discourse-analytic techniques and the qualitative analysis of naturally occurring talk, conversations and writing (Potter & Wetherell 1987; Hyrkstedt & Kalaja 1998). While this argument would appear to point the way back to an adoption of one of the two simpler methods of studying attitudes towards language I mentioned earlier (P. 61), it remains to be seen whether such a discourse analytic technique will dispense with the MGT altogether.

The transience of attitudes and their potentially ephemeral nature also underpins the criticism that the reliability of many experiments using the MGT is questionable whenever conclusions are drawn, and generalisations made, about the impact of an accent where only one instantiation of that variable is witnessed (Cargile & Giles 1997:200). Take the Welsh theatre study and the Danish cinema

study: would the same results have been repeated on a different day of the week, with a different audience come to watch a different programme? Moreover, what was the population being targeted, and in what sense were the informants a random sample of it? Several studies are needed with several speakers producing the same combinations of guises, and with several population samples, in order to obtain corroborative results. Such confirmatory studies would justify the fundamental assumption of studies of attitudes towards language, viz. that there is a homogeneous set of shared attitudes towards language throughout a speech community (Starks & Paltridge 1996:218) and that the attitudes are relatively stable. Despite this exhortation for supportive studies, the body of MGT research carried out so far is remarkably consistent in its findings: there is not a mass of contradictory results but a solid accumulation of data all pointing to the same general conclusions (Honey 1989:58). Nevertheless, restraint, caution and circumspection are still needed when making generalisations, and the target population, sampling and statistical methods must not be neglected if experiments are not to degenerate into anecdotal exercises.

One further methodological weakness has persisted: the disregard for speech diversity:

> This technique . . . has not taken into account the social meaning of people's ability in many contexts, and at many linguistic levels, to modify their speech styles. (Giles & Bourhis 1976:297).
>
> An individual will use different pronunciation patterns . . . in different circumstances. (Wells 1982:25)

The way in which the MGT has generally been used has implied that individuals and populations have only one language variety at their disposal and possess no speech style repertoire. This is obviously not the case for native speakers/ listeners. In normal and dynamic social situations people not only adapt their speech in order to project the self-image they wish their interlocutor(s) to see, but they also 'react,' either converging in terms of dialect, accent, utterance length, speech rate, volume and so on, or occasionally diverging. Studies using the MGT make no allowance for these phenomena. More seriously, most authors conspicuously fail to recognise the fact that *non*-native speakers/listeners do *not* have a range of language varieties at their disposal, cannot vary their speech at will, consciously, in order to cause a particular effect or to respond to a particular cue, and cannot easily identify and interpret shifts in the speech of native speakers.

4.4. Theorising prompted by the matched-guise technique

Criticisms of the MGT has led not only to more carefully constructed experiments but also to theorising about the phenomenon of distinctive reactions to accents, the origins of accent prestige and accent loyalty, and the processes underlying accent accommodation and language shifts. Theories have also been presented to explain the mutual influence of interactants in communication, introducing the ideas of speech style accommodation, speech divergence and convergence, integrative and dissociative speech style selection, upward and downward speech style shifts, and difference versus deficit in non-standard usage. Discussions have crossed the borders of sociolinguistics, sociology and

psychology, and many experiments have been conducted to test the various hypotheses presented. However, the implications for second-language speakers, foreign-language teaching and acculturation have not always been recognised.

One puzzle which quickly received attention was the fact that some accents are afforded more prestige than others. That such prestige exists (RP, for example, clearly holding supremacy over other British accents, Parisian French over Canadian French, standard American over Black English, standard Danish over non-standard Danish, classical Arabic over colloquial Arabic [El-Dash & Tucker 1975]) has been amply demonstrated, and attempts to explain this prestige phenomenon soon led to two opposing hypotheses: the **Inherent Value Hypothesis**, and the **Imposed Norm Hypothesis**. The first claims that an accent rises to a level of prestige simply because it is the 'best' accent: unlike other accents, its sophistication is such that it can satisfactorily perform all the demands placed upon it; it is aesthetically pleasing; it succeeds. However, the second claims that an accent attains widespread respect and status only because of chance extra-linguistic factors affecting society and the march of history (e.g. the rise of royal courts of patronage, of economic power bases): the accent accedes to the prestige which is accorded to its users simply because of association.

Some apologists took a strong social and/or educational bias, claimed absence of prestige equated with 'linguistic deprivation,' and advocated 'language extension' programmes for the so-called linguistically deprived. Thus 'deficit' theorists, pointing out that lower-class children who use non-standard speech generally do badly at school, argued that non-standard speech is inferior, linguistically underdeveloped and deficient; and 'difference' theorists, stressing that middle-class teachers do not use and cannot understand non-standard speech, argued that non-standard speech is not inadequate, merely different.

Giles recognised how the deficit theory supports the inherent value hypothesis, and how the difference theory tallies with the imposed norm hypothesis, and set about trying to obtain empirical evidence which would bolster one or the other position.

One MGT study had Welsh listeners evaluate standard and non-standard accents of a language they did not know (French, but presented with six other 'filler' languages) on the assumption that:

> If listeners could differentiate speech styles in terms of perceived pleasantness and status in the same direction as speakers of this language, then this would, to a certain extent, substantiate the INHERENT VALUE hypothesis. However, if non-native listeners could not differentiate the speech styles along these dimensions then the data would be in line with the IMPOSED NORM hypothesis. (Giles et al. 1979:592)[6]

The results showed that the informants failed to differentiate between the French accents in the same way as French Canadians do (Lambert et al. 1960; d'Anglejan & Tucker 1973; Bourhis et al. 1975). They bestowed no more prestige to one guise or another, nor more favourable personality characteristics to

[6] This article, despite its date, is a congress paper presented in 1973.

one speaker or another. The imposed norm hypothesis and the difference theory appeared to be upheld.

A second MGT study conducted to test the same hypotheses used two varieties of Greek: Athenian and Cretan (Giles et al. 1974).[7] British undergraduates, with no knowledge of the Greek language, listened to the two matched guises plus four non-Greek 'fillers.' The results revealed no significant differences in the ratings of the two unidentified Greek guises, thus lending further support to the imposed norm hypothesis and the difference theory.

More studies using languages that listeners cannot identify at all are needed for corroborative purposes, but curiously no one has yet called for parallel studies to explore how listeners react to various non-identifiable foreign accents of their (the listeners') own mother tongue.

Explanations were also sought for why it is that speakers of standard accents are generally found to be upgraded for competence and confidence (e.g. Sebastian & Ryan 1985) and for substance and authority in their messages (e.g. Giles 1973b; Giles & Powesland 1975; Giles & Sassoon 1983), whereas speakers of non-standard accents are upgraded for integrity and social attractiveness. One suggestion is that this dichotomy results from the widespread preconception that the standard speaker is committed to personal social advancement and socio-economic improvement, and so lacks the honesty and human warmth of the non-standard speaker more committed to community-oriented values. Thus non-standard speakers are afforded low status, but if their accent resists the attraction of the standard and persists, then it must nevertheless attract 'covert prestige' (Trudgill 1972; Ryan 1979). The respect for the standard accent is a 'real' or 'dominant' attitude, 'flaunted' and 'tolerated' openly in public (Kristiansen & Giles 1992), but the stubborn retention of a non-standard accent reveals a private attitude, negative and shameful, good enough only for the home, hidden from public scrutiny. The curious thing is that when a non-standard variety starts to spread (as the Copenhagen accent is), the accent increasingly being accepted and used in public, then covert prestige has turned into overt prestige.

Another issue which exercised people's minds was the attempt to provide (and test) descriptive models for the contextual and interpersonal constraints affecting phonological usage, i.e. for people's accent shifts within their accent repertoires. Studies soon became divided into 'attitude and behaviour static research' and 'attitude and behaviour dynamic research' (Giles & Bourhis 1976). The speech accommodation processes in the latter were divided into examples of *convergence* and *divergence*, upwards or downwards in terms of prestige (Giles 1973a), but it was argued that such **accent mobility** is constrained by receiver attributes, situational determinants, social attitudes (e.g. ethnocentricity), cognitive styles (e.g. field independence, where a person may maintain attention to particular parts of a perceived world regardless of confusing surroundings) and even 'empathic' capacity. By standardisation of pronunciation (i.e. a move away

[7] The publication date of this paper is 1974, and so *follows* the congress presentation of Giles, Bourhis and Davies (1979). However, the actual experiments reported are probably contemporaneous.

from a basilect towards the acrolect), speakers increase their perceived status and competence, and the perceived quality of their message and its persuasiveness, in the presence of an RP receiver. The importance of social context in the speech accommodation process has been widely recognised, but the fact that accommodation (whether competition or co-operation) in dyads is not necessarily the same process with the same results as accommodation (usually conformity) in small groups (Tajfel 1972) has frequently been overlooked: no allowance is usually made for the phenomenon of 'risky shift,' where groups generally favour riskier courses of action than individuals, or conversely where individuals are more conservative than groups.

These theoretical models of accent shifts in interactive dynamic social speech led to the hypotheses that convergent and divergent speech accommodation would occur both within and *between* language (ethnolinguistic) communities, and that the greater the perceived effort put into accommodation, the more favourable the impression formed by the listener and the more willing the listener would be to reciprocate with similar accommodation.

One study focusing on the bicultural setting of French Canada proved that this was indeed the case: a bilingual French Canadian (FC) described a picture to bilingual English Canadians (ECs) in French (no shift), in a mix of French and English (partial shift), in fluent English but with an FC accent (full shift), and in non-fluent English. When EC students drew the described picture and rated the speaker for various traits, it was found that 'a speaker's perceived considerateness and effort in bridging the cultural gap are a function of his perceived effort in accommodation' (Giles et al. 1973:183), and the perceived effort to accommodate provoked a welter of accommodation–non-accommodation tactics on the part of the listeners. It was deduced that there is a continuum of strategies available for people in different situations, though no explanations were put forward as to why one speaker should choose one level of reciprocal accommodation and another speaker another. However, a *quid pro quo* arrangement would seem to be the norm: if listeners suspect that speakers of another ethnolinguistic group are accommodating for them, then they themselves are more disposed to accommodate in turn. The authors recognise the complexity of speech accommodation processes in a bilingual society, in intermixing ethnolinguistic communities, especially where 'imbalanced bilinguals' converse in their first and/or *second* languages, but they make no reference to situations in a monolingual/monocultural society where a native speaker of the society's language converses with a foreigner speaking in a *foreign* language.

Effort, however, contrasts with motives: it was necessary 'to determine whether accommodation is differentially perceived and is consequently distinct depending on the motives attributed to it by the listener' (Simard et al. 1976: 376). Is there an interaction between the *effort* expended by speakers in accommodating their speech style, the perceived *ability* of speakers to accommodate their speech style and the perceived external *pressure* to do so (the last two conspicuously absent from the experiment in Giles et al. 1973)? It is this interaction

that results in listeners adopting accommodative or non-accommodative speech styles.[8]

The idea was tested in Simard et al. (1976) with French Canadian listeners knowing or not knowing (a) the language ability of an English Canadian speaker, (b) the external pressure put on this speaker to switch to faulty English-accented French, and (c) the giving or withholding of permission to answer in English. The dependent variables included number of errors made in interpreting the message, and choice of language for returning the communication. The results, bolstered up with theories of attribution processes, cognitive consistency and cognitive dissonance (Festinger 1957), led the authors to claim that perceived non-accommodation is likely to foster a relatively unfavourable evaluation of the speaker and induce listeners to maintain their regular speech style. However, if linguistic inability or external pressure is known to be, or suspected of being, the cause for non-accommodation, then the reaction is still negative and unhelpful but less so. Moreover, if convergence is attributed by listeners to effort rather than external pressure, then the message is understood more clearly and accurately, and reciprocal convergent accommodation is more likely. The important accommodation model which is presented (a revised diagram appears in Giles & Powesland 1975:164),[9] depicting a speaker's speech act and subsequent listeners' perceptions and responses, at last takes some account of foreign-accented speech, for it recognises that listeners may perceive a speaker's non-accommodation as resulting from inability to accommodate.

This paper opened the way for people to focus attention not just on the speech accommodation phenomena of interacting bilinguals and diglossic/bidialectal monolinguals, but also on how listeners perceive foreign-accented speech, on how listeners respond to the non-accommodation of non-native speakers, on models of the speech behaviour of interacting native and non-native speakers. However, the notion of linguistic inability was not elaborated further, and the obvious remained unsaid: foreign learners of a language are generally incapable of accommodating their speech style on the segmental pronunciation level (whether consciously or unconsciously), and have only restricted abilities across the remaining gamut of phonetic speech resources such as pause and vocalisation durations, speech rate, vocal pitch and vocal intensity (Street & Giles 1982:193). Non-native speaker non-accommodation on at least the phonemic level is non-remediable. Alert learners may be able to pronounce a foreign language better than tired learners, but it is axiomatic that few can shift easily (if at all) along the continuum of accurate–inaccurate (native-like–foreign-accented) pronunciation. Speech accommodation, whether divergent or convergent, is essentially a discourse strategy which is available only to native speakers, and non-native speakers are obliged to resort (if they can) to other compensatory strategies such

---

[8] This hypothesis owes much to theories of interpersonal accommodation and of causal attribution (Kelley 1973; see also Kelley & Michela 1980). However, there is a serious typing mistake in Simard et al. (1976), occurring at least three times: Kelley's *causal* attribution theory is referred to as *casual* attribution theory (P. 376, P. 383). The reader needs to beware.

[9] The revised diagram in Giles and Powesland (1975) was published *before* the article of Simard et al. (1976). Hence the apparently impossible publication dates.

as avoidance, paraphrase, conscious transfer, appeals for assistance and mime (Tarone 1977). Thus foreign speakers lack one of the tools for producing and conveying the image, affect and message which they may wish and intend, and the listener's response rests that much more beyond their control, influence and manipulation. Non-accommodation may mistakenly be perceived as negative intent, as negative disposition, and the response, based on a false interpretation of the speakers' *actual* intent, may therefore be equally negative. Lack of ability, therefore, when interpreted as lack of effort, may lead to dire consequences in terms of interpersonal attraction, mutual intelligibility and social approval.

Speech accommodation theory (SAT) was further refined in Giles and Smith (1979), Giles (1980), Street and Giles (1982), Thakerar et al. (1982) and Beebe and Giles (1984),[10] but all these studies would seem to take as their starting-point the premise that speech accommodation is—to quote one from them all—

> a linguistic strategy whereby individuals adapt to *each other's* speech by means of a wide range of linguistic features, including speech rates, pause and utterance lengths, pronunciation, etc. (Thakerar et al. 1982:207. My emphasis.)

The basic postulate is that:

> communicators are motivated *to adjust their speech styles with respect to one another* as a means of expressing values, attitudes and intentions. (Street & Giles 1982:205. My emphasis.)

This stance is inappropriate, I would claim, for a dyad of native speaker and non-native speaker: in such a situation the accommodation is one-sided or at best lop-sided, since non-native speakers are *ipso facto* underachievers (even non-performers) in the accommodation of speech style. For the foreign language speaker convergence is not an option, and divergence may mean renouncing the conversation by reverting to the L1. Native speaker *linguistic* accommodation, both convergent and divergent, can rarely be reciprocated by the non-native speaker: the accommodation is not mutual; it's not reflexive; there's no give and take. The learner's desire to accommodate may be present, may indeed be strong, because the social motivation is strong, but this psychological dimension is not enough to provide the linguistic tools necessary for linguistic accommodation (see Thakerar et al. 1982:221–2).

Although this deficit in the linguistic abilities of foreign language speakers is not generally conceded, Beebe and Giles are blunt in admitting 'the embryonic status of SAT' (1984:10). They claim the terms 'convergence' and 'divergence' are 'far too imprecise linguistically' (*ibid.*), and hence they repeat Street and Giles's (1982:218–20) call for more empirical research into the multiple linguistic levels of accommodation, with a focus both on which speech features are *accommodated* (or not), when, how and why, and on which speech features are *reacted to* (or not), when, how and why. In addition the idea of positive v. negative reaction needs clarifying, as also the fact of actual v. perceived accom-

---

[10] Convergence was variously labelled or associated with: complementarity, dissociation, compensation, congruence, reciprocity, synchrony, symmetry and pattern matching. Theories of mutual influence in communication set the speech accommodation model against models such as the activation model, the communication model and the discrepancy-arousal model. See Street & Giles (1982).

modation. Furthermore, even the identification and labelling of a linguistic shift as accommodation is problematic (see also Meyerhoff 1998) since some other influence may be at work: a broader theoretical base is needed to furnish us with 'an "interpretive" framework . . . for understanding the complex interweaving of different communication strategies for normative, self-presentational, negotiative and accommodative (among other) purposes' (Beebe & Giles 1984:11). Moreover, general awareness of the strategies people use (or do not use) in accommodating their speech style remains largely unmapped.[11]

An extension to SAT is **ethnolinguistic-identity theory** (ELIT), an attempt to specify the conditions necessary for ethnic speech divergence, to identify the speech strategies and the language attitudes of speaker groups, and to characterise the sociostructural/cultural pressures exerted on these groups (Beebe & Giles 1984). Despite the enormous problems of generalisation from one ethnic context to another, ELIT aims to provide a social psychological theory of language and ethnicity (see Giles et al. 1977). In addition, because of its focus on principles of ethnolinguistic identity, ELIT has important implications for the intergroup theory of second-language acquisition (SLA) (Giles & Byrne 1982) for it helps explain the selection of integrative or dissociative speech styles. However, in the discussion of the range of factors affecting SLA, it is clear that Beebe and Giles do *not* distinguish between *second* language and *foreign* language acquisition. This is regrettable since the learning of English, for example, by the Spanish-speaking teenage child of a Spanish-speaking family living in Los Angeles is not comparable to the learning of Tibetan by the Gaelic-speaking teenage child of a Gaelic-speaking family living in the Outer Hebrides. This oversight is all the more surprising as Giles and Byrne (1982:37) stressed the distinction between second and foreign language learning. A theory of speech accommodation, I would argue, cannot be the same for the second-language learner and for the foreign-language learner. There will be areas of overlap, but many factors will be different; and the same applies for theories of second-language acquisition and foreign-language acquisition.

As more and more empirical studies were conducted, so there was a growing recognition of the enormous complexity of the phenomenon referred to as 'language attitudes.' Cargile et al. (1994) lists the daunting wealth of factors which influence the language a speaker uses, the nature of a listener's attitudes towards that language, and the outcomes (in terms of evaluations, communicative strategies and behaviours) of those attitudes. In addition the authors stress that attitude formation in conversational interactions is volatile, that the variables affecting attitudes are reciprocal, symbiotic and recursive. Attitudes are not the outcome of a simple static input–output mechanism. To understand attitudes we need to understand the social dynamic processes involved in their formation, their motivational and affective constituents, the situations and interpersonal

[11] Meyerhoff (1998) uses the image of 'mapping' to explain the convergence and divergence strategies of accommodation: a speaker's personal and group identities, all interacting and non-discrete, influence the linguistic behaviour. Milroy (1987a, 1987b) talks of '(social) networks.'

relationships behind them. This is as true for native accents of a language as for foreign accents of it.

4.5. Refinements to the matched-guise technique

The combination of direct criticism following the growing awareness of the failings and short-comings of some studies, and the appearance of theories of language interaction, have prompted people to be more careful in the design of their empirical experiments. The abundance of features influencing the perception of speakers is widely recognised:

> It is not unreasonable to assume that all available levels of linguistic expression can and will be used by hearers as diagnostic of the speaker's identity. This will include, among other things, lexical choice, morphological and syntactic variation, phonetic/phonological variation and tempo, in addition to paralinguistic features such as voice quality. . . . Stereotyped features do not occur in a linguistic vacuum; they occur in context with a whole range of other features. (Romaine 1980b:229)

Clearly no study can embrace and control all of the relevant factors, but attempts have been made to focus on various aspects in turn, on characteristics of the voice and on those of the listener.

4.5.1. Characteristics of stimulus voice

The characteristics of the stimulus voice(s) which have received most attention are speaker gender, speaker age, the precise phonetic description of the accent, voice set, tempo, overall pitch and pitch range, voice quality, speaker's presumed job and social class, context and speech content.

4.5.1.1. Gender of voice

It was surprisingly common—until perhaps about the late 1970s—for linguists to neglect gender as a variable in their studies, to fail to recognise that language use and language processing may be affected by the gender of both speaker and listener. For example, in 1972 Labov had no qualms in writing:

> There is a growing realisation that the basis of intersubjective knowledge in linguistics must be found in speech-language as it is used in everyday life by members of the social order, in that vehicle of communication in which *they argue with their wives*, joke with their friends, and deceive their enemies. (Labov 1972:13. My emphasis.)

Even Bourhis and Giles's (1976) otherwise excellent study with Cardiff theatre audiences (see above, P. 70) betrays male chauvinism and blindness to gender differences in language use, for the authors talk of 'the status or the ethnicity of a speaker as symbolized by *his* speech style' (P. 16. My emphasis). The subjects of linguistic investigation are clearly assumed to be male: women's use of language is of no interest, is merely identical to that of men. This delusion was symptomatic of the male-dominated times: even TV news directors firmly believed—erroneously—that audiences preferred the sound of a male newscaster to a female one (Stone 1973). Significantly hardly any of the initial studies of speaker perception utilised female voices (Anisfeld & Lambert 1964 and Cheyne 1970 are rare exceptions): see Table 4.1 below. However, gender blindness was

Table 4.1 Gender of voices used in various studies

Gender of voice(s)	Study
Male	Strongman & Woosley 1967[a]; Giles 1970; Giles 1972b; Giles 1973b; d'Anglejan & Tucker 1973; Wölck 1973; Aboud et al. 1974; Bourhis et al. 1975[b]; Brennan et al. 1975; Ryan & Carranza 1975; Ryan et al. 1975; Bourhis & Giles 1976; Ryan et al. 1977; Apple et al. 1979; Brown 1980; Fielding & Evered 1980; Giles, Smith, Ford et al. 1980; Ryan & Sebastian 1980; Brennan & Brennan 1981a, 1981b; Thakerar & Giles 1981; Ryan & Bulik 1982; Stewart & Ryan 1982; Giles et al. 1983; Giles & Sassoon 1983; Seggie 1983; Paltridge & Giles 1984; Anderson-Hsieh & Koehler 1988; Gordon & Abell 1990; Giles et al. 1995; Cargile & Giles 1997
Female	Anisfeld & Lambert 1964; Cheyne 1970; Bourhis et al. 1975[b]; Elyan et al. 1978; Giles, Smith, Browne et al. 1980; Ball et al. 1982
Male and Female	Mulac et al. 1974; Giles & Marsh 1979; Romaine 1980b; Berry 1992

[a] Presumed gender of voice(s) is male.
[b] 'Study II' uses one male speaker, 'Study I' one female speaker.

eventually diagnosed, rightly deplored (Henton 1989), and a re-evaluation of basic gender premises was prescribed.

Studies quickly demonstrated that people harbour clear stereotypes of male and female speech behaviour, embracing both subject matter, content and manner of delivery. These stereotypes are firmly rooted in people's minds, being 'part of our social heritage' (Kramer 1977:159), though the stereotype characteristics might not of course correspond to any actual differences in men's and women's speech. However, with increasingly more scrutiny being paid to the perception of accented speech in men as opposed to women, it was deemed important to ascertain whether the upgrading of competence and socio-economic status (intelligence, ability, status, etc.) in men with a high-prestige accent, and the upgrading of integrity and social attractiveness (friendly, trustworthy, honest, etc.) in men with a low-prestige accent (see *inter alia* Strongman & Woosley 1967; Cheyne 1970; Giles 1971b, 1971c, 1972a; Bourhis et al. 1975; Elyan et al. 1978; Giles et al. 1981; Edwards 1982), are replicated in women.

Various studies confirmed that women are evaluated broadly in the same way as men. For example, in one study of RP and Lancashire accents in women alone, the RP guises were upgraded for competence (intelligence, self-confidence) and communicative skills (fluency, clarity), but downgraded for social attractiveness and personal integrity; by contrast the Lancashire guises were upgraded for sincerity and likeability, downgraded for aggressiveness and egotism (Elyan et al. 1978). In addition, however, the RP female guises were upgraded on certain occupational dimensions (the likelihood of having a job which was well-paid and prestigious), and upgraded on the likelihood of having an egalitarian relationship with their spouse and of being childless. At the same time they were thought to be more masculine on certain recognised male traits (adventurousness, indepen-

dence, self-esteem, strength) even though, curiously, they were upgraded for femininity. This last finding provoked the contentious conclusion that 'the female RP accent [is] the voice of perceived androgyny' (P. 130), a conclusion that tore the blinkers of gender blindness from many people's eyes.

Attempts to confirm this female androgyny finding with RP set against a Welsh accent, and with a check to see whether the *male* RP voice is also perceived androgynously, proved equally startling: 'RP speakers, irrespective of their sex, were rated [by both men and women] as more competent, liberated and masculine . . . than South Welsh speakers' (Giles & Marsh 1979:307). However, unlike the first study, there was no relative upgrading of the female RP guises for femininity (the original source of the androgyny claim). Moreover, confirming the previous study, the non-prestige accent (South Welsh here) was *not* found to be associated with masculinity, and was *not* upgraded relative to RP for social attractiveness. The issue of whether androgyny is a general feature of both male and female RP speakers remained unsettled though intriguing.

Certain asymmetries in the data of these two 'androgyny' studies prompted suspicions that the key factor in the perception of androgyny in RP-accented women might be social class rather than gender since 'the "masculine" items chosen to characterize female RP speech might also be middle-class stereotypical' (Giles, Smith, Ford et al. 1980:263–4). A re-analysis of the data with a list of adjectives carefully vetted for being typically male, or female, or middle-class, or working-class revealed that the female androgyny interpretation indeed stemmed from the use of adjectives which unwittingly had not just masculine but also middle-class connotations. This re-reading was confirmed with evidence from RP and Yorkshire guises also rated with carefully chosen adjectives: the male and female guises were upgraded on sex-appropriate adjectives; the RP guises were upgraded on middle-class adjectives; and the Yorkshire guises were upgraded on working-class adjectives. Speaker gender appeared to be less salient than accent and inferred socio-economic status.

The concern for gender bias even led to a re-evaluation of a finding that female accents may disclose the pro-feminist views of the speaker (Giles, Smith, Browne et al. 1980): can listeners spot pro-feminist views in *male* voices? An experiment with male voices revealed that

> 'the voice of feminism' . . . is not unique to women who are attempting to redefine their status in society. . . . The so-called 'feminist' voice is more likely to be causally linked to a more broadly-based liberal ideology than to feminism per se. (Giles, Smith, Ford et al. 1980:272–3)

The wider implications of this unforeseen influence of voice gender on people's reactions to speech is clear: not just gender but also other unexpected variables such as class, age, ethnicity and economic background—and even the choice of adjectives for the measuring instrument—may be hidden factors distorting our understanding of a situation.

From the 1970's onwards voice gender has generally been recognised as an important factor influencing people's attitudes towards accent. Nevertheless, hardly any investigations have yet been conducted into the perception of speakers

whose biological gender is different from their professed (or suppressed) gender identity or gender orientation. This is surprising in view of the number of cross-gender and homosexual roles and drag-queens that have been portrayed on the screen in recent years (e.g. *Mrs Doubtfire, Tootsie, The Crying Game, In and Out, Bound, Priscilla Queen of the Desert, Victor–Victoria*, Danny la Rue, Dame Edna). Some work has recently been done with the characteristics of 'the gay voice' (e.g. Gaudio 1994), and it has also been established that the perception of speakers prompted by vocal cues conforms to some extent to speakers' own perception of their gender identities (Smith 1980). However, there would appear to be a huge area here which remains uncharted. New work may in fact put in question the validity of some studies undertaken so far: the identification of factors—such as age, socio-economic class, subject matter, paralinguistic features—said to have influenced the perceptions of speakers may have been mis-identification or distortion arising from the unsuspected interfering factor of the speaker's gender identity and orientation.

4.5.1.2. Age of voice

Teenage speakers, student speakers and adult speakers are variously used as stimuli voices in matched-guise studies. However, few studies include speaker age as an independent variable, or examine how speaker age interacts with other variables to form listeners' attitudes toward the speaker, or consider the perception of the age of foreigners/non-native speakers as worthy of attention.[12]

Nevertheless, it has been shown that (native) listeners can recognise the age of native speakers from vocal cues alone with 'impressive accuracy' (Ptacek & Sander 1966), though there is a centring tendency, higher ages being underestimated and lower overestimated (Ryan & Capadano 1978). On the other hand, the effect of age designations (stated age, not real age) as opposed to unknown age upon evaluations of a speaker's personality traits would appear to be minimal (Bell & Stanfield (1973): listeners trust their ears and are not fooled by false information about age.

The age of a voice (real or perceived) clearly affects personality ratings (Ryan & Capadano 1978; Stewart & Ryan 1982). Generally speaking, university undergraduates have been found to hold a negative stereotype of elderly speakers: the older speakers are thought to be, the more old-fashioned, lower in social class, less competent, more reserved, more passive, more out-of-it and less flexible they are rated as being.

However, perceived age would seem to have a lesser effect on speaker evaluation than speech tempo (which nevertheless itself affects perceived age). For example, students can reliably guess the age of speakers, but older people are judged to be older still—and afforded more negative traits—when speaking slowly (Ryan & Capadano 1978; Stewart & Ryan 1982). On the other hand, overall, older/faster speakers are perceived as being more benevolent than younger/slower speakers.

---

[12] For a brief review of the physiology of the ageing voice see Laver and Trudgill (1979:9–11).

The single most important non-verbal vocal indicator of a speaker's age is probably the decline in speech rate in the elderly (Helfrich 1979). However, foreign speakers of a language also regularly speak slowly on account of their deficient phonological (and semantic, stylistic, syntactic) skills, yet few studies have specifically examined how tempo affects the perception of the age of foreign language learners.

4.5.1.3. Accent broadness

Common to all studies of attitudes and reactions towards accented English is the problem with how to define the broadness–mildness of an accent. Two different studies may examine attitudes towards, for example, French-accented English, but there has been little attempt to define the degree of accentedness in the samples used. Do the same segmental and suprasegmental features of accent occur? Are the occurrences of each *potential* feature of mispronunciation equally frequent, and if they are, are they *actually* mispronounced equally frequently? I could well imagine that a Frenchman with a poor command of English will sound *less* French-accented when saying a sentence containing *no* words which should begin with / h / than when saying another equally long sentence containing *lots* of words which should begin with / h /. Moreover, when a bi-dialectal or multi-dialectal speaker produces strongly and mildly accented versions of a text for listener reactions, can we in fact be sure that the features of pronunciation characterising the versions conform with any real degree of accent broadness–mildness as opposed to an artificial stereotype of it? There is presumably in the real world a continuum of accent broadness–mildness, with different segmental and suprasegmental pronunciations appearing and disappearing in a relatively specific order and with an increasing or decreasing frequency of occurrence.[13]

Some studies have tried to solve the problem of accent broadness by including mild and broad versions of an accent for evaluation. For example, one early study of the evaluation of the aesthetic, status and communicative content of three regional British accents (South Welsh, Birmingham and Irish) set accent broadness (mild and broad) against both listener age (21-year-olds v. 12-year-olds) and regional membership (South Welsh v. S.W. England) (Giles 1972b). The results revealed that the older listeners were more sensitive to accent broadness than the younger ones, but that the younger ones were more discriminating in their social evaluations: increasing maturity brings increasing sensitivity and tolerance. More generally the study showed that the broader an accent is perceived to be, the less favourable the evaluation it receives, but the problem of defining the broadness of accents in phonetic terms remained unsolved.

A study with a similar goal used different degrees of Los Angeles Chicano-accented English, the authors being careful to point out that for most speakers Chicano English is a native first language (Arthur et al. 1974).[14] The study, which

---

[13] Whether this order is comparable to the order of acquisition of 'native' speech sounds in infants is of course a further issue which needs to be addressed.

[14] Some studies (e.g. Brennan & Brennan 1981a, 1981b) regrettably fail to say whether the accented speech they use should be considered as the speech of foreign-accented non-native speakers or of ethnic-minority native mother-tongue speakers.

was one of the first to establish that the broadness or mildness of an accent is clearly distinguishable by naive listeners, found that each pair of contrasting guises (mild–broad) evoked clear clusters of highly correlated traits which suggested different and distinct personality types. However, the combined rankings of each trait for both guises of each speaker were so discrepant between speakers that the raters must have been responding to individual voice characteristics as well as to accent differences. More generally the Chicano guises were upgraded for fun and carefree, and consistently downgraded for everything else (intelligence, honesty, reliability, etc.). Stereotyping was obviously at work, but the stereotyping documented here is not a straightforward reaction to race or ethnicity, since all the speakers were identified as Mexican-Americans at commencement of the rating sessions, and even the most Standard American guise still had phonetic features which marked it as the voice of a Mexican-American: the negative reaction was not towards Mexican-Americans, but towards Mexican-Americans with particularly strong non-standard Chicano accents.

The results were also significant for another reason. The range of mean scores for the standard guises of the four speakers was consistently wider than that for the Chicano guises, suggesting that listeners were able to make more subtle evaluations of accents closer to standard American English than to Chicano accents: the stronger the Chicano accent, the more blanket-like the stereotyping (a point confirmed in Brennan & Brennan 1981a, 1981b). This finding supports the 'gradualist' view of reactions to accent broadness and undermines the 'all-or-nothing' school of thought which claims that a minimum amount of non-standardness in pronunciation is sufficient to elicit the complete stereotype of a speaker with a particular non-standard accent, i.e. listeners experience a *categorical* reaction to accented speech, not a gradually changing one (Ryan 1973). However, this latter *Gestalt* view has recently received some corroboration in a study of the degree of pleasure felt in listening to Japanese-accented English: people seem to respond not to the level of accentedness or proficiency of speech but to the generic accent (Cargile & Giles 1997:208).

Identifying an accent is one thing, but quantifying its broadness or strength another: linguistically trained judges can certainly rate the relative frequency of specific pronunciations (or mispronunciations) in a sample of speech labelled as 'accented,' but are the measurements of its perceived *overall* accentedness or nonstandardness as made by naive listeners equally reliable? Two psychophysical methods have been tried: magnitude estimation (ME) and sensory-modality matching (SMM). ME means that listeners assign to the first accented speaker they hear any number that seems appropriate for the amount of accentedness in their speech, and then to successive speakers any number proportional to the impression of accentedness relative to the first speaker. SMM, in contrast, involves listeners squeezing a hand dynamometer with a force matching the accentedness of the speakers they hear. A high degree of interscale agreement has been found when these two techniques are used with listeners rating a wide range of Spanish-accented English (Brennan et al. 1975). The results achieved with ME in fact show a high correlation with the total number of mispronunciations per

speaker as judged by professional linguists.[15] A high correlation is also found between the assessments of naive listeners using a simple Likert-type scale and the detailed phonetic analysis of phoneticians (Brennan & Brennan 1981a and 1981b; Cunningham-Andersson & Engstrand 1989). Naive listeners, therefore, are as good at accent quantification as the professionals. However, it is still possible that certain features of mispronunciation are stronger cues of accentedness than others (Cunningham-Andersson & Engstrand 1989) and this important issue awaits further investigation. If evaluative reactions to accented speech are uniquely categorical in nature (a holistic response), then accented speech merely serves to identify the speaker's group membership, whereas if reactions gradually change with perceived degree of accentedness (an atomistic response), then the features of mispronunciation which prompt that change are of importance far beyond the matter of group identification. The implications for foreign learners (and their teachers) are colossal.

The feature of accent broadness has not only been examined as *between*-speaker broadness, but also as *within*-speaker broadness: how do people react to a person's shifts in accent, to their accent accommodation, whether convergence or divergence? Two studies illustrate this approach (Bourhis et al. 1975): one is set in French-speaking Canada (the stimulus voice shifting from informal Canadian-French, to formal Canadian-French, to Parisian French), and one is set in south Wales (the stimulus voice shifting from broad Welsh-accented English, to mild Welsh-accented English, to RP). Both involve interview situations, with the speaker to be rated converging toward, or diverging away from, the accent of the interviewer. Basically, in Study I, the Parisian French guise was upgraded for competence (intelligence and education) with the informal Canadian-French being downgraded the most. All three French guises were rated equally for affective ratings (sociability, likeability, trustworthiness, etc.). In Study II, RP similarly increased perceived competence (intelligence and socio-economic class), but the shift to broad Welsh increased perceived social attractiveness (trustworthiness and kind-heartedness). Clearly, when people broaden their accents, then listeners perceive them in a different way.

The broadening of one's accent, however, is not necessarily the result of accommodation to an interlocutor: it can also result from a change in speaking style, e.g. reading aloud and free conversation. This special sort of broadening has been tested in Edinburgh with six speakers—all born, raised and educated within the city—taken from different socio-economic backgrounds and reflecting a range of accents ranging from RP through Scottish Standard English to Scots (Romaine 1980b). With only ten listener-evaluators rating the speakers (no effort being made to find a larger, let alone stratified, random sample of subjects) the results are merely suggestive: the speakers were downgraded on the sample of free conversation relative to the reading passage; all the speakers (except the RP

---

[15] The twelve features used in Brennan et al. (1975) were: stressed / æ /; stressed / ɪ / as in bit; stressed / ɒ / as in hot; stressed schwa as in some, stuff; / z /; epenthetic / e / introduced before initial / s / as in stomach, stuff; / v /; fortis / tʃ / and / ʃ /; / j / and / dʒ /; / θ / and / ð /; final / d / as in fried, requested; final consonant clusters, such as discouraged, appeased, last.

speaker) are perceived as sounding more Scottish when speaking than when reading; and none of the listeners thought they spoke like the RP speaker (nor did they want to speak like him!). The results also revealed that informants did not interpret the dimension 'mild–broad' in the same way, that any accent that differs greatly from one's own would seem to be perceived as 'broad': the linguist's and the lay person's understanding of 'broad' is different.

The main theoretical and practical problem here is the definition of broadness: how broad is broad? It is particularly acute in Cheyne (1970) where the accents 'English' and 'Scottish' are left with no more precision than that (see above P. 65): are the accents *mildly* English, *strongly* Scottish, or what? And how mild, how strong? This issue pervades all accent studies, for the phonetic sciences do not as yet allow us to define an accent with a degree of economy which is more manageable or attractive than a subjective statement given in terms of a *Gestalt* (mild, strong); nor do they offer us a satisfactory method for confidently describing *degrees* of accentedness (quite strong, very strong). Moreover, social psychology provides no detailed account for the interaction, influence and accumulation of accent features on our *perception* of the degree of accentedness. And yet the attestable fact remains that an accent can indeed be placed on a continuum of broadness–mildness, which would seem to imply that there is an order of appearance–disappearance of the phonetic features making up the accent, and that each feature contributes in some way to the total reaction of listeners to the accent. Such an atomistic approach to accents, and to listeners' reactions and attitudes to accents, contrasts sharply with the holistic approach underpinning the MGT used by most research in the field: the MGT treats accents as monoliths, clearly defined, consistent, and invariable, even though they are not.

One important question remains unasked: what happens with a noticeably non-native accent—however broad or narrow, strong or weak—which is *unidentifiable*? Both the 'gradual' and the 'all-or-nothing' reaction to a generic accent, I submit, depends both upon the accent being strong enough to be noticed and upon the precision (and confidence) with which listeners can identify it. If the accent cannot be identified, then the stereotype conjured up presumably remains vague: the speaker is merely an unspecified but recognisable foreigner, an outsider.

Thus the exhortation is that studies of reactions to accent must include some control for, and specification of, accent broadness, though there is a crucial distinction to be made between the different degrees of accentedness to be found in regional and social native accents of a language and in foreign accents of a language: shifts between native áccents of a language typically affect a whole spectrum of phonetic details (e.g. consonant articulations, vowel qualities and durations, stress placement, rhythm, pitch movement), each minor shift being a complex combination of interlocking subtly blending elements, but shifts between degrees of foreign-accentedness are usually much more stark, atomistic, one could even say 'pointillist.' In addition, whereas a native speaker of a non-standard accent may move (consciously or unconsciously) a certain extent along

the continuum of standard–non-standard, non-native speakers don't (and usually can't) adapt their *phonetic* foreign speech style (whether consciously or unconsciously) in order to slide along the continuum of sounding more–less like a native-speaker. Sometimes foreigners may perform worse than at other times, but usually this is a result of non-linguistic and non-discourse factors (e.g. tiredness). Just very occasionally one may come across highly proficient foreigners who are able to vary their pronunciation to some minor extent *within* the target language, like native speakers, for social, or situational, or regional reasons (e.g. a foreigner speaking English may want to sound more American and less British). But there are probably few learners of a language (certainly few beginners) who have the phonetic skills to slide smoothly *into* and *out of* a foreign language in order to sound less–more foreign-accented.

4.5.1.4. Tempo, pitch, volume and voice quality

Criticisms were voiced about the fact that when a speaker reads or 'performs' a text for evaluation, the speaker's subjective manipulation of speed in the two or more guises does not result in identical speech rates. This fact prompted hesitations in the interpretation of results (e.g. Stewart & Ryan 1982:106). However, with sophisticated computer programs for sound editing now more readily available, it is possible to manipulate the tempo of recordings very easily, and the results sound—at least to my ear—remarkably non-distorted. On the other hand, technical problems remain.

For example, the manual for the SoundEdit 16 computer program (Version 1.0, published by Macromedia) gives no details as to how the adjustments to speech tempo are effected: is a gross global adjustment made (as seems technically most likely), or (and this seems less likely) is special allowance made for the widely held view (see e.g. Brown 1980:294) that vowel sounds in speech are more elastic than consonants, and are lengthened and shortened more than consonants? Enquiries with the manufacturers of computer programs could of course clarify this point, but the issue is important for it has been suggested that mispronounced consonants are judged to be more serious by native listeners than mispronounced vowels (Johansson 1978:111). If this is correct, is a negative/positive evaluation of a voice/guise the result of anomalies in the discriminant lengthening and shortening?

Some phoneticians also claim that changes of tempo are achieved in English by preferentially shortening or lengthening unstressed syllables, and/or by reorganising syllable structures (Lehiste 1970:38; Laver 1994:544). In other words, tempo embraces a range of phonetic features: vowel durations, Weak Forms (their vowel qualities and durations), elision, syncope, compression and rhythm. Native and bilingual speakers producing voice guises in their mother tongue(s) automatically bring all these various forces into play when making temporal changes, but foreign learners don't. They can't. A major problem for studies of reactions towards foreign-accented speech is how to isolate the influence of all these different features.

However, whether personal manipulation or computer manipulation is used to produce standard tempos, there is the further matter of actually defining tempo, **articulation rate** not being the same as **speaking rate**:

> Articulation rate describes the tempo of articulating an utterance, excluding any silent pauses, but including non-linguistic speech material such as filled pauses and prolongations of syllables. [It] thus refers to the tempo of performance of all audible, 'vocalised' speech within an individual utterance, whether that speech consists of the manifestation of linguistic units or of paralinguistic signals of hesitation. (Laver 1994:539)

Speaking rate, on the other hand, is the overall tempo of performance with all the utterances, filled pauses, prolongations of syllables and silent pauses within the speaking turn.

Laver suggests (1994:541) that a classification consisting of three steady tempos (slow, medium and fast) is probably sufficient for describing most (presumably conversational) articulation rates. However, there are problems here. For example, Goldman-Eisler (1968:24) considers a medium rate of articulation in English (for spontaneous utterances in an interview situation) to be between 4.4 and 5.9 syllables per second; Cruttenden (1994:53) says 'around four syllables per second' for British English; but Ramsaran (1978), as quoted in Cruttenden (1994:267),[16] places a figure of 3.1 sylls/sec on slow spontaneous conversational speech, and (disconcertingly) 5.4 sylls/sec on fast spontaneous conversational speech (i.e. medium tempo = 4.3 sylls/sec). As for reading a text aloud, Calvert (1980:178) suggests that the average speed for an adult is 150 to 180 words per minute, but Grosjean (1980:309) mentions a figure for medium tempo of 224 words per minute. Laver concludes by saying that it might be easier (presumably again for conversation rather than for reading a text aloud) to consider a speaking rate of over 240 words per minute as 'notably fast' and a rate of under 160 words per minute as 'notably slow' (Laver 1994:542).[17]

Despite the problems, the influence of tempo on image formation has been investigated in several matched-guise studies and a wide range of results obtained. For example, evaluations of speaking rate are remarkably consistent across listeners (Markel 1965); faster tempos make both men and women sound more animated and extrovert (Addington 1968); rapid diction in TV news personalities irritates TV viewers (Shosteck 1973); a fast delivery increases perceived intelligence, knowledge and objectivity, i.e. rapid speech (within normal limits) increases credibility and hence enhances persuasion (Miller et al. 1976; Lee & Boster 1992); slow male talkers are downgraded for truthfulness, fluency and persuasiveness, and upgraded for potency and passiveness (slower, colder, weaker) (Apple et al. 1979); an inverted-U pattern appears for the correspondence between speech rate and credibility, fluency, persuasiveness and

---

[16] I have not yet seen Ramsaran's doctoral dissertation (1978).

[17] Notice that Ramsaran's figures for syllables per second (slow: 3.1; medium: 4.3; fast: 5.4) suggest a medium rate +/- 25% for fast and slow, whereas Laver's figures for words per minute (slow: 160; medium: 200; fast: 240) suggest a medium rate +/- 20% for fast and slow. Whether this discrepancy is a substantial difference or merely a result of the use of syllables per second as opposed to words per minute is unclear.

'emphaticness' (*sic*), normal speed being upgraded, fast and slow being downgraded (*ibid.*); both fast and slow speech rates are upgraded for nervousness (*ibid.*); slower speech rates are generally more acceptable than faster ones, though rates similar to one's own, or marginally faster, are the most preferred (Street et al. 1983); old slow males are generally rated least favourably on most scales, the slow delivery apparently augmenting the negative stereotype of old people (Stewart & Ryan 1982).

The perception of competence and benevolence as functions of tempo seems to vary. Generally a fast speaking rate decreases benevolence ratings (e.g. kind, polite, sincere, dependable) and a slow rate competence ratings (e.g. ambitious, intelligent, confident, active) (Brown et al. 1974; Smith et al. 1975; Brown 1980; Stewart & Ryan 1982; Brown et al. 1985). However, the association of slow rate with low intelligence and ambition is lost when the attributional vacuum is eliminated and a speech context is introduced (Brown et al. 1985). Moreover, unusually, young slow males are significantly downgraded for benevolence, probably a result of the disconfirmation of the stereotype that young people speak quickly (Stewart & Ryan 1982); moderate to relatively fast speech rates (actual and perceived) increase ratings of competence and social attractiveness, but extremely fast speech renders speakers less socially attractive than speakers with rates at moderate to moderately fast levels (Street et al. 1983).

Few studies have looked at tempo in foreign-accented speech. Those which do inevitably stumble over the fact that foreigners typically speak more slowly than native speakers: slow speech in a foreigner may be incredibly slow relative to a native speaker (25 words/min, or 2.5 sylls/sec), and fast speech may be slow (39 words/min, or 4.5 sylls/sec): see Gynan (1985) and Anderson-Hsieh and Koehler (1988) respectively. Generally speaking, the issue of whether tempo in what is obviously foreign-accented speech is perceived in the same way as tempo in native-speaker speech remains unaddressed. We still know little about whether the perception of a foreigner alters with speech tempo in the same way as the perception of a native speaker. However, one extensive study of Chinese-accented English has found that high speed decreases comprehensibility, that accented English is—as expected—consistently less comprehensible at whatever speed than native-speaker English, that the faster the speech and the heavier the accent, the greater the incomprehensibility (i.e. speaking rate is critical for heavy foreign-accented English), and that the more severe the accent, the faster the perceived relative speed of speech (Anderson-Hsieh & Koehler 1988).

The perception of tempo in foreign-accented speech may naturally be affected by the presence of silent pauses, though one study claims that the removal of pauses does not in fact affect people's perceptions of the degree of foreign accent (Flege 1988). This finding is somewhat surprising, particularly for anyone with experience of Finnish modes of discourse: silence is a noticeable feature of Finnish 'conversation' (Lehtonen & Sajavaara 1985), and the presence of silent pauses—their frequency and duration—may be of particular importance in the perception of Finnish speakers of English.

Pitch has received some attention in studies of reactions towards accents, though considerably more in studies of emotional state as opposed to long-term personality traits. Listeners can certainly identify pitch reliably (Markel 1965); increased pitch variety in men is perceived as more dynamic, feminine and aesthetically inclined, in women as more dynamic and extrovert (Addington 1968); and higher pitch generally results in speakers being deemed less competent and less benevolent (Brown et al. 1974). The fact that pitch is significantly higher in subjects who are lying curiously does not lead to the perception and attribution of deception: either the pitch change is too small, or else listeners use such things as nervousness, nonjuncture pauses, gaze aversion and facial shielding as cues to deception, but not pitch (Streeter et al. 1977).

High pitch, of course, is usually associated with stressful situations, particularly with anger and fear, low pitch (and slow delivery) with 'low energy' states such as sorrow and indifference, i.e. with temporary emotional states, but men with high-pitched voices (the pitch having been raised and lowered by acoustic manipulation but in such a way as to remain sounding 'natural') have been found to be downgraded for truthfulness (credibility), persuasiveness, emphasis and potency (smaller, thinner, faster) and upgraded for nervousness (Apple et al. 1979).

Volume seems to receive only cursory attention: jokes are made about loud-mouthed American tourists, and bland comments are made about the fact that people do not usually shout in church and at funerals although they do shout at football matches and in political arguments; but few studies seem to focus on the image that loud speakers and soft speakers (particularly loud and quiet foreign speakers of English) create in listeners (see Mallory & Miller 1958, Davitz & Davitz 1961, Markel 1965, Phillis 1970).

Voice quality has received scant attention despite considerable classificatory/physiological work in the area (Laver 1968, 1974, 1980; Laver & Trudgill 1979). Nevertheless, several voice qualities have been found to influence listeners' perceptions of speakers (Addington 1968: see Table 4.2, P. 91; Pennington 1996: 158). Apart from evidence suggesting that alterations in female voice quality are more effective in changing personality ascriptions than those in males, there is also reason to believe that interactions between voice qualities and other signals (e.g. vocabulary, speech content, visual cues, etc.) may be non-additive, and that reactions may be curvilinear. Certainly visual cues may override some speech cues in impression formation (Williams et al. 1971), but when, where and to what extent remains obscure.

It has also been found that: voice quality is directly related to the appeal of TV personalities (Shosteck 1973); German men with breathy voice are perceived as neurotic, dominant and unsociable, tense voice being associated with emotional instability, extroversion and dominance (Scherer 1979a); creaky, nasal and tense voices are negatively related to solidarity judgements, while nasal voice is negatively related to status judgements, with the result that the 'ideal' voice is perhaps characterised by small amounts of several voice qualities, particularly breathy, creaky, nasal, tense and whispery voice (Pittam & Gallois 1986).

Table 4.2 Reactions to increases in specific voice qualities in men and women

Voice quality	In men	In women
Breathiness	younger, more artistic	more feminine, prettier, more petite, effervescent, highly strung, shallower
Thinness	∅	immaturity on social, physical, emotional and mental levels; more humour and sensitivity
Flatness	more masculine, sluggish, withdrawn and colder	
Nasality	'A wide array of socially undesirable characteristics.' (P. 502)	
Tenseness	older and more unyielding, i.e. cantankerous	younger, more emotional, feminine, highly strung, less intelligent
Throatiness	older, more realistic, mature, sophisticated, more well adjusted	less intelligent, more masculine, lazier, more boorish, unemotional, ugly, sickly, careless, naive, humble, neurotic, quiet, uninteresting, apathetic, i.e. cloddish/oafish
Orotundity	more energetic, healthy, artistic, proud, sophisticated, interesting, enthusiastic, i.e. hardy and aesthetically inclined	more lively, gregarious, aesthetically sensitive, but also more proud and humourless

(Adapted from Addington 1968)

No studies seem to have examined the influence of voice quality on native listeners' perceptions of a foreigner speaking their (the listeners') language with a foreign accent, though foreign language teachers have been urged to instil in learners that voice quality settings imported into an L2 from the native L1 may evoke unfavourable responses from L2 native listeners (Esling & Wong 1983; Celce-Murcia et al. 1996:28; Pennington 1996:162–5).

4.5.1.5. Background information, context and speech content

The fact that a listener's perception of a speaker is influenced by context, i.e. non-phonetic and non-linguistic 'background' information concerning the speaker (e.g. employment and social position), even visual cues, kinesics and proxemics, received increasing attention following Agheyisi and Fishman (1970). Attempts were made to ascertain whether vocal cues are over-ruled by non-vocal cues or vice versa.

The effect on speaker-perception of background information concerning the social class of a speaker remains unclear. One non-MGT study found that Spanish-accented English coupled with the 'pre-information' that the speaker is from a low social class produces inevitable derogation of the speaker (Ryan & Sebastian 1980). However, when American listeners evaluated standard Ameri-

can and German-accented English speakers labelled as upper middle class or lower class, the social class information affected the perception of social status (intelligent, successful) but not that of solidarity (friendly, trustworthy) (Ryan & Bulik 1982). Nevertheless, when British listeners were given similar social class 'pre-information' about RP and Cockney guises, no interaction was found at all: accent alone determined perceived status regardless of 'known' socio-economic standing (Giles & Sassoon 1983).

'Post-information' has also been found to have an effect on speaker evaluations. In other words, a 'retroactive speech halo effect' may affect people's judgements, the speech stereotypes that listeners harbour for people of differing class and status being activated retrospectively such that these stereotype images corrupt (within certain limits, presumably) the memory of how a speaker actually sounded (Thakerar & Giles 1981; Ball et al. 1982). Since the 'reality' of a speech style may be distorted by a belief planted just moments later in the minds of listeners, one may well wonder how or when reactions to speech styles are *not* 'contaminated.'

Context has been approached in various ways. In one study the home domain (mother preparing breakfast) and the school domain (teacher giving a history lesson) were contrasted in order to focus on the attitudes of Mexican American (MA) and Anglo American (AA) Chicago students towards the use of standard Spanish and English (Carranza & Ryan 1975). The results showed that the MAs and the AAs upgraded the English speakers both for status and, unexpectedly, for solidarity. However, an examination of the interaction of 'factor x context x language' revealed the hidden detail that both MAs and AAs upgraded Spanish in the home context but English in the school context: in other words the listeners reacted to the appropriateness of the language for the particular situation.

A similar study set a school environment against a youth club, and RP against Welsh-accented English (Creber & Giles 1983). The results revealed that regardless of setting Buckinghamshire school children upgraded RP for *status* traits, but—unexpectedly—they did not upgrade the Welsh-accented English for *solidarity* traits. However, the upgrading of the social traits for RP was more extreme in the formal school setting than in the informal youth club setting: social setting has a clear effect upon children's language attitudes. 'Children have differing evaluative sets in formal versus informal settings' (P. 159).

Simulated radio interviews have also been used in an attempt to increase ecological validity. Informants heard two interviews supposedly recorded after a diving competition, the athlete changing or not changing speech style to accommodate to the interviewer (Bourhis et al. 1975). Basically the athlete's upward convergence in both Canada and Wales led to an increase in perceived intelligence, though in Wales it also led to a decrease in perceived trustworthiness and kindness. Regrettably only one female interviewer was enlisted, thus excluding gender control. In their conclusion the authors noted that the listeners' perceptions of the targeted speaker (the athlete) may have been distorted by the inter-

viewer's voice (i.e. experimenter effect): the total speech event affects perceptions of individual participants in the event.

The context of speaker evaluation was extended to the courtroom with a matched-guise study exploiting speech style rather than accent (Lind & O'Barr 1979). Testimony using so-called 'power' speech (infrequent use of intensifiers, empty adjectives, hyper-correct grammar, polite forms, gestures, hedges and rising intonation, and a narrow range of intonational patterns), in both men and women, was found to render a witness more competent, attractive, trustworthy, dynamic and convincing than 'powerless' speech. Simulated jury discussions have also been studied: voice quality and speech style are powerful tools in influencing the opinion of co-jurors (Scherer 1979b). The serious implications for trial outcome and justice were confirmed with an MGT accent study (Seggie 1983): Australian students are more likely to attribute guilt for white collar crimes (like embezzlement) to RP speakers than to non-RP speakers, and guilt in crimes of violence to non-RP speakers (Broad Australian or Malaysian Chinese accents) than to RP speakers.

The medical context of the doctor–patient diagnostic interview has also been exploited (Fielding & Evered 1980). When a rural regional (Southwest England) guise was set against an RP accent in one MGT study, the speaker (a male 40 year-old) was identified as lower or middle class respectively, and exactly the same symptoms voiced in the regional accent were considered as less psychosomatically and more physically determined than those voiced in RP. The patient perceived as being lower class (the Southwest guise) was seen as having difficulty in communicating and detailing his symptoms, the one perceived as being middle class as being emotionally unstable.

Several studies have investigated the role of accents in simulated employment settings. One early work which presented standard English, Black English, Spanish-influenced English and Southern-white-dialect English in job interview situations surprisingly found that ethnicity is *not* important for employment decisions (Hopper & Williams 1973). By contrast, Spanish-accented prospective employees are downgraded by South Floridan employers (Rey 1977). English Canadian students have shown similar discrimination against foreign-accented speakers (Italian, Greek, Portuguese, West African and Slovak accents) as opposed to Canadian-English-accented speakers in simulated job application situations at four levels of job status (Kalin & Rayko 1978). The same interaction between accent and job status was also found in a study of applicant suitability for jobs of various status (Kalin et al. 1980): jobs ranging from high to low status equated with the order of accents English English, German English, South Asian English and West Indian English. It has further been found that evaluations of speakers are 'stricter' in a supposedly formal interview situation than in an informal conversational setting, a more standard accent prompting upgrading, a non-standard accent downgrading (Street et al. 1984). The dynamics of job interviews also affect applicant evaluations: interviewees are upgraded by active interviewers but downgraded by non-participant observers (Street 1985).

Listeners are also more aware of slow and fast extremes in an employment interview than in a conversation setting (Street et al. 1983).

Dress style was introduced as an independent non-linguistic factor in a study where a bidialectal woman in her early twenties, speaking Cockney or RP, and dressed smartly or casually, conducted a live face-to-face survey (Cf. Giles et al. 1975) requiring open-ended written responses from house-wives in a middle-class suburb in Berkshire (Giles & Farrar 1979). Listeners' reactions to the combination of clothing and voice guise were assessed by the length and style of written answers (cf. the theatre audience survey of Bourhis & Giles 1976). The results showed that dress style had no effect on length at all, though it did interact with accent guise to influence the writing style of respondents: the RP guise and smart appearance elicited a formal style; the RP guise and casual appearance elicited an informal style; while the Cockney guise, whether with smart or casual dress, triggered a uniform writing style midway between the two former extremes.

Dress style has also been presented in photographs of speakers, together with factory or office background, to suggest socio-economic class (middle class or lower class). The results, disconfirming predictions based on consistency theory (Helson 1947, 1959; Festinger 1957), showed that a 'discrepant' stimulus person (working class + standard accent; middle class + non-standard accent) was upgraded relative to a 'consistent' person (Aboud et al. 1974). Thus behavioural responses to accent depend not just on vocal cues or even social context, but also on other perceptual cues.

The crucial point here, that the whole context of speech interaction needs to be considered, was reiterated in the Welsh/RP study of Giles and Marsh (1979) but extended to embrace an even wider sphere: linguistic attitudes are 'the product of a particular sociohistorical context' (P. 308). Even once we have described the stimulus voice, speaker and informant characteristics as precisely as possible, social, political, ideological and temporal factors should still prompt us to treat the generality of findings with suspicion and to draw generalisations with caution.

Surprisingly speech content (*what* is said) has not always been recognised as a moderating variable in the evaluation of voices, despite the early demonstration of its effect on the evaluative factors of the semantic differential (Markel & Roblin 1965), and despite proof that lexical diversity interacts with social status (as perceived subjectively through accent) to affect impression formation advantageously in some contexts but detrimentally in others (Giles et al. 1981). Protestations as to the neutrality of texts used have regularly been made, and some authors have even resorted to a recitation of the alphabet (Berry 1992), but few authors have conducted parallel experiments in order to obtain corroborative results. One exception is a study of the interaction of accent (weaker/stronger) and fluency (better v. worse vocabulary and grammar) in Japanese-accented English (Cargile & Giles 1997): accent seemed to be more salient than fluency. Another exception investigated the interaction of accent (standard v. non-standard) and message (pro v. anti) in California, and found that speakers were

considered more persuasive when defending a viewpoint contrary to the one expected (Giles et al. 1995).

Speech content—whether viewed in terms of subject matter, grammar, vocabulary, style or register—always influences the reactions of listeners, but the implications for foreign language learners are enormous, since restricted lexical variety and syntactic deficiencies, together with accent, are the conspicuous hallmarks of their speech. In studies of attitudes towards foreigners speaking a foreign language, more attention needs to be given to background information, context and speech content (Giles & Coupland 1991b; Cargile et al. 1994:214), with special care being given—I believe—to the distinction between (a) a foreigner of known origin and ethnicity, with a known mother-tongue, with a known socio-cultural background, and conversing for a particular reason, and (b) a foreigner of *unknown* origin and ethnicity, with an *unknown* mother-tongue, with an *unknown* socio-cultural background, and conversing with various degrees of linguistic deficiency for as yet to be ascertained reasons.

4.5.2. Informant characteristics

The informant characteristics which have received particular attention are: age, gender, social class, socio-economic background, ethnic background, regional membership, social attitudes, mastery of languages (mother-tongue and foreign languages), and knowledge concerning the voices/speakers being presented for evaluation.

4.5.2.1. Listener's age

Since the first adoption of the MGT, the age of informants has been recognised as being important, different age groups being used and care being taken to name them. The majority of studies have used students, though some have consciously avoided such informants: e.g. Anisfeld and Lambert (1964) used 10-year-olds, Lambert et al. (1966) three groups within the 10–17 year age range, Giles (1972b) 12 and 17 year-olds, Creber and Giles (1983) 12–14-year-olds, Paltridge and Giles (1984) 12, 30 and 72 year-olds, Norell (1989) 13–17 and 21–61-year-olds.

Most studies have noted that children's reactions (up to about 14 years of age) are distinct from those of adults, the supposition being that 14–15-year-olds are too young to be aware of the social significance of accents (Labov 1965:91). However, it is still unclear *when* youngsters assimilate the norms of older people and society concerning prestige speech. For example, when evaluating RP, South Welsh and Somerset accents, 17-year-olds rate RP more favourably on five traits (good looks, ambition, intelligence, self-confidence, determination), and less favourably on four (seriousness, talkativeness, good-naturedness and humorousness), than 12-year-olds, presumably reflecting social maturity and/or experience (Giles 1971c). On the other hand, 7-year-olds downgrade an RP speaker relative to a South Welsh one for both competence and social attractiveness, whereas 10-year-olds do not (Giles et al. 1983).

At the other end of the age spectrum ageism has reigned: few studies have been conducted with people over 30 years of age, though one non-MGT study

which looked at attitudes towards French regional accents showed that the elderly (70-year-olds) were generally more reluctant to derogate non-standard regional-accented speakers, and less concerned with status connotations, than their younger counterparts, 12-year-olds and 30-year-olds (Paltridge & Giles 1984). More attention must be given to e.g. young-old v. middle-old v. old-old informants, retired v. non-retired informants (e.g. active septuagenarians), and chronological age v. subjective age (*ibid.* P. 80).

4.5.2.2. Listener's gender

The gender of informants was quickly recognised as a variable potentially affecting ratings of accents (Markel & Roblin 1965), but few studies have produced noteworthy results. For example, although men and women reveal clear-cut differences in their perception of male and female speech characteristics when replying to a written questionnaire (Kramer 1977), male and female listeners respond in the same way to actual voice quality (Addington 1968) and perceive personality traits of actual foreign accents/speakers in the same way (Giles 1970; Mulac et al. 1974; Gordon & Abell 1990). This may be why some empirical studies have downplayed informants' gender: some have used one gender but not the other; some diligently investigate the effect of informant gender on responses; and some have mixed men with women in equal (or unequal) proportions but not used listener gender for the statistical analyses (see Table 4.3). The

Table 4.3 Gender of informants in various studies

Informants' gender	Study
Male	Wölck 1973[a]; Ryan & Capadano 1978 (85% male)
Female	Lambert et al. 1966; Williams et al. 1971 (95% women); Ryan & Carranza 1975; Masterson et al. 1983
Male v. Female (Genders set against each other for analysis)	Stagner 1936; Markel & Roblin 1965; Addington 1968; Tucker & Lambert 1969; Giles 1970; Giles et al. 1974; Mulac et al. 1974; Carranza & Ryan 1975; Bourhis & Giles 1976; Simard et al. 1976; Kramer 1977; Elyan et al. 1978; Giles & Marsh 1979; Gallois & Callan 1981; Paltridge & Giles 1984; Street et al. 1984; Berry 1992; Kristiansen & Giles 1992
Male/Female (Genders not used for analysis)	Lambert et al. 1965; Giles 1972b; d'Anglejan & Tucker 1973; Giles 1973a, 1973b; Hopper & Williams 1973[b]; Arthur et al. 1974; Bourhis et al. 1975; Brennan et al. 1975; Flores & Hopper 1975; Ryan et al. 1975, 1977[c]; Kalin & Rayko 1978; Giles et al. 1979; Giles, Smith, Ford et al. 1980; Romaine 1980b; Brennan & Brennan 1981a; Ryan & Bulik 1982; Stewart & Ryan 1982; Creber & Giles 1983; Giles et al. 1983; Brown et al. 1985; Gynan 1985; Pittam & Gallois 1986; Anderson-Hsieh & Koehler 1988; Norell 1989; Giles et al. 1995; Cargile & Giles 1997

[a] 'A few women.'
[b] Gender of informants is not actually stated but presumed.
[c] 'Mainly male.'

pressure from the 'new' disciplines of gender studies and women studies has still to make an impact. No studies would appear to have investigated the effect of the gender identity and orientation of informants on attitudes to accents.

4.5.2.3. Listener's background

Some interest has been shown in the possible effect on attitudes towards language of the listeners' own social background, ethnic status, language use, beliefs and so on, but only meagre results have been obtained. Work in this area reflects an increasing awareness that the population under investigation has to be clearly defined. A convenience sample has become increasingly unacceptable.

One potential moderating influence on attitudes towards language is the social and socio-economic class of informants (Ryan 1973; Ryan & Carranza 1977), but theories of social stratification and social class—whether of the American structural-functionalist tradition or of the British neo-Weberian and neo-Marxist tradition—and the procedures for social class identification and the assignment of informants into classes pose serious problems (Barber 1957, Oakley & Oakley 1979, OPCS 1980 and 1986, Cameron 1985, Abbott & Sapsford 1987, Graddol & Swann 1989, Saunders 1990). It is insufficient to compile lists of jobs, equate occupations with social and/or socio-economic class, and then blindly claim those classes affect language use (as did Bernstein 1962a and 1962b, Labov 1972, Trudgill 1972 and 1974b, Macaulay 1976 and 1977) let alone attitudes towards language. Over half the population of every Western civilised country (the underclass of children, students, 'house-spouses', retirees, the unemployed, the permanently sick, etc.) is economically inactive, and it is illogical to say their language habits are 'dependent' upon the employed minority. In brief the employment-based gender-biased methods adopted by sociologists for classifying families for *global* purposes (standards of living, income, health, life expectancy, etc.) and their approach to social stratification are unsuitable for classifying individuals for *linguistic* purposes. Thus if social class *is* to be examined as a possible influence on attitudes towards language, linguists have to decide on which definitions of social class and/or socio-economic grouping are germane to this phenomenon, on how many levels of class are catered for, and on the validity and reliability of the instrument used to measure them.

The few studies which have broached this subject vary considerably. One study involving adolescent informants is open to the charge of gender bias, the status being partly based on the father's occupation (Ryan, Carranza & Moffie 1975); another study (Aboud et al. 1974) had college student informants rate themselves; and the authors of a third study (Giles & Sassoon 1983) 'informally assessed' student informants themselves (they were all middle class). However, the socio-economic status of informants was only used in these three studies for delimiting the population and not for the analyses.

One study to have clear-cut results indicating the major influence of social class on perceptions of personality traits—and the greater importance of social class than age on such perceptions—used 10–17-year-old children whose social status was mainly determined by father's occupation (Lambert et al. 1966). The

upper middle-class children all exhibited in-group downgrading whereas the lower class/lower middle-class children revealed in-group upgrading up to about 14 years of age.

There is a distinct dearth of studies examining the importance or otherwise of informant social class on attitudes towards language. It is possible that education, beliefs and socio-political context may prove more valid and reliable influences, particularly for the perception of foreign-accented speech which may involve different social, cultural, religious, political and ideological perspectives.

For example, as predicted by ELIT (see above P. 78), strong feelings of ethnocentrism have been shown to prompt down-grading of regional accents but upgrading of the standard social accent RP (Giles 1971b), and 'less ethnocentric [listeners] tend to be socially less cynical . . . prepared to perceive some favourable qualities in all speakers' (Giles 1972a:170). However, ethnolinguistic vitality has little influence on a person's readiness to converge to the language of their interlocutor (Côté & Clément 1994).

One surprising finding has been that linguistically experienced informants (i.e. speech therapy students) evaluated three forms of Irish-accented English in the same way as linguistically naive informants (i.e. natural science students) (Masterson et al. 1983).

The importance of the socio-political context of informants as a moderator variable has already been referred to in connection with, for example, the Arabic/Hebrew study of Lambert et al. (1965), and the Welsh/RP study of Giles and Marsh (1979). Another study which vividly highlights this factor examined the effect of speech styles on memory recall (*not* speaker perception) in Catholic v. Protestant family background 10–11-year-old children in Northern Ireland: the recall of Catholic children was worse than that of Protestant children when they listened to an RP guise, and worse than that of Catholic children who had listened to a Belfast guise; Protestants who had listened to a Dublin guise recalled less than Catholics who had heard a Belfast guise (Cairns & Duriez 1976).

The regional origin of listeners has also been found to have some effect on the perception of speakers (Paltridge & Giles 1984).

4.6. Foreign-accented accents

Despite Sapir's warnings seventy years ago concerning the difficulties faced by investigators of person perception in cross-language studies, studies of attitudes towards accented language have regularly failed to separate regional and/or social *native* accents from non-native *foreign* accents, and to distinguish bilingual and bicultural *second*-language situations from monolingual and monocultural *foreign*-language situations. Thus Lambert's first studies using French accents and English accents blurred the distinction between *second* and *foreign* language, and between *second*-language-accented and *foreign*-accented speech, and no special account was taken of whether informants knew, or were familiar with, the 'other' language. Similarly, some studies which have looked at Spanish-accented English have failed to define whether the accent is being treated as a native accent (the speaker's mother tongue is English), or a foreign

accent (the speaker's mother tongue is Spanish),[18] and whether the situation is a second language situation (the informants live in a bilingual English/Spanish community, and know Spanish) or a foreign language situation (the informants don't live in a bilingual community and don't know Spanish). The process of accent and speaker evaluation which is called into play for the different situations has generally been assumed to be identical.

Only one study clearly seeks to move the spotlight from non-standard regional/social *native* accents to non-native *foreign* accents (Ryan 1983), though even this study fails to distinguish the terms 'second' and 'foreign' language. Ryan (1983) specifically pondered the social psychological mechanisms underlying native speaker evaluations of non-native accents, proposing that evaluative reactions to foreign-accented speech be interpreted with (a) three essentially cognitive social psychological mechanisms, (identification, stereotyping and competence inferencing), and (b) one non-cognitive mechanism (generalised negative affect arising from frustration, discomfort, extra effort, uncertainty, etc.), this last point being partly based on Sebastian et al.'s (1980) finding that the negative affect caused in a listener by the effort and frustration of trying to understand Spanish-accented English leads to derogation of the speaker on several personality and social dimensions (e.g. status, social class and attitude/belief similarity).

To elaborate briefly, Ryan (1983) claims that within the archetypal MGT experiment, speaker evaluation is assumed to entail two consecutive steps: first social group identification (age, gender, ethnicity, social class, etc.) based on voice and speech cues, and then stereotyping arising from the social categories established. However, these two stages (also identified in Delia 1972, and in Bradac & Giles 1991) and their chronological sequencing are not automatic and obvious where non-native speech is concerned: the two processes are more likely to be complementary and interactive, and the stereotyping is not necessarily based exclusively on assumed social group membership but on that and speech/phonetic cues as well. Nevertheless, the identification process (which is often dependent on auxiliary non-phonetic information) probably involves—says Ryan—up to five stages, and the success with which listeners navigate their way through them determines the level and type of stereotyping which takes place (Table 4.4).

Table 4.4 Levels of specificity in the identification of non-native speakers

Level 1:	Is the speech non-standard?
Level 2:	Is the speaker a learner?
Level 3:	Which basic language group does the speaker belong to?
Level 4:	What is the speaker's native language?
Level 5:	What regional/class dialect of that language does the speaker use?

(Adapted from Ryan 1983:151)

Particularly problematic for the evaluation process—again according to Ryan—is the misidentification of a native non-standard accent as a foreign one.

[18] A clear exception is Arthur et al. (1974).

However, whether the accent is correctly identified or not, the levels of tolerance as revealed on the two basic evaluative dimensions of social status and group solidarity are strongly influenced not only by the context of the linguistic deviations (e.g. school v. home, which usually equates with status v. solidarity) but also by the purpose of the conversation. Further variability in listeners' evaluations of a foreign-accented speaker may arise as a result of the listeners' social identities (ethnicity, social class, gender, age, etc.) and the mismatch between the listeners and the speaker.

The third cognitive mechanism to impinge uniquely on the evaluation of non-native speakers—according to Ryan, who draws on theories of causal attribution here—is competence inferencing. The presumption of non-native competence (and the absence of linguistic accommodation) and the presumed degree of incompetence typically prompt downgrading of traits (and overgeneralising of weakness in language skills to weakness in cognitive skills![19]), though some positive results (e.g. increased tolerance) may occur, especially when auxiliary non-language information is available.

The effects of negative inferences on the behaviour of the listeners themselves is relatively well known—especially 'foreigner talk' with its simplified syntax, vocabulary and content, louder volume, slower speech rate, and exaggerated intonation—yet little is known about the level of ungrammaticality which is tolerated (and expected) for a given level of non-native accent, or about the advantages of *retaining* a foreign accent (and ungrammaticality, limited vocabulary, restricted politeness markers, etc.), and so on.

> A certain amount of nonstandardness (e.g. a language learner's accent) can sometimes attenuate the impact of another aspect of nonstandardness (e.g. grammatical or sociolinguistic errors). Further investigation of the specificity v. generality of competence inferences, their interactions with other inferences, and their complex social meanings would be particularly valuable for understanding reactions of others toward language learners and for understanding the learners' own self-perceptions and motivations. (Ryan 1983:157)

However, negative deviations from expectations doubtless lead to irritation, confusion and insecurity, which lead to the generalised negative affect which constitutes Ryan's fourth non-cognitive mechanism exerting an influence upon speaker evaluation; and negative affect leads to further derogation of the speaker.

Ryan's insistence on the role of negative affect is repeated in the so-called process model of language attitudes where affective responses constitute one of several links in a chain of dynamic 'virtually endless' recursive loops (Cargile et al. 1994:214–5), and it gains some justification from empirical studies. For example, the effort and frustration of listening to English which is punctuated with bursts of white noise has been shown to lead to negative affect and derogation of the speaker (lower status, solidarity and social class, and greater social distance) (Sebastian et al. 1980). Contrariwise a positive attitude towards

---

[19] Charles Dickens offers an apt quotation here: '. . . and there was a droll disposition, not only on the part of Mr. Podsnap, but of everybody else, to treat [the foreign gentleman] as if he were a child who was hard of hearing.' (*Our Mutual Friend*)

foreigners and foreign speech *increases* comprehensibility, particularly in heavily accented and fast speech (Anderson-Hsieh & Koehler 1988:590). This implies that if a listener is supportive towards the speaker, is sympathetic and 'other-centred,' then the evaluations of the speaker will be more positive than if the listener is non-supportive, non-sympathetic and self-centred.

The role of negative affect in attitude formation is also mirrored in the social identity theory that asserts that the stronger one's feelings of group identity, and the more salient that identity (both of which foster the support of one's own group and antipathy towards an outgroup), then the more likely one is to downgrade a speaker with a foreign accent (necessarily an outgroup member), particularly if the message heard is perceived as being aggressive to one's own viewpoint (Cargile & Giles 1997:199). Negative affect would also seem to be present in people with low feelings of group identity who indulge in self-downgrading, as witnessed by French-speaking Canadians downgrading French speakers relative to English speakers (Anisfeld et al. 1962), and Spanish-speaking Mexican-American listeners downgrading Spanish speakers relative to English speakers (Flores & Hopper 1975).

My attempt at a diagrammatic generalised model of the process underlying the perception of foreign speakers based on their language is presented in Figure 4.1, P. 102. The line of cause and effect goes horizontally across the page (language → attribution → evaluation; independent variable → dependent variable), and the moderating variables (which may form recursive loops, though this is not shown here) rise vertically (↑). The final evaluation may swing more positive or more negative (↙ ↗) depending on the orientation of the listener ('other-centred'–self-centred).

Ryan's (1983) initiative remained broadly unnoticed and a trawl of the literature reveals very few studies which examine native-listener attitudes towards unspecified and unidentifiable foreign accents. Some studies, while focusing on foreign accents, simply fail to disclose whether informants can (or did) identify the accent (e.g. Norwegian, Italian, Czechoslovakian, Polish and Russian in Mulac et al., 1974; German-accented English in Ryan & Bulik, 1982; Japanese-accented English in Cargile & Giles, 1997). One study even used a 'feigned foreign accent' without saying what it was supposed to be (Varonis & Gass 1982). One study, however, confirms the obvious point that some foreign accents are more familiar, and are thus easier to identify, than others (Gallois & Callan 1981).

In short, I would hypothesise that if the reason for an accent or mispronunciation is known to be (or suspected of being) non-native speaker incompetence deriving from an unspecifiable interlanguage, then evaluations of the speaker will be different than if the speaker is believed to be a native speaker albeit of unidentifiable region, nationality or ethnicity. Of course the whole range of factors impinging on the perception and evaluation of native accents already discussed is still important (viz. the speaker's and listener's age, gender, and social status; the speech content; the voice quality; tempo; the context of the speech event), but in

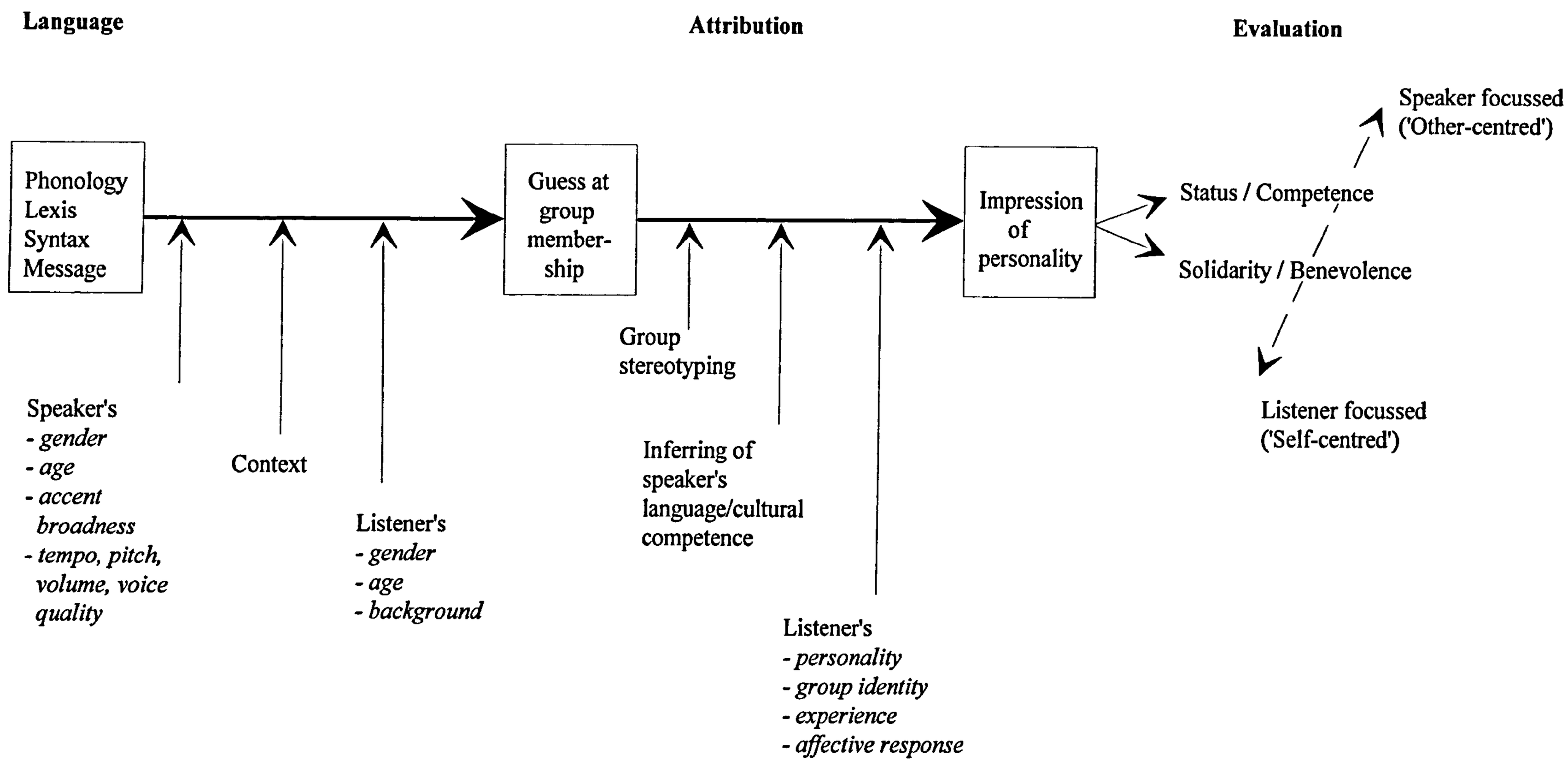


Figure 4.1 Theoretical model of foreign speaker perception

addition the following factors necessarily affect the evaluation process: overall comprehensibility (Varonis & Gass 1982); the type of pronunciation deviance, whether segmental, prosodic, epenthetic, deletive or metathetic (Anderson-Hsieh et al. 1992); the degree of accentedness (Ryan 1973) and the range of accentedness being heard within a group of speakers (Flege & Fletcher 1992); the listener's linguistic experience, particularly in listening to foreign accents (Wingstedt & Schulman 1984:343); lexical choice and variety (Cargile & Giles 1997); compatibility of ungrammaticality to degree of mispronunciation (Cargile et al. 1994); and the listener's knowledge of and/or ability to identify the social/national identity of the speaker.

Since it is known that the stronger a foreign accent is, the greater the downgrading of the speaker (Ryan et al. 1977), it is essential to be able to define and quantify degrees of accentedness (Ryan 1973). However, as yet we have no tools for labelling, measuring and reporting accentedness. We do not even know which if any features of mispronunciation intrude more on listeners' perceptions of foreign speakers. Giles, Smith, Ford et al. (1980) confidently defend a holistic approach to *native* speech, because

> one does not . . . hear isolated speech rates and vowel or consonant sounds, fluctuations in pitch, etc. under normal conditions. One hears speech, a conglomerate of potentially informative stimuli all emanating from the same source. (Giles, Smith, Ford et al. 1980:274)

Nevertheless, I sympathise more with an atomistic approach for *non-native* speakers and *foreign*-accented speech because mispronunciation may turn the attention of listeners away from the content of the speech to a particular feature of it, segmental, intonational, pitch or whatever. For example, listeners hearing [zə], [zɪs], [zæt] for *the*, *this*, *that* may be so distracted that they fail to register the full content of the message being transmitted. Moreover, although it has been found that the effect of adding foreign mispronunciation features (to Swedish) is cumulative, the fact remains that 'the various deviant features are not equally powerful when it comes to creating the impression of a [foreign] accent' (Cunningham-Andersson & Engstrand 1989:146). I would contend that the holistic model of attitude formation (and the *Gestalt* perception of personality and character traits) is more suitable to the native accent situation where conversation chugs along without conscious attention necessarily being paid to the vehicle of communication, the conversational strategies being followed, the psychological and sociological game plans being played, and so on, but that an atomistic approach tallies better with reality for foreign-accented speech where the conversation may stumble along with one being painfully aware of the phonetic deficiencies of one's interlocutor.

Sometimes particular phonetic mistakes may permit the identification of the speaker as French, or Spanish, or whatever, and this prompts national stereotyping. However, if the accent remains unidentified, then a national stereotype cannot be invoked, and the perception of the speaker rests on phonetic cues alone. But these cues are interpretable only by the speaker's socio-phonetic framework which is unknown and inaccessible to the listener. Thus the question

is: are phonetic cues consistent across languages, e.g. does a non-aspirated plosive always suggest characteristic X? Unfortunately the phonetic details of particular foreign-accented speech—whether French, or Spanish, or Quechua (Wölck 1973), or German (Ryan & Bulik 1982), or Japanese (Cargile & Giles 1997)—and the frequency of occurrence of specific misarticulations[20] are rarely given[21] and this makes any comparison of study results impossible.

The results of studies are indeed limited and dominated by studies of Spanish-accented and French-accented Englishes on account of their relatively wide-spread presence on the North American continent. Generally speaking, men and women would seem to have similar views about foreign accents: American men and women (Mulac et al. 1974) and Australian men and women (Gallois & Callan 1981) rate the personality traits of foreign speakers in the same way. Overall, all foreign accents would seem to be equally foreign—though variations may result from an interaction of speaker gender and mother-tongue—and all Englishes accented by some European language consistently elicit lower personality ratings (socio-intellectual status, aesthetic qualities, dynamism) than native American English (Mulac et al. 1974). Broadly, foreign female speakers are downgraded relative to foreign men on dynamic traits (active, strong, powerful, etc.) but upgraded on evaluative traits (helpful, friendly, educated, intelligent, etc.), this being true across all foreign accents and for both male and female listeners, though again speaker gender may interact with nationality to influence listeners' judgements (Gallois & Callan 1981). Generally, foreign-accented speech would seem to be rated higher (i.e. deviations are tolerated more) in an informal setting than in a formal one (Ryan & Carranza 1975).

Much work has been done in the field of attitudes towards accents, as my review attests, but not much in the field of attitudes towards foreign accents. There is indeed one enormous void here, a black hole unnoticed for too long now: the examination of the attitudes of informants from a dominant language community towards the accent of fellow L1-speakers speaking a mutually *foreign* language. Studies of the attitudes of minority ethnic groups towards in-group members speaking the language—their *second* language—of the dominant out-group, where the results usually reveal that low self-esteem leads to self-deprecation and self-abasement (e.g. Chicanos/Mexican Americans: Flores and Hopper 1975; Ryan & Carranza 1975; Ryan et al. 1975, 1977), are not the same thing. Nor are studies involving the attitudes of members of a dominant language community towards accented members of minority groups within that community. Nor are studies of the accented speech of the complementary language community within a bilingual/diglossic/bicultural society. Attention has yet to turn towards how L1 listeners perceive L1 speakers from the same community as themselves when speaking a foreign language in a *foreign* language situation,

---

[20] Brennan et al. (1975:33) unusually makes reference to this point.

[21] Cunningham-Andersson & Engstrand (1989) is a notable exception.

e.g. Anglo-Americans speaking Spanish and Scots speaking French. How different are attitudes towards such speakers from those towards them when they are speaking the common mother-tongue? How sensitive are L1 listeners to the non-native L1-accented 'foreign' pronunciation of L1 speakers speaking an L2? Is there derogation or approbation of such speakers? How much embarrassment or pride is felt at the L2 accent skills of L1 co-speakers? Many such questions seem never to have been raised, let alone addressed. Nevertheless, they are as important for foreign language teachers and learners as those concerning the ways foreigners perceive foreign-accented speakers.

Teachers still have little to work on when advising students as to their pronunciation; students still have nothing to refer to to know what effect their foreign-accented speech has on their interlocutors, whether they be L2 or L1 listeners. What we need are facts as to which mispronunciation features are disliked more than others and which features affect the perception of personality in which ways. Studies with second-language accents form a good foundation for the work that still needs to be done: investigations of foreign accents, especially of unidentifiable foreign accents. However, such examinations into the complex socio-psychological reactions of listeners to foreign-accented speech perhaps need to be pursued in conjunction with the development of more investigatory techniques. The two experiments which are reported in the present study are one small contribution to this field.

5. AIMS OF THE PRESENT STUDY

> L'accent du pays où l'on est né, demeure dans l'esprit et dans le coeur comme dans le langage.
>
> —La Rochefoucauld, *Maximes*

The problems associated with investigating and measuring attitudes in general, with the use of the matched-guise technique for accessing attitudes towards accents and with the statistical techniques available for analysing data about attitudes are well-known. Nevertheless, as long as the problems are recognised and addressed, as long as the limitations of the investigatory techniques are acknowledged, and as long as circumspection is maintained in drawing one's conclusions from the data gathered, it is still possible to use the MGT, or modifications of it, to conduct valuable research investigating attitudes towards accented speech. Thus the main aim of the present study was to examine people's reactions to and evaluation of Finnish-accented English. I decided to follow two tracks: error evaluation, a study of the pronunciation mistake itself, and speaker evaluation, a study of the image created by the mispronunciation. Thus the study falls clearly into two complementary halves: Experiment 1 and Experiment 2.

5.1. Error evaluation / Alternative mispronunciations

The aim of the first study, Experiment 1, was to test the reactions of Finnish-speaking, Swedish-speaking and English-speaking (British) listeners to male and female Finnish voices saying various English sentences in two different Finnish-accented ways. Listeners would be forced to choose which of two ways of saying a sentence is better, the null hypothesis being that the preferences would be evenly distributed, 50/50. The range of sentences and their versions would embrace six phonemic consonantal errors typical of Finnish-accented English, and the intention would be to integrate the results of the separate dichotomous choices into an overall error hierarchy that would indicate whether one particular error is more acceptable (i.e. less serious) than another. The variables involved would be:

Independent variables: phonemic error; gender of speaker.
Dependent variable: response better/worse.
Controlled variables: listener's mother tongue, gender, age.

Notice that whereas most studies of attitudes utilising the MGT focus attention on the image of the speaker as perceived through the speech or accent, this study focuses uncompromisingly on the mispronunciation, the error, not on the speaker. Moreover, by presenting listeners with alternative pronunciations of a

sentence one after the other, the study is not faithful to the classic model of MGT studies.

The specific research questions for which answers would be sought are:

1) Which phoneme errors are tolerated more readily, which less readily? Specifically, what is the hierarchy of phoneme error tolerance as far as / v, θ, ʃ, p, tʃ, dʒ / are concerned?
2) Is the tolerance of Finnish-speaking listeners, Swedish-speaking listeners and English-speaking (British) listeners the same or different?
3) Are there particular patterns of tolerance associated with the gender and age of the listener?
4) Does the gender of the speaker prompt different patterns of tolerance?

5.2. Speaker evaluation

The aim of the second study, Experiment 2, was to examine, with the use of fourteen 7-point bi-polar adjective scales, how Finnish-speaking and English-speaking (British) native listeners evaluate the personality traits of male and female Finnish speakers reading a standard English text with various degrees of Finnish-accentedness. Each of five speakers would present the same text in two different ways, more and less Finnish accented as measured in terms of the total number of consonantal mistakes committed for the text. The various versions of the five speakers would overall represent four levels of accentedness labelled for reference purposes as 'excellent,' 'good,' 'fair' and 'poor' (see section 6.1.2.2 for details). It was hypothesised that the reactions of listeners would be influenced by various factors:

Independent variables: degree of accentedness; gender of speaker.

Dependent variable: responses on fourteen 7-point ordinal scales.

Controlled variables: listener's mother tongue, gender, age.

This time, as with most studies using the MGT, the voice itself would *not* be evaluated in the main part of this study (unlike Experiment 1) but rather the voice would be used as a means of evaluating (indirectly) the personality (the personality traits) of the speaker. Moreover, in a novel methodological departure from most if not all previous MGT studies, the study would be conducted with two sample groups from the population under investigation, and listeners in one sample would hear one version of each of the five speakers, and listeners in the other group would hear the second version. The null hypothesis would be that there is no difference between the two groups in the evaluations of the different versions of the speakers that they hear.

Two preliminary research questions would be addressed before the main part of the experiment:

1) How good are informants at guessing the age of Finnish speakers speaking English with various degrees of Finnish accentedness?
2) Do evaluations of Finnish-accented English on a 7-point scale ranging from 'awful' to 'excellent' reflect the real degree of accentedness of the text (as measured in terms of the number of specific phonemic errors)?

If the answer to this second question is negative (i.e. if informants cannot identify a particular reading of a text as being more or less Finnish-accented), then there is no basis for continuing with the study, nor for posing the following four research questions that the study would seek answers for:

3) Is the degree of Finnish-accentedness of English-speaking Finns reflected in the personality trait ratings given them on bi-polar adjective scales?
4) Do Finnish-speaking listeners and English-speaking (British) listeners rate the personality traits of Finns speaking English with different degrees of accentedness in the *same* way or in *different* ways?
5) Does the gender of the evaluating listener affect the rating given the speaker?
6) Does the age of the evaluating listener affect the rating given the speaker?

6. METHOD

> These things belong only to pronunciation, which is the least part of grammar.
>
> —George Eliot, *Middlemarch*

For the sake of economy I shall present the method for the two experiments which I conducted as far as possible together. The materials that were used for the testing (the texts for the recordings; the speakers, the tapes and the editing; the questionnaires) will have to be handled separately, but the informants who were interviewed (the selection of informants, their biographical background and the problem of socio-economic classification, the protocol adopted for the interviews, the equipment used) can be presented as one section.

6.1. Materials

6.1.1. Experiment 1: Materials for error evaluation/alternative mispronunciations

6.1.1.1. Texts

Twenty-three sentences were constructed to enable two contrasting mispronunciations which would clearly illustrate two of six consonant articulations of English typically prone to mispronunciation by Finnish learners: / v, θ, ʃ, p, tʃ, dʒ /.[1] The mispronunciation of these phonemes is, in my opinion, characteristic

[1] A note on labels for places of articulations in English and on transcription is essential here. I shall use the symbols / ʃ /, / tʃ / and / dʒ / to indicate respectively the fortis palato-alveolar sibilant and the palato-alveolar affricates in English. The tradition for many years has been to describe these English sounds as having a *palato*-alveolar place of articulation (e.g. Ladefoged 1975:141; Roach 1991:48–52; Cruttenden 1994:160–2,169–73), though a recent innovation is to label them as *post*alveolar (e.g. Jones et al. 1997:x). Notice that the IPA uses the symbols / ʃ /, / tʃ / and / dʒ / to indicate standard (i.e. non-language specific) postalveolar articulations, and / ç / and / ʝ / to indicate palatal articulations. Since it is generally claimed that the Finnish language 's'-sound is post-alveolar (i.e. [ʃ] if one uses the standard IPA symbols), I find it more convenient to describe English / s / and / ʃ / as respectively alveolar and palato-alveolar, so keeping three clearly separate terms for the three clearly different articulations. Now since the typical Finnish speaker's mispronunciation of English / ʃ / falls, I would claim, *between* English / s / and / ʃ /, I suppose logically, considering the matter from the English speaking listener's viewpoint, one could label the misarticulation as fronted palato-alveolar, and transcribe it as [ʃ̟] (the articulation sounds 'too fronted'). However, I prefer to consider the matter from the Finnish speaker's viewpoint, label the mispronunciation as postalveolar and use—for the sake of convenience (and clarity)—the symbol [s], or, if a closer transcription is required, [s̠], a retracted alveolar articulation. (See Laver 1994:410 for a brief presentation of denti-alveolar (fronted/advanced alveolar) [s̪] or [s̟], standard alveolar [s] and retracted alveolar (pharyngealised; postalveolar) [s̠].) I shall follow a parallel convention for the typically fronted mispronunciation of the English palato-alveolar affricates / tʃ / and / dʒ / as produced by Finns:

of Finnish-accented English,[2] but nothing should be read into the order of my list. The only reference to support my opinion is Mazzarella (1971), but I have no data and no references purporting to show how frequently each of these phonemes is mispronounced by which Finns at what stage of learning. As for the frequency of *correct* occurrence of these phonemes in a 'general text' in native (British) speech, Cruttenden (1994:196) suggests the following figures: / v / 2.00%, / θ / 0.37%, / ʃ / 0.96%, / p / 1.78%, / tʃ / 0.41%, / dʒ / 0.60%. I could have chosen other features of mispronunciation (particularly the production of / ð / as a stop, or of dentalised / t /, the occurrence of the standard fricative / ð / and of the alveolar / t / in RP being 3.56% and 6.42% respectively), but I did not as I had to draw the line somewhere. Other pronunciation teachers may of course disagree with me on my choice (particularly as British listeners may consider dentalised / t / as being effeminate in male speakers and childish in female speakers in the same manner as dentalised alveolar sibilants: Wells 1982:21–2).

The phonetic description of the typical Finnish-accented misarticulations is seen in Table 6.1.

Table 6.1 Finnish-accented misarticulations of RP phonemes

	RP articulation	Finnish-accented misarticulation
/ v /	lenis labiodental fricative	voiced labiodental frictionless continuant [ʋ]
/ θ /	fortis dental fricative	voiceless (unaspirated) dental plosive [t̪]
/ ʃ /	fortis palatoalveolar sibilant	voiceless postalveolar sibilant [s̪]
/ p /	fortis (aspirated) bilabial plosive [pʰ]	voiceless (unaspirated) bilabial plosive [p]
/ tʃ /	fortis palatoalveolar affricate	voiceless postalveolar affricate [t̪s̪]
/ dʒ /	lenis palatoalveolar affricate	voiced postalveolar affricate [d̪z̪]

Each of the 23 sentences afforded 3 contrasting pairs of pronunciation: accurate v. phonemic mistake 'X'; accurate v. phonemic mistake 'Y'; mistake 'X' v. mistake 'Y.' For example, (a) v. (b); (a) v. (c); (b) v. (c):

(a) / ˈdʒɒn eɪt ə ˈlʌntʃ ðət wəz ˈmʌtʃ tuː ˈlɑːdʒ /

(b) / ˈd̪z̪ɒn eɪt ə ˈlʌntʃ ðət wəz ˈmʌtʃ tuː ˈlɑːd̪z̪ /

(c) / ˈdʒɒn eɪt ə ˈlʌnt̪s̪ ðət wəz ˈmʌt̪s̪ tuː ˈlɑːdʒ /

However, to avoid ending up with an unwieldy questionnaire of 69 pairs, a limited set of 30 contrasting pairs was established: all the major contrasts of one phonemic mispronunciation versus another phonemic mispronunciation were maintained (giving 15 pairs); eight pairs were repeated with a second different sentence (15 + 8 = 23 semantically different sentences); and six of these 23

I shall label the misarticulations as postalveolar (rather than as fronted palato-alveolar) and transcribe them as / ts / and / dz /, or more closely [t̪s̪] and [d̪z̪].

[2] For a full description of the segmental mispronunciation of English by speakers of Finnish see Morris-Wilson (1992). See also e.g. Wiik (1965), Tommola (1975), Moisio & Valento (1976), Suomi (1976, 1979, 1980), Lehtonen & Koponen (1977), Lehtonen et al. (1977), Hänninen (1979), Lamminmäki (1979), Niemi (1979), Marjomaa (1985a, 1985b), Laaksonen & Lieko (1988).

sentences were chosen for presentation of the six mispronounced phonemes set against an error-free version (the mispronounced / p / being set twice): 23 + 6 + 1 = 30. The final set of sentence pairs thus consisted of 30 contrasting pairs, the maximum number I thought possible to present to informants without the test becoming tediously long (see Appendix A).

The meaning of the sentences was kept as 'neutral' as possible (uncontentious, non-provocative), though it is of course arguable that no text is ever 'neutral' (see section 4.5.1.5.).

The order of the 30 pairs of sentences was randomised but manipulation proved necessary to ensure that no two consecutive pairs of sentences contained the same mispronunciation feature (e.g. no consecutive pairs both illustrated the mispronunciation of / tʃ /). This manipulated randomised order provided the set used for the questionnaires (see Appendix B).

6.1.1.2. Tapes

Thirteen first and second-year university students of the Department of English of Oulu University, Finland, seven men and six women, attempted to record the 23 sentences in the three different ways required. None of the speakers had noticeable voice defects, or a cold; all had normal voice quality, none having a creaky or breathy voice, etc.

Efforts were made to keep the age of the speakers as similar as possible in order to neutralise the effect of the real age of the speakers on the evaluation of the alternative mispronunciations produced, the supposition being that a particular mispronunciation may be acceptable, for example, in a young person but not in an old person (a hypothesis not tested in the present study). However, the necessity of finding Finnish people who could produce both authentic-sounding RP and specific authentic-sounding Finnish-accented mispronunciations of English resulted in the age of the speakers varying quite widely. The average age of the speakers on the two tapes which were made (see below) is presented in Table 6.2.

Table 6.2 Average age of the speakers on the tapes

	N	Average age	sd	Range
Tape 1				
All	13	22.3	1.86	6.2
Women	6	21.6	1.56	3.9
Men	7	23.0	1.97	5.0
Tape 2				
All	12	22.4	1.90	6.2
Women	5	21.7	1.73	3.9
Men	7	23.0	1.97	5.0

The recordings were made in the sound-proof recording studio of the Department of Finnish, Saami and Logopedics of the University of Oulu, using a Casio DA-7 portable DAT (digital audio tape) recorder and a Sony electric condenser

microphone (ECM-55B).[3] Care was taken to ensure the highest quality of recording as possible (N.B. the concern of Sebastian and Ryan for this point: 1985:124 ff.). Some sentences and some phonemic mispronunciations proved more difficult to produce than others, and some speakers couldn't produce some of the required mispronunciations, but eventually, after careful coaching, much laughter and sudden strings of expletives, a sufficient number of natural-sounding versions of the required mispronunciations were obtained. Considerable time, effort and attention was paid by the speakers and the recordist (me) into ensuring that as far as humanly possible the pairs of contrasting sentences were as identical as possible for pitch range and variation, tempo, volume, rhythm and stress placement, and differed essentially only in their intentionally mispronounced phonemes.

It should be noted that the difficulties that the students experienced in producing the desired mispronunciations of the texts were not of the same kind as native speakers may have when trying to produce two or more regional and/or social accents of their mother tongue. Whereas variations in regional and social native accents cover a whole gamut of phonetic detail and articulatory settings, and may therefore necessitate a change of several details within, let's say, a ten-word phrase, the present speakers had to change *only* one or two phonemes within a ten-word phrase. The task was similar to that of a native English speaker saying 'The puppy barked at the purple bubble' and then saying 'The buppy barked at the burple bubble.' Everything except the intended consonantal mispronunciations is held constant. The speakers found the recording task difficult simply because they were unused to producing 'mistakes' to order, and producing them for the teacher who had always insisted on them producing a *non*-Finnish-accented pronunciation. It was the unfamiliar element of purposeful mispronunciation which caused the need for numerous repetitions and led to the hilarity, frustration and swearing. Occasionally speakers just could not produce a specific mispronunciation because their native-like pronunciation was too well established; and occasionally speakers stumbled into producing more than the one mispronunciation wanted in a phrase, the desired single mispronunciation contaminating other phonetic features. However, whenever a particular sentence appeared to be causing insurmountable problems, that sentence was dropped and the next one was tackled, and eventually each student produced sufficient contrasting recordings to allow me to go ahead with the tape-editing.

A master tape (Tape 1) was made using the *SoundEdit 16* computer program (Version 1.0) published by Macromedia on a Power Macintosh (16 bit resolution for the sound; 44 kHz sound sampling rate). For the sake of variety, men's and women's voices were distributed throughout, no two consecutive pairs of sentences using the same speaker, and no more than two consecutive pairs of sentences having the same gender of voice. No speaker appeared on the tape more than three times, i.e. of the 23 sentences that the students tried to produce

---

[3] My sincere thanks to Pentti Körkkö of the Department of Finnish, Saami and Logopedics of the University of Oulu for assisting me in the setting up of the equipment.

in three different ways, the contrasting pairs of only three sentences at most per speaker were used.

A second master tape (Tape 2) was made with a male voice replacing a female voice and vice versa.

In addition, in a misguided attempt to minimise the effect of order of presentation on informant reactions, Tape 2 was also made by *reversing* the order of Tape 1. In other words, 'Tape 1 Sentence 1 *male* voice' did not become 'Tape 2 Sentence 1 *female* voice' but 'Tape 2 Sentence *30* female voice,' and so on, and a correspondingly reversed questionnaire was also drawn up (Questionnaire 2). This was a methodological mistake: the two tapes produced and used were not two identical series of test sentences with simply the gender of voices inverted, but two entirely separate tapes. Thus although the intention had been to have two tapes clearly contrasting merely on the gender of the voices, the tapes in fact afford no safe check on the influence of the gender of the speakers since the responses are open to possible contamination from the order of presentation of the speakers. It would have been better to have four tapes: Tape 1a, male voices, forwards; Tape 1b the same but in reverse order; Tape 2a, female voices, forwards; Tape 2b, the same but in reverse order (cf. the use by e.g. Anderson-Hsieh et al. (1992) of only male speakers).

However, I shall make the major (perhaps unjustified) assumption that the influence of the order of presentation of test sentences is minimal and has no serious or significant effect on the reaction of listeners. In addition, since listeners were assigned randomly to hear either Tape 1 or Tape 2, i.e. a male or female voice presenting a particular contrast of mispronunciation, I shall assume that it is safe for analysis purposes to use between-samples statistical tests.

Obvious pauses in mid sentence were edited out, as also laughter, swearing, hiccups, smacking of lips and such like, but pitch and tempo were left untouched in the editing process and were *not* manipulated. Thus the same sentences with the same phonemic mistakes as produced by the male voice on Tape 1 and the female voice on Tape 2 (and vice versa) were different in terms of tempo and syllables per second. Volume was consistently adjusted to sit comfortably at the optimum recording level for re-recording onto the master tape.

The master tapes included audible prompts ('Number 1,' 'A,' 'B,' 'A,' 'B'; 'Number 2,' 'A,' 'B,' 'A,' 'B'; etc.) to help informants keep track of where they were, a woman's voice being used for all the tapes (one woman for the English version, and another for both the Finnish and Swedish versions). In short 6 master tapes were made: two for English informants (E1 and E2), two for Finnish informants (F1 and F2) and two for Swedish-speaking informants (S1 and S2), E2, F2 and S2 being the opposite gender (woman for man, and man for woman) and the reverse order (end to beginning, and right to left) of E1, F1 and S1 respectively. Each tape lasted approximately 14½ minutes. Informants were to be divided into two groups at random and would hear either Tape 1 or Tape 2, not both. Phonetic transcriptions of the experiment sentences on Tape 1 and Tape 2 are presented in Appendix A.

6.1.1.3. Questionnaires

The questionnaires were constructed using the *PinPoint 3 for Windows* computer program published by Longman Logotron (see Appendix B).

An introductory set of instructions was written to explain to informants what they were being asked to do; informants were to read these instructions, but would not hear them on the tape. Informants were specifically instructed '[not to] worry about the meaning of the sentences. Just listen to the voice and to the pronunciation, and say which version sounds better,' though naturally one cannot determine to what extent informants succeeded in blotting out the meaning of the sentences they heard. English, Finnish and Swedish versions of these instructions were written for the Questionnaires destined for the English-speaking, Finnish-speaking and Swedish-speaking informants, though the texts of the sentences they would hear remained of course in English.

Each of the thirty sentences to be listened to was printed once with two boxes alongside marked A and B: informants were asked to put a cross in the box corresponding to the version they thought was 'better.'

Two versions of the Questionnaire were used for each of the three language-informant groups. Questionnaire 2 presented the test sentences of Questionnaire 1 in reverse order, and thus it corresponded to the order of presentation as heard on Tape 2. Informants, who were assigned Tape 1 or Tape 2 at random, were naturally given the appropriate Questionnaire 1 or Questionnaire 2.

Neither the prompt voice on the tape nor the written instructions on the questionnaire paper provided information as to who said the sentence, to whom, where, when and why: all the sentences were presented 'out of context.' There was no locus of discourse. In addition, it was obvious from the wording of the instructions, from the presentation of the answer boxes, and from the tape recordings themselves (the consecutive presentation of the two versions of each sentence, A, B, A, B) that the same speaker was saying the two versions for evaluation of the same sentence. Thus, strictly speaking, Experiment 1 was *not* a study using the matched-guise technique (see the beginning of section 4.2).

6.1.2. Experiment 2 : Materials for speaker evaluation

6.1.2.1. Texts

A short text of English prose (9 sentences, 110 words, 141 syllables) was carefully constructed in order to contain 47 instances of 7 specific phonemes which typically cause problems for Finnish speakers of English (see Appendix C). The phonemes included the six used in Experiment 1 (/ v, θ, ʃ, p, tʃ, dʒ /) plus / ð /: see Table 6.3, P. 115. When *mis*pronounced by native speakers of Finnish, these RP English phonemes are usually produced in the way illustrated for Experiment 1 (see Table 6.1, P. 110), where in addition the RP lenis dental fricative / ð / is mispronounced as a lenis alveolar plosive [d].

The length of the text was strongly influenced by the time factor: if the version produced by each speaker is repeated once, and if there are five speakers, how long will the whole tape last? Since informants will hear the text ten times, how long can informants be reasonably expected to survive before their tolerance

Table 6.3 The number of tokens of each phoneme appearing in the whole text

RP articulation		Number of tokens
/ v /	lenis labiodental fricative	8
/ θ /	fortis dental fricative	5
/ ð /	lenis dental fricative	10
/ ʃ /	fortis palatoalveolar sibilant	6
/ p /	aspirated fortis plosive	7
/ tʃ /	fortis palatoalveolar affricate	7
/ dʒ /	lenis palatoalveolar affricate	6

wanes? Basically the text needs to be long enough to give an impression of the speaker, but not longer than is necessary to fill in the questionnaire.

The choice of phonemes to be mispronounced and the number of tokens of each phoneme in the whole text was arbitrary, though strongly determined by my twenty years of teaching experience in the classroom. The exclusion of alveolar / t / (frequently mispronounced as dental [t̪] in the speech of Finnish speakers of English) was one of several regrettable omissions though necessary because of the constraints of time.

The meaning of the text was once again kept as 'neutral' or non-provocative as possible.

6.1.2.2. Tapes

Nine undergraduate university students of the Department of English of Oulu University, Finland, attempted to record the text in two of four different Finnish-accented ways so as to provide between them an overall range of four versions: 'excellent,' 'good,' 'fair' and 'poor'. Only five students managed to perform the task satisfactorily (for details of their gender and age see Table 6.4 and Table 6.5 below), and only their recordings were retained for editing for the master-tape. Unfortunately the 'fair' version ended up being split between a version (a) and a version (b) which were almost identical but not quite, the unfortunate result of not being able to arrange further recording sessions either to 'corrupt' a correct

Table 6.4
Overall age of speakers

	N	Age	sd
All	5	23.3	3.0
Women	3	21.3	1.5
Men	2	26.3	0.3

Table 6.5
Age of individual speakers

		Age
Women		
	x	23.0
	y	20.1
	z	20.75
Men		
	a	26.1
	b	26.5

pronunciation or to correct a mispronunciation. Thus the experiment was in fact conducted with five different 'versions' representing five different levels of Finnish-accentedness, though only four 'labels' ('excellent', 'good', 'fair' and 'poor') are used to refer to the basic levels of Finnish-accentedness. The number of tokens of each phoneme mispronunciation in each version, and the ratio of each particular phoneme mispronunciation to the others, was purely arbitrary (see Table 6.6, and Appendix C for the full transcripts).

Table 6.6 Total number of phonemic mispronunciations in each version

Version	Total No. of mispronounced phonemes	No. of tokens of each mispronounced phoneme						
		/v/	/θ/	/ð/	/ʃ/	/p/	/tʃ/	/dʒ/
1 – 'Excellent'	0	0	0	0	0	0	0	0
2 – 'Good'	11	2	1	1	2	0	4	1
3 – 'Fair' (a)	31	3	4	4	6	3	6	5
'Fair' (b)	33	4	4	5	6	4	6	4
4 – 'Poor'	47	6	5	10	6	7	7	6

None of the speakers had any noticeable voice defect, or a cold; all had normal voice quality, none having a creaky or breathy voice, etc.

The recording sessions were held in the sound-proof recording studio of the Department of Finnish, Saami and Logopedics of the University of Oulu, using a Casio DA-7 portable DAT (digital audio tape) recorder and a Sony electric condenser microphone (ECM-55B).[4] Some sentences of the text and some phonemic mispronunciations proved more difficult to produce than others, but eventually, after a certain amount of coaching and much laughter, sufficiently natural-sounding versions of the required mispronunciations were obtained. Again I wish to stress—as I did for the recordings made for Experiment 1 (P. 112)—that the variations in pronunciation that the students were asked to produce were relatively minor in phonetic terms, involving only 3–4 alterations per phrase in the two versions they produced.

The recordings were edited with the SoundEdit 16 computer program on a Power Macintosh (16 bit resolution for the sound; 44 kHz sound sampling rate). Coughs, stutters, repetitions, swearing, smacking of the lips, overlong pauses,[5] etc. were deleted, though tempo was not otherwise manipulated.

Two master tapes were made for each language-informant group (English-speaking and Finnish-speaking listeners), each tape presenting the same five speakers for evaluation in the same order (female, male, male, female, female) but with a different degree of accentedness. Each voice was presented twice in immediate succession so as to give listeners time to react to the voice, to form an

[4] I am indebted once again to Pentti Körkkö for his help in setting up all the equipment for me.

[5] Note that Flege (1988:72, 75–6) claims that the removal of pauses does not affect degree of perceived foreign accent, i.e. he claims that articulation rate is the determining factor, not speaking rate.

impression of the speaker, and to fill in the accompanying questionnaire. The time durations of the ten recorded versions, calculated from the beginning of the first word of the text to the end of the last word of the text and including silences and pauses between sentences and phrases, are shown in Table 6.7 below.

Table 6.7 Time durations of the various recordings

Speaker	Version	Duration (seconds)	Sylls/sec	Wds/min
		Tape 1		
1	1 – 'Excellent'	42.76	3.30	154.3
2	3 – 'Fair'	44.65	3.16	147.8
3	2 – 'Good'	44.01	3.20	150.0
4	5 – 'Poor'	45.18	3.12	146.1
5	1 – 'Excellent'	42.80	3.29	154.2
		Tape 2		
1	3 – 'Fair'	42.79	3.30	154.2
2	1 – 'Excellent'	44.78	3.15	147.4
3	5 – 'Poor'	44.24	3.19	149.2
4	2 – 'Good'	44.77	3.15	147.4
5	4 – 'Fair'	42.79	3.30	154.2

Any possible effects of the order of presentation of the speakers, or of the order of the degrees of accentedness, on listeners' reactions were not catered for. I recognised that the order of presentation of the five speakers is significant in that listeners may progressively become more attuned to the accents, more accustomed to the exercise in hand, and hence more (or less) positive towards the successive speakers. I also recognised that Speaker 1 has no reference point for evaluation: Speakers 2–5 can be evaluated with reference to the preceding speaker or speakers, but Speaker 1 is necessarily 'a shot in the dark' (and the interpretation of the results for Speaker 1 has to bear this in mind). However, because I needed two versions of each speaker, and presentation of only one of these two versions at random to two different groups of informants (Groups A and B), I decided to maintain the same order of presentation of speakers so that any 'order effect' is consistent for both groups of listeners. Otherwise the research design would have required each listener group to be divided in two, with a 'forwards' and 'backwards' tape for each listener group, i.e. Listener Group A with Versions 1–5, Listener Group A with Versions 5–1, Listener Group B with Versions 1–5, Listener Group B with Versions 5–1. This solution was rejected for practical reasons.

No native English speaker voice was inserted for evaluation before or between the five 'test' texts; hence there was no native 'reference' voice for the texts to be

compared with, and no native 'reference' voice to reduce the influence that one non-native voice may have on the next one to be rated.[6]

The master tapes included audible prompts ('Speaker 2, First Reading,' 'Second Reading,' etc. in English or Finnish) to help informants keep track of where they were, a Scottish woman being used for both English tapes, and a Finnish woman for the Finnish tapes.

In short, four master tapes were made: two for English-speaking (British) informants (E1 and E2) and two for Finnish-speaking informants (F1 and F2). Informants were to be given one or other appropriate language version at random, E1 or E2, F1 or F2. Each tape, from first introductory words to final thank you, lasted approximately 9½ minutes.

6.1.2.3. Questionnaires

Two versions of two Questionnaires (Q1 and Q2) were constructed using the *PinPoint 3 for Windows* computer program published by Longman Logotron, one version in English, the other in Finnish. See Appendix D.

The first question of both Q1 and Q2 asked informants to estimate the age of the speaker. A small-scale pilot study conducted with Finnish students had established that separate boxes for every year was too precise and cumbersome a scale for estimating people's ages. Hence boxes were used which corresponded to two-year periods, e.g. 20-21, 22-23, etc. It was anticipated that mode and median would be the most useful indicators of central tendency and distribution, though mean (taking precisely 21.0 years as the value of the 20-21 box, 23.0 years as the value of the 22-23 box, etc.) might also be useful.

The second question of both questionnaires asked for an evaluation of the overall standard of pronunciation of the speaker. A simple 7-point scale[7] with the ends marked by adjectives forming a bi-polar pair (excellent pronunciation v. awful pronunciation) was used, but the poles were reversed for the two questionnaires. No attempt was made to disguise the evaluative nature of the question by using anodyne labels such as 'native-like competence' and 'strongly foreign-accented.'

The third question of both questionnaires consisted of a set of 14 carefully selected bi-polar adjective pairs referring to personality and other characteristics,

---

[6] See Anderson-Hsieh et al. 1992 who catered carefully for this 'reference' issue.

[7] Some people have used 9-point scales (e.g. Munro & Derwing 1994), many have used 7-point scales (e.g. Carranza & Ryan 1975, Ryan & Carranza 1975, Ryan, Carranza & Moffie 1977, Brennan & Brennan 1981a), a few have used 6-point scales (e.g. Arthur et al. 1974, Paltridge & Giles 1984) and some have used 5-point scales (e.g. Dimitrijevic & Djordjevic 1971, Johansson 1978, Cunningham-Andersson & Engstrand 1989). Landén & Trankell (1975: reported in Johansson 1978:16–17) exceptionally used 15-point scales. Flege & Fletcher (1992) have argued convincingly that a 4-point scale is not delicate enough to capture the sensitivity of listeners to accent; and Busch (1993) has argued strongly for the acceptance of 7-point scales as the standard in accent research. Indeed a 7-point standard seems to have established itself. Notice that these scales are often referred to in the sociolinguistic and sociophonetic literature as Likert scales, but they are not: each scale is handled separately, and no one calculates the average of the sum of the scores on the separate items. These scales have more to do with Osgood's semantic differential than to Likert scales, even though the semantic differentials of Osgood's scale are usually factor analysed to discover the limited set of factors or dimensions underlying the whole thing. A better label might be 'Likert-type scales.'

the fourteen chosen traits falling into two equal groups: status/competence traits ('work' traits), and solidarity/benevolence traits ('social' traits). Informants had to indicate on 7-point scales the impressions they had formed of the speakers according to these adjectives. The bi-polar adjective pairs used are presented in Table 6.8 below, the choice being strongly influenced by the work of several authors[8], by popular opinions concerning Finnish stereotype personalities and by translation difficulties.

Table 6.8 Trait labels in English and Finnish

English	Finnish
Status/Competence	
Ambitious – Unambitious	Kunnianhimoinen – Ei kunnianhimoinen
Educated – Uneducated	Sivistynyt/koulutettu – Sivistymätön/kouluttamaton
Hardworking – Lazy	Ahkera – Laiska
Intelligent – Unintelligent	Älykäs – Tyhmä
Rich – Poor	Rikas – Köyhä
Self-confident – Hesitant/unsure	Itsevarma – Epävarma
Solidarity/Benevolence	
Friendly – Unfriendly	Ystävällinen – Epäystävällinen
Honest – Dishonest	Rehellinen – Epärehellinen
Kind – Cruel	Lempeä – Julma
Reliable – Unreliable	Luotettava – Epäluotettava
Serious – Fun-loving	Vakava – Leikkimielinen
Sympathetic – Unsympathetic	Miellyttävä – Epämiellyttävä
Trustworthy – Untrustworthy	Luottamusta herättävä – Ei luottamusta herättävä

The choice of bi-polar adjective pairs used for evaluating listeners' perceptions of speaker characteristics has varied quite widely, and has sometimes been dictated by the contextualisation (real or artificial) of the voices to be evaluated. While my study remains strictly non-contextualised (no information is given as to who is speaking to whom, where, when or why), the choice of adjectives was complicated by the necessity of finding adjectives and characteristics which are acceptable culturally in both Britain and Finland, and adjective pairs which are clearly bi-polar where the 'optimum' point is not in the centre (e.g. medium volume is arguably the 'optimum' of the bi-polar pair 'a loud voice' and 'a soft voice'). The ideal of conducting pre-trials in both Finland and Britain, with listeners describing a set of voices with whatever adjectives they choose and then establishing a set of adjective-pairs based on the most frequently used adjectives, was dismissed as impracticable.

An added complication was the need for adjective pairs which have more or less clear cultural and translation equivalents in English and Finnish. This is not an easy matter: even such apparently simple pairs like *rich–poor* (*rikas–köyhä*) have vastly different semantic fields and socio-cultural connotations in the two

[8] See in particular Carranza & Ryan (1975), Ryan & Carranza (1975), Ryan, Carranza & Moffie (1977), Brennan & Brennan (1981a, 1981b).

languages.[9] Translations from English into Finnish are particularly difficult for two adjective pairs which I felt obliged to use because they have been included in so many studies on attitudes to accents, viz. *reliable–unreliable* and *trustworthy–untrustworthy*.[10] It would appear that the distinction between *reliable* and *trustworthy* is relatively culture-dependent, for both are normally translated into Finnish as *luotettava*: the Finnish language does not readily make a distinction between someone who is dependable (i.e. they will do something because they have said they will) and someone who is worthy of trust (i.e. they will not reveal a confidence). There are no simple monolexemic equivalents in Finnish. Presumably the two concepts have not been lexicalised separately because the culture does not make, and does not need to make, the distinction very often.[11] However, wishing to keep the two pairs of English adjectives, I decided to use *luotettava* for *reliable* and translate *trustworthy* as *luottamusta herättävä* (lit. 'trust/reliability prompting'). In addition I decided to keep the two pairs of adjectives next to each other on the questionnaire in the hope that informants would recognise that I am trying to get at two different personality traits.

A further complication arises when one begins to wonder whether there are some traits of the stereotypical Finn—possibly unique to Finns—which must be included. This consideration led me to include the *self-confident–hesitant/unsure* adjective pair in an attempt to tap (however obliquely) the widely held impression that Finns are shy (*ujo*). Shyness is perhaps the one trait that the stereotypical Finn is universally thought to display. It is rarely (if ever) tested for in sociophonetic research into attitudes towards accents in Britain and America, but it would be foolish and foolhardy to omit some reference to it here. Does this allegedly common Finnish trait come through in the accent? If it does, do both Finns and non-Finns notice it? However, rather than using the adjective poles *shy* and *not shy*, which I felt uncomfortable with, I chose to use a compromise solution (still somewhat clumsy): *self-confident* v. *hesitant/unsure* (*itsevarma* v. *epävarma*). I would contend that this solution partly embraces the two other adjectives frequently used to describe Finns: *diffident* and *self-effacing*.

Similarly I wished to determine whether the general opinion held by foreigners that Finns are somewhat dour and taciturn, not exactly full of fun (except when drunk), is also carried through the accent. This trait is perhaps a further interpretation of the shy trait, or misinterpretation of it: hesitant shyness is

---

[9] I am grateful to several people whom I consulted about this thorny problem of translation equivalence, but I owe particular thanks to Ilkka Marjomaa. The problem is not discussed at all in Paltridge and Giles (1984) despite being a report in English of an experiment held in French.

[10] For example, *reliable–unreliable* is used by Giles, Baker & Fielding (1975), Giles (1971c), Strongman & Woosley (1967), Lambert, Giles & Picard (1975), Asch (1946); *trustworthy–untrustworthy* is used by Stewart & Ryan (1982), Paltridge & Giles (1984), Brennan & Brennan (1981a, 1981b), Ryan & Carranza (1975), Carranza & Ryan (1975), Creber & Giles (1983), Tucker & Lambert (1969); *dependable–undependable* is used by Street, Brady & Putnam (1983), Gallois & Callan (1981), Brown, Giles & Thakerar (1985), Ball, Byrne, Giles et al.(1982), Thakerar & Giles (1981); *dependable–unreliable* is used by Arthur, Farrar & Bradford (1974).

[11] For an interesting discussion of language-specific and culture-specific concepts, and for the lexicalisation of concepts in different languages, see Wierzbicka (1986).

perhaps mistakenly read as reticent withdrawing coldness. Hence the adjective pair *serious-minded* v. *fun-loving* (*vakava* v. *leikkimielinen*).

In a sense there are two facets to the problem of equivalence of questionnaires here. One side of the coin is precision in the linguistic translation of terms generally used to describe personality traits: do the Finnish/English words which one can so easily take as being translation equivalents share in fact the same semantic fields? And the other side of the coin is the problem of capturing those personality traits (possibly stereotypical traits) which are generally thought to be of importance in the socio-cultural environments of Britain and Finland (regardless of whether British or Finnish people think of them as somehow primary within their own socio-cultural environments, and regardless of whether other studies have used them).

Linguistic translation equivalents *are* available, just as personality trait equivalents *are* available, but it must be borne in mind that equivalents (particularly single-word equivalents) are rarely exact. They differ semantically, and they differ in social/cultural import. One possible solution to this issue of cross-linguistic and cross-cultural comparability would be to invent 'a language-independent semantic meta-language' (Wierzbicka 1986:585), but while we cognoscenti might understand the esoteric niceties and subtleties of our private language, our informants would be utterly lost and unable to inform us of anything.

Valo (1994) underlines a closely related further problem, perceptual equivalence, suggesting (Pp. 56–7) that the results of many American studies cannot be generalised to the Finnish speech-communication situation because of socio-cultural differences: what may be considered a positive voice or personality characteristic in the States is not necessarily so in Finland. She also draws attention to the fact that many adjective pairs used in American studies (e.g. *informal–formal*, *fast–slow*, *verbal–quiet*, *talkative–silent*, *adventurous–cautious*, *ambitious–not ambitious* and *dominant–submissive*) are firmly based on a local (i.e. New England, Mid West, etc.) or national (American) ideal of a communicator or radio celebrity and that this ideal does not sit comfortably with the bi-polar positive–negative norms found in Finnish culture:

> Henkilökuva on . . . vahvasti kulttuurisidonnainen. . . . Yhdysvaltalaisen tutkimuksen tulokset eivät siten ole yleistettävissä suomalaiseen puhekulttuuriin. Yhdysvaltalaisten positiiviseksi arvioimaa henkilökuvaa ei välttämättä vastaavassa suomalaisessa kontekstissa pidettäisi suotuisana, koska puhe- ja viestintäkäyttäytymistä arvostetaan eri tavalla. . . . Yhdysvaltalaisten tutkijoiden henkilökuvamittareissaan usein käyttämät piirrenimet *informal–formal, fast–slow, verbal–quiet, talkative–silent, adventurous–cautious, ambitious–not ambitious* ja *dominant–submissive* perustuvat sikäläiseen kommunikoija-ihanteeseen, eivätkä ne toimi positiivisuus–negatiivisuus-ulottuvuuksina suomalaisessa puhekulttuurissa. (Valo 1994:56–7)

Valo points out in particular (P. 57) that unlike the situation in the States, high points scored on *dynamic* and *energetic* do not necessarily strengthen the image of trustworthiness (this problem word again: *luotettavuus*!) conveyed by a

Finnish politician in a television appearance. *Presence of mind* (*cool-headedness*) and *prudence* (*careful thought and reflection*) are valued more highly in Finland:

> Esimerkiksi suomalaisen poliitikon televisioesiintymisen luotettavuusvaikutelmaa eivät välttämättä vahvista 'dynaamisuus' ja 'energisyys' vaan esimerkiksi maltti ja harkitsevuus. (Valo 1994:57)

Valo therefore offers wise words of caution which I have tried to heed. Basically, my two questionnaires, one in English and one in Finnish, are necessarily different linguistically; the English and Finnish adjectives which I have used—supposedly translation equivalents—do not have the same semantic fields; my choice of traits is a compromise between two languages and two different socio-cultural views of what is appropriate; and the informants (Finnish and British) answer their respective questionnaires using as a reference point for their personality ideal a different socio-cultural baseline. The analysis of the data collected and the interpretation of the results must take all this into consideration.

In an attempt to minimise the probable influence of order of presentation of the traits on the questionnaire, the 14 bi-polar adjective pairs on Q1 were presented in reverse order on Q2 (the 1st became the 14th, the 2nd the 13th, etc.) and also inverted mirror-style (i.e. the right-hand pole became the left, and vice versa). This is an important point of my research design, since I assume (correctly or incorrectly) that the creation of the image of a speaker is a step-by-step accumulative process in which a decision taken on one trait will influence the decision taken on a trait to be considered later on. Rifkin and Roberts (1995: 523–4) call this the 'order effect,' and Valo (1994:50) calls it the 'ensisijaisuus-efekti' (lit. 'the effect of the first in line of order').[12] From a methodological viewpoint it is inevitable that the presentation of adjective pairs is sequential, but this sequential build-up of traits may not correspond to the natural formation of person perceptions in real life situations, for there the perception may be global rather than atomistic, some features helping to form the overall image but not consciously being tuned in to. Hence the two questionnaires. Hence a further reason for worrying about the artificiality of the task informants were asked to perform.

The same introductory set of instructions (in English or Finnish) was appended to both questionnaires to explain to informants what they were being asked to do. See Appendix D. Note that similar to Experiment 1 neither the prompt voice on the tape nor the written instructions on the questionnaire paper provided information as to where the text came from, who the two characters referred to in the text were, what their relationship was or anything else: the text was presented 'out of context.'

Informants were to be given the appropriate English-language or Finnish-language Q1 or Q2 at random, irrespective of which Tape they were randomly given to listen to. In effect this provided four random groups of informants for each language group:

---

[12] Valo refers to Smith, Giles & Hewstone (1980), Amabile & Glazebrook (1981), Frandsen & Clement (1984), Wyer & Srull (1989).

1: Tape 1 + Q1
2: Tape 1 + Q2
3: Tape 2 + Q1
4: Tape 2 + Q2

Experiment 2 exploits the classic MGT methodology where the same speaker(s) produces the same text(s) using two (or more) different accents, but unlike most (if not all) previous MGT studies the two accent versions were to be given to two different informant samples. Informants were to hear only one version of each speaker, and until the postmortem informants were totally unaware that some people would hear and evaluate one version and some people a second version.

6.1.3. Statistical note

The major assumption underlying both Experiment 1 and Experiment 2 is that the samples of language being used—the recordings—can be treated as representatives of the complete range of vocal phonetic variants possible, thus allowing generalisation of the results beyond the present data. I neither seek nor offer proof for this claim to representativeness: I did not make *n* recordings with *n* different speakers and then choose the few recordings I needed from the total batch at random; I have no evidence to confirm that different recordings made with different speakers and presented to different informants would produce the same or similar results. This flaw in my base hypothesis—which admittedly is found in nearly all studies like the present one—lays me bare to the accusation of committing a statistical error, what Clark (1973) calls 'the language-as-fixed-effect fallacy.' (See Section 2.2.3.) I treat my language samples as a fixed rather than a random effect. Despite this, rightly or wrongly, I shall, with due reservations, make generalisations.

6.2. Informants and test protocol

6.2.1. Convenience samples

The population (or sampling frame) chosen for investigation for both experiments was 'Finnish/British university students'. However, a probability sample was not feasible for reasons of practical organisation: neither simple random sampling nor stratified random sampling could be used. Hence, a non-probability (quota) sample was aimed at, the basic minimum requirement for inclusion in the study being 'university student' or 'university graduate'. While this minimum requirement was rigorously adhered to, the likelihood of each individual in the quota specification group participating was by no means equal: the essential feature of random selection was not there. The sample, therefore, is best described as **a convenience sample** (or **opportunity sample**) taken from a loosely defined population of unknown distribution.

All informants, therefore, were either university undergraduates (over 97%) or graduates. The Finnish informants covered a wide range of disciplines from the Arts and Sciences (foreign languages, history, logopedia, education, physics, engineering, chemistry, computer sciences, etc.), but the English-speaking

informants were from a much narrower subject background (foreign languages, English, film and media, psychology, sociology). For Experiment 1, error evaluation, the Finnish informants included both Finnish-speaking and Swedish-speaking Finns, but for Experiment 2, speaker evaluation, there were no Swedish-speaking Finns. The English-speaking informants in both experiments came from a wide geographical area embracing the whole of Scotland and England, but there was also a handful of people from Wales and Ireland (plus a few Americans and Canadians who had lived in Britain for several years).[13] For Experiment 2 in order to find English-speaking informants I was obliged to include some adult evening class students attending non-degree Spanish classes.[14]

Although informants were not told the nationality and native language of the speakers they heard on their tape, it is a fair assumption that all the Finns (both Finnish-speaking and Swedish-speaking native listeners) guessed that the speakers were Finns. The English-speaking informants, however, had no idea who the speakers were other than that they were non-native speakers of English: the voices were unidentified foreigners. In Britain I was careful never to divulge my normal place of work, and where hints were occasionally dropped accidentally by contact helpers, I made clear that I taught English abroad, had been teaching in many different countries, etc. I usually added that I was now studying in Stirling after many years abroad. (After a test-session in Stirling a couple of informants came up to me—separately—and said they had been amused by 'the Arabic speaker' and 'the Spanish speaker.')

6.2.2. Co-operation rates

Initially there were no rewards for participation, and although of little consequence in Finland this certainly led to problems in Britain.

The co-operation rate of people approached in Finland (both Finnish and Swedish native listeners) was consistently high, over 99%. Informants were basically compulsory volunteers: teachers gave me the last 20–30 minutes of a lesson/lecture without having warned people in advance that anything unusual would happen. People were given the chance to leave without participating but only one Finnish-speaker (for Experiment 1) did so.

---

[13] Of the 106 English-speaking informants in Experiment 1, two (1.9%) were Irish (though their childhood had been spent in Southeast England) and one (0.9%) was American. Of the 143 English-speaking informants in Experiment 2, six (4.1%) were Irish, four (2.8%) were American and three (2.1%) were Canadians. For the sake of brevity I refer throughout this thesis to the informants as being 'British' or 'GB.' This is not intended as a slight to those nationals of Ireland, the United States and Canada who participated. I am truly grateful to all those non-British nationals who helped me and beg them (and my readers) not to take offence at my shorthand. Similarly I hope those Scots who claimed they were not British but Scottish, and who maintained their mother-tongue was not English but Scots or Lallans, will also forgive me for 'doctoring' the information they offered me. I consciously chose not to use the word 'Anglophone' in order not to lose sight of the fact that this thesis is firmly based in Britain, uses the RP accent as the standard against which the Finnish-accented speech is implicitly contrasted, and uses predominantly British informants.

[14] For Experiment 2, fourteen of the 143 English-speaking informants (9.8%) were adult evening class students: mean age 55.4 years, range 39–69, *sd* 10.02. This necessarily raised the overall mean age to above that of the Finnish-speaking listeners.

The co-operation rate of British informants was in contrast extremely low. All informants were 'willing' volunteers, being invited to participate in the experiment and free to come along or not as they wished. In Hull and Edinburgh approximately 90–95% of students in classes agreed to participate, either inside or outside class time, in Canterbury and Glasgow about 50%. On the other hand, in Stirling the participation rate was only 10–15%. Because of the low response in Stirling during Experiment 1, I handed out mini-Mars bars or mini-Snickers bars as a reward while conducting Experiment 2, but to no avail: the response remained dismally low.

One English informant was obviously 'hostile' to Experiment 1, refusing on two of the alternative pronunciation questions to mark a preference, giving no indication of her age ('over 21') or her language experience background, and writing in a comment that the test was too long and that the repetitions (there weren't any in fact) were unnecessary. Fearing for the validity of her responses, I felt obliged to omit her incomplete data from the study.

6.2.3. Biographical background questionnaire

Once they had completed the evaluation test, all informants were asked to fill in a background questionnaire about their personal/biographical details (see Appendix E). This questionnaire was also written with the use of the *PinPoint 3 for Windows* computer program published by Longman Logotron. Informants were not asked to give their names, and all participants were assured of the anonymity of the exercise. Questions focused on: gender, age, mother-tongue, school/academic background, language skills and experience, and socio-economic background. Finnish and Swedish translations were made for the Finnish-speaking and Swedish-speaking informants respectively, though the questions were slightly different from the English version: whereas the English version referred to foreign-language skills and experience only briefly and in general, the Finnish and Swedish versions referred specifically and in more detail to skills and experience in the English language.

The biological gender of informants was recorded, following the tradition of many sociolinguistic studies, with the aim of discovering whether the gender of language users influences preferences in mispronunciation. However, I am well aware that the approach is seriously flawed: *biological* gender almost certainly has no moderating influence whatsoever on people's linguistic behaviour. It is of course possible that the female brain *is* intrinsically and fundamentally different from the male brain as far as language processing and language perception is concerned (and we all know the jokes and the folklore which thrives on this very assumption), but I remain sceptical. However, I readily admit that biological gender is the easiest (socially unobtrusive) way to access the closely related factor that (I believe) almost certainly *does* flavour people's linguistic behaviour: a person's adopted *gender role* and/or *gender orientation.* The analyses undertaken, therefore, should be interpreted with caution. I am completely unable to

suggest how opaque the window of biological gender is for looking at the *social* gender I wish to observe.[15]

The age of informants was to be used as a control variable, informants 22 years of age and under forming one group, those 23 years of age and over forming another. However, a small-scale pre-trial test-run conducted with Finnish students established that at least Finns find a request for precise date of birth too intrusive. Hence only year of birth was asked for, and age was calculated by subtracting the year of birth they provided from the year of participation in the test. Age, therefore, as thus calculated, is not precisely '22 and under' v. '23 and over': some informants classified as 22 may in fact be 23, and some classified as 23 may be 22. The analyses undertaken using age of informants as a variable are thus based upon only a rough guide to *actual* age of informants, but the degree of imprecision was considered tolerable for the purposes of this study.

The position of the dividing line for 'older' and 'younger' is purely arbitrary. However, the 22/23 border (or even 21/22) is perhaps the usual point of *compulsory* graduation for full-time British students, though one has to bear in mind that there are a considerable (and increasing?) number of 'mature' students in Britain, people who do not start university studies immediately after school but some years later. In Finland, in contrast, most students seem to start university immediately after school (except men who often do their military service before starting their studies), study full-time for 2–3 years, then at 21–22 years of age, perhaps 22–23, they *unofficially* become part-time students with a paid job and a family, and they take—depending on what they *choose* to do—at least five years in all, often 8 or 9 years, sometimes even longer, to graduate. There is no time limit and no compulsory cut-off point for university study in Finland. The socio-cultural life-pattern found in Britain of *first* school *then* university *then* job and family is not mirrored in Finland where the last two stages (university, then job and family) clearly overlap. Whether this difference between the two countries is significant when considering attitudes towards language and accents I have no idea, but I suspect that psychologists, sociologists and culturalists would have a field day with inter-psychological/social/cultural studies of adolescence, adulthood, maturity, independence and so on. This difference (if it indeed exists) needs to be borne in mind when using real chronological age as a control variable, but I cannot investigate it.

All informants whose mother-tongue was neither Finnish, nor English, nor Swedish were systematically excluded from the study.

The data collected from all the informants concerning school/academic background, language skills and experience, and socio-economic background were not analysed for the two present studies. No one (apart from the 'hostile' English informant mentioned above) seemed to be put out by the range and type of questions asked.

---

[15] See Cameron and Coates (1988), and Graddol and Swann (1989; particularly Ch. 2), for a discussion of biological/anatomical gender v. social gender.

Once informants had completed the biographical background questionnaire, a short post-mortem discussion was held, allowing me to explain the general purpose of my study and to thank informants for their co-operation.

6.3. Listening equipment

The listening equipment that was used varied quite considerably but the quality was maintained as high as possible (see Samarin & Kalmar 1979:184 who stress the need for good quality play-back equipment). Approximately 97% of the informants in Finland and 40% of those in Britain were interviewed in groups and used standard language laboratory equipment produced by Auditek, Sony or Tandberg. In all the laboratories, I checked the headsets and tapes of each booth to ensure that the quality of sound production was acceptable (no faulty headphones, no poor quality tape, etc.): this was a rough and ready check on quality, but it was the best that I could do. There is here a possible source of measurement error (and test unreliability). The remaining informants, interviewed individually, used a small portable (Walkman) digital audio tape (DAT) mini-cassette Sony 'tape-corder' with headphones: the quality of sound production was excellent. I shall *assume* that the differences in equipment used had no effect at all on the responses.

7. RESULTS AND ANALYSIS – EXPERIMENT 1 : ERROR EVALUATION / ALTERNATIVE MISPRONUNCIATIONS

> Eh pien, si ces tames feulent fus dennir gambagnie, dit Nucingen, che fus laiserai sèle, gar chai drop manché.
>
> —Balzac, *Splendeurs et misères des courtisanes*

In examining the data collected for Experiment 1, it should be borne in mind that Tape 1 and Tape 2 are two entirely separate tapes: there are essentially two different though complementary versions of the same experiment here (see 6.1.1.2. above). However, since informants were presented at random with either Tape 1 or Tape 2 for evaluation, I shall assume that it is safe for the two data sets to be analysed in parallel. Unfortunately the result is not entirely satisfactory since the number of informants for the two groups was not the same: I am obliged to reveal, repeatedly, which 'tape group' is being referred to so that the sample size is clear. I could have omitted the last one or two informants of particular groups, in order to balance the respective groups, but did not, preferring to keep my informant sample as high as possible.

In addition, my basic approach will be to analyse the data as three separate blocks of data: data from the English-speaking informants, the Finnish-speaking informants and the Swedish-speaking informants. I shall bring the three sets together only for cross-tabulation correlation purposes where informant mother-tongue will act as the independent variable.

7.1. Error free pronunciation v. a particular phonemic error

One might perhaps expect the error free version of a sentence to be preferred always, without exception. However, this is not the case, as Table 7.1, P. 129, reveals.

> (N.B. The following conventions are used in the tables: / Ø / = no phonemic mispronunciation; the contrast / Ø - v / means the contrast between accurate pronunciation and *mis*pronounced / v /. The *first* symbol/phoneme in the contrast column is the 'preferred' or 'better' pronunciation, and the percentage figure given is that of listeners preferring this first phoneme: so for the contrast / Ø - v /, the accurate pronunciation is preferred [by e.g. 96.3% of listeners] and mispronounced / v / is rejected.)

Table 7.1 Percentage of preferences for error-free pronunciations

	Female voices				Males voices			
	GB[a]	SF	SW		GB	SF	SW	
Contrast	%	%	%	Tape	%	%	%	Tape
/∅ - v/	96.3 ***	73.2 ***	70.0 *	T2	100.0 ***	62.6 *	53.7	T1
/∅ - dʒ/	100.0 ***	97.0 ***	97.6 ***	T1	100.0 ***	91.8 ***	100.0 ***	T2
/∅ - tʃ/	96.2 ***	70.7 ***	92.7 ***	T1	94.4 ***	71.1 ***	97.5 ***	T2
/∅ - θ/	94.2 ***	97.0 ***	97.6 ***	T1	92.6 ***	85.6 ***	72.5 **	T2
/∅ - ʃ/	100.0 ***	76.3 ***	100.0 ***	T2	94.2 ***	91.9 ***	97.6 ***	T1
/∅ - p/	98.1 ***	94.8 ***	97.5 ***	T2	88.5 ***	94.9 ***	97.6 ***	T1
	96.3 ***	92.8 ***	92.5 ***	T2	67.3 *	88.9 ***	92.7 ***	T1

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB T1, $N = 52$; SF T1, $N = 99$; SW T1, $N = 41$.
GB T2, $N = 54$; SF T2, $N = 97$; SW T2, $N = 40$.

[a] 'GB' is shorthand for 'English-speaking': the vast majority of informants were from Great Britain, but there were a few Irish and American informants. See section 6.2.1.

Chi-Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Looking first at the two GB columns (English-speaking informants), it is clear that the native English listeners hear the mispronunciations and reject them, though it is also clear that one of the unaspirated mispronunciations of / p / (as produced by the male voice) prompted fewer rejections than the other mispronunciations (67.3%, $p < .05$). The level of acceptance for this male mispronunciation of / p / is so low relative to the Finnish-speaking and Swedish-speaking listeners that crosstabulations performed for mother-tongue of informants (English-speaking v. Finnish-speaking, GB/SF, and English-speaking v. Swedish-speaking, GB/SW) produce significant values for Cramér's *V*: .26 ($p < .01$) and .31 ($p < .01$) (row 5, Table 7.2, P. 131).

(N.B. Whereas the order of phonemes in the contrast column of the tables of preferences for mispronunciations is crucial, it is of course totally irrelevant in all the correlation tables for Cramér's *V*.)

Looking next at the two SF columns for Finnish-speaking informants, the rejection of mispronounced / v / is highly significant for the female voice (73.2%, $p < .001$) but only mildly significant for the male voice (62.6%, $p < .05$). Both of these figures for mispronounced / v / are also so low compared to those for the English-speaking informants (100% and 96.3% respectively) that crosstabulations for the mother-tongue of informants here (English-speaking v. Finnish-speaking: GB/SF) produce relatively high values for Cramér's *V*: .28 ($p < .001$) for the female voice, and .41 ($p < .001$) for the male voice (row 1, Table 7.2, P. 131): the difference is significant.

As for the Swedish-speaking informants, the rejection of / v / is just significant for the female voice (70%, $p < .05$) but not significant at all for the male voice (53.7%, $p > .05$). These figures are also so low compared to those for the English-speaking informants that crosstabulations for English-speaking listeners v. Swedish-speaking listeners (GB/SW) again produce high values for Cramér's *V*: .37 ($p < .001$) for the female voice, and a dramatic .57 ($p < .001$) for the male voice (row 1, Table 7.2, P. 131): the difference is clearly significant.

The rejection of mispronounced / tʃ / is highly significant ($p < .001$) for all three informant groups for both female and male voices (row 3, Table 7.1). However, the level of rejection among the Finnish-speaking listeners is lower than that among the English and Swedish-speaking listeners. As a result the reactions to this / ∅ - tʃ / contrast is significantly different between the English and Finnish-speaking listeners and between the Swedish and Finnish-speaking listeners, but *not* so between the English and Swedish-speaking listeners (row 6, Table 7.2).

Of all the mispronunciations only mispronounced / dʒ / was equally firmly rejected by everyone (more than 91%) for both female and male voices (row 2, Table 7.1). Crosstabulations (using Cramér's *V*) for mother-tongue of informants (English v. Finnish; English v. Swedish; Finnish v. Swedish) confirmed this unanimity, revealing no significant correlations here at all (last row, Table 7.2, P. 131).

Table 7.2 Correlations (Cramér's *V*) of preferences for error-free pronunciation to mother-tongue of informant

	Female voices				Male voices			
Contrast	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape
/∅-v/	.28 ***	.37 ***	.03	T2	.41 ***	.57 ***	.08	T1
/∅-θ/	.07	.08	.02	T1	.10	.27 **	.15	T2
/∅-ʃ/	.32 ***	—[a]	.29 ***	T2	.04	.08	.10	T1
/∅-p/	.08	.02	.06	T2	.12	.17	.06	T1
	.07	.08	.00	T2	.26 **	.31 **	.06	T1
/∅-tʃ/	.30 ***	.08	.24 **	T1	.28 ***	.07	.29 ***	T2
/∅-dʒ/	.10	.12	.02	T1	.18	—[a]	.16	T2

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB/SF T1: $N = 52/99$; GB/SW T1: $N = 52/41$; SF/SW T1: $N = 99/41$.
GB/SF T2: $N = 54/97$; GB/SW T2: $N = 54/40$; SF/SW T2: $N = 97/40$.

[a] NDA: No value calculated for Cramér's *V* since percentages identical (100%:100%).

*** $p < .001$; ** $p < .01$.

7.2. Hierarchies of tolerance towards mispronunciations

The null hypothesis proposes that all the mispronunciations are equally preferred, that the six mispronounced phonemes appear on a level scale with no phoneme rising 'above' or falling 'below' the others through preference or rejection. If, however, clear patterns of preference *are* manifested, with each mispronounced phoneme being more or less tolerated than the other five, then a hierarchy of six levels will appear, with each mispronounced phoneme fixed in that hierarchy by five 'anchors' (giving a total of 15 anchors). The more levels there are in the hierarchy, and the more anchors binding the mispronunciations into the hierarchy, the more marked or well-defined the attitudes towards the mispronunciations.

The data did in fact reveal clear hierarchical patterns of preference (tolerance/ intolerance) towards the six mispronounced phonemes. This is best presented and appreciated initially by focusing attention on one phoneme, / v /, before examining the other mispronunciations, for the main finding in this study of alternative mispronunciations was a striking difference in the hierarchy of tolerance towards mispronounced / v / between the English-speaking listeners on the one hand and the Finnish and Swedish-speaking listeners on the other hand.

7.2.1. Mispronounced lenis labiodental fricative / v /[1]

The English-speaking listeners (both groups, Tapes 1 and 2) clearly disliked the Finnish-accented mispronunciation of / v /, i.e. [υ], many people (unsolicited by me) even remarking on the mispronunciation in the postmortem after the test. Mispronounced / v /, regardless of whether a man or a woman was speaking, was consistently rejected when contrasted with the other five mispronounced phonemes, and the Binomial test and Chi Square analysis revealed rejection at the $p < .001$ level for all five pairings. See Table 7.3, P. 133 and Table 7.4, P. 134.

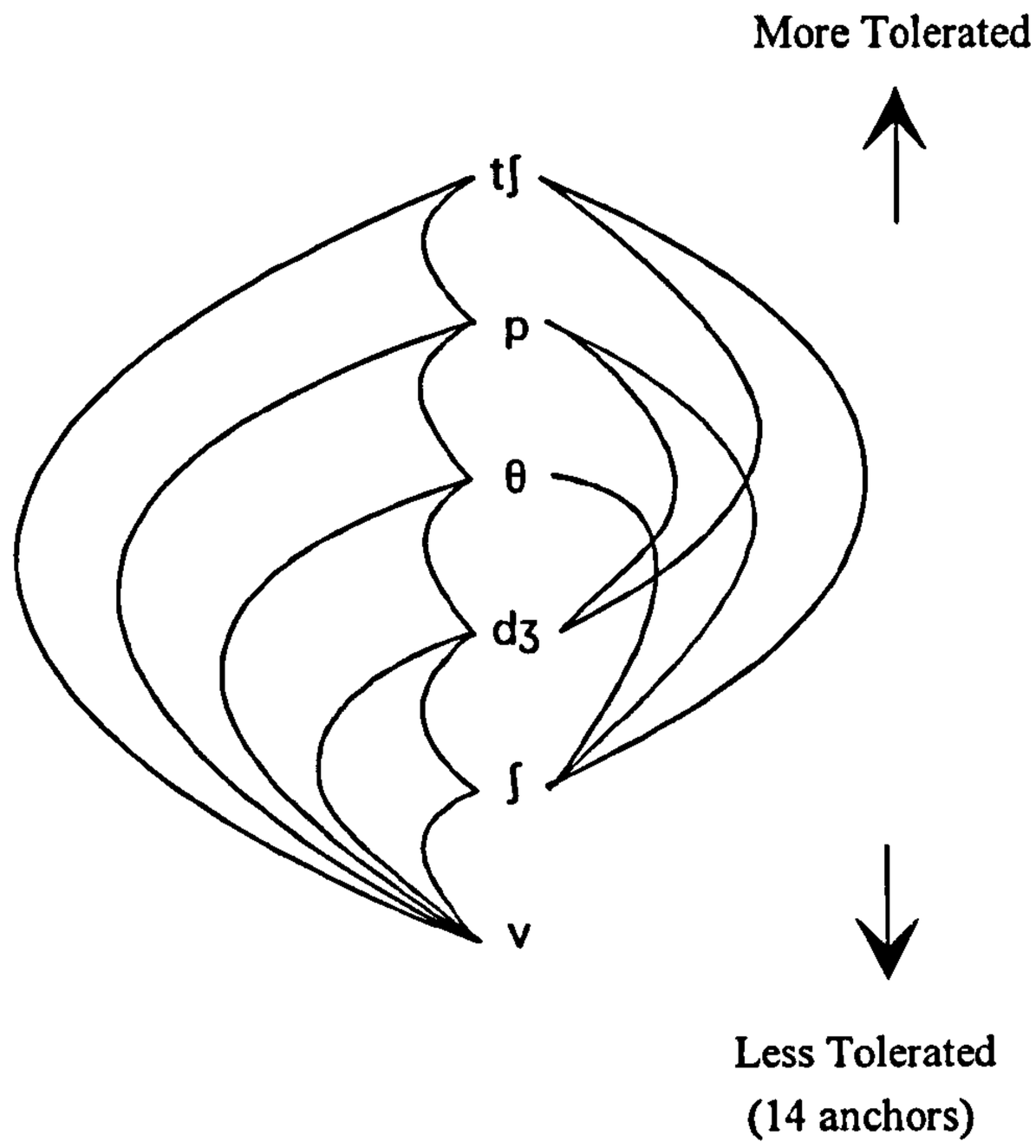
> (N.B. the contrast / θ - ʃ /, row 1 Table 7.3, means the contrast between *mis*pronounced / θ / and *mis*pronounced / ʃ /. The *first* phoneme in the contrast column is the 'preferred' or 'better' pronunciation, and the percentage figure given is that of listeners preferring this first phoneme. The diagram presented—for the sake of convenience—within each table represents the logical hierarchy resulting from the 'anchors' between the various mispronounced phonemes: e.g. if A is preferred to both B and C, and if C is preferred to B, then the logical hierarchical order is necessarily: A → C → B. Thus, since / θ / is preferred to both / ʃ / and / dʒ /, while / dʒ / is preferred to / ʃ /, the logical hierarchical order here is / θ / → / dʒ / → / ʃ /.)

These first two tables show that the hierarchy for the phonemes is clear-cut for the female voices (14 statistically significant relationships, or 'anchors', and 6 levels), but is not so well defined for the male voices (12 anchors, 5 levels) where mispronounced / tʃ / and / ʃ / would seem to be equally acceptable. While

[1] Most of the results reported in this section 7.2.1. are presented in Morris-Wilson (1997).

Table 7.3 English-speaking listeners listening to female voices

Contrast	%	Tape
/ θ - ʃ /	100.0 ***	T2
/ p - v /	100.0 ***	T2
/ p - ʃ /	98.1 ***	T2
	98.1 ***	T2
/ p - dʒ /	98.1 ***	T1
/ dʒ - v /	98.1 ***	T1
/ ʃ - v /	96.3 ***	T2
	68.5 **	T2
/ tʃ - dʒ /	96.2 ***	T1
	65.4 *	T1
/ tʃ - v /	94.2 ***	T1
	92.3 ***	T1
/ θ - dʒ /	88.9 ***	T2
/ θ - v /	88.5 ***	T1
/ tʃ - p /	76.9 ***	T1
/ p - θ /	73.1 ***	T1
/ tʃ - ʃ /	70.4 **	T2
/ dʒ - ʃ /	70.4 **	T2



Notes: For T1, $N = 52$; for T2, $N = 54$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: / tʃ - θ / 73.1 *** T1
/ θ - tʃ / 68.5 ** T2

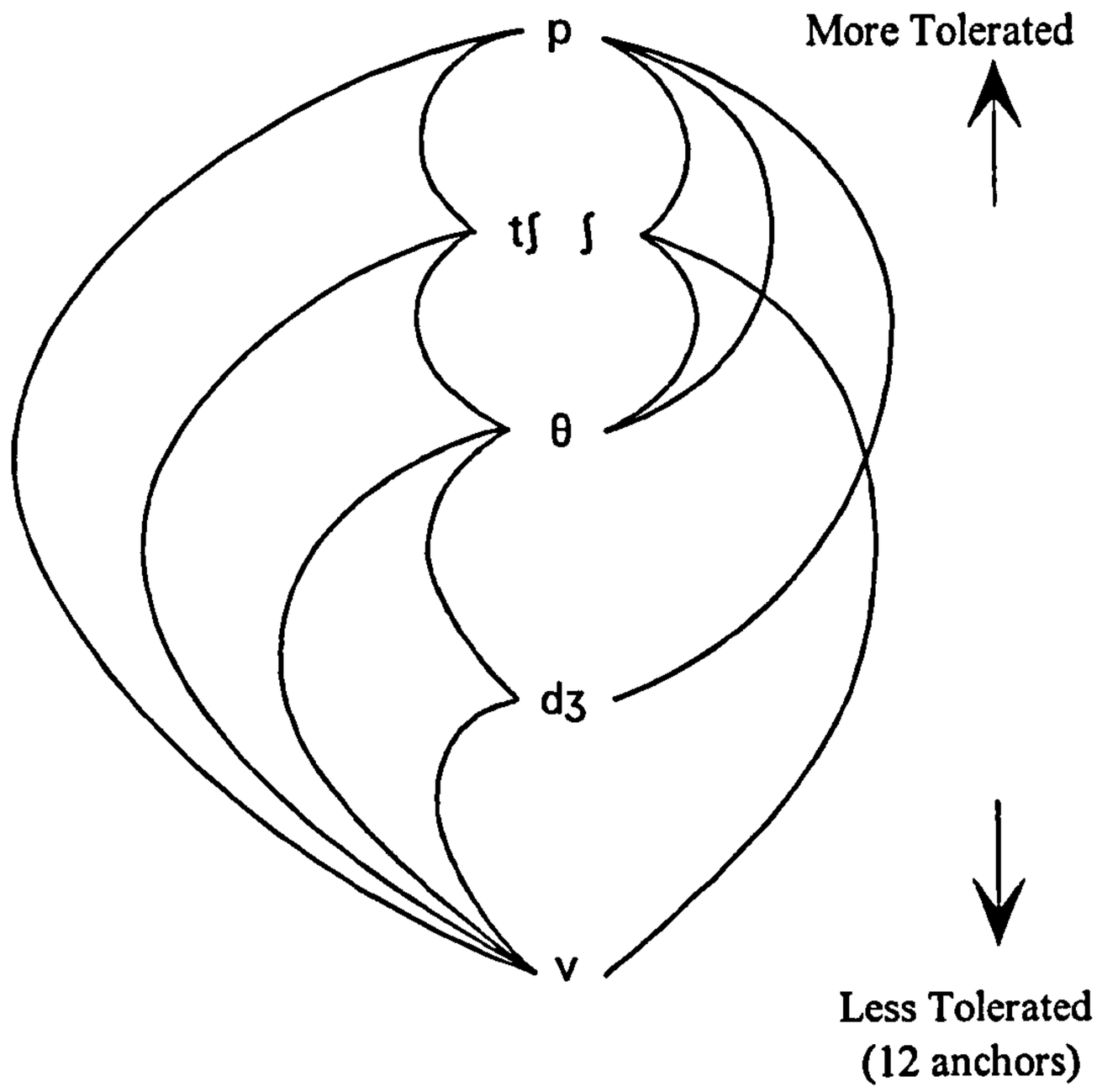
mispronounced / v / is the least tolerated misarticulation for the women's and men's voices, the two most tolerated were mispronounced / tʃ / and / p /.

The self-contradictory results noted in each table here (and also in Table 7.5, P. 135) are an anomaly I cannot explain: I have checked the recordings, data and analysis, and can find nothing at fault.

> (N.B. Strictly speaking the Chi Square figures—the significance values—only allow one to say whether there is a significant relationship or not: they do not reveal the *strength* of the relationship. My interpretation of the relationships as *more* and *less* tolerated is based on my extrapolated hierarchy of anchors, not on the Chi Square values.)

Table 7.4 English-speaking listeners listening to male voices

Contrast	%	Tape
/ tʃ - v /	98.1 ***	T2
	98.1 ***	T2
/ dʒ - v /	98.1 ***	T2
/ p - θ /	96.2 ***	T1
	72.2 ***	T2
/ ʃ - v /	96.2 ***	T1
	94.2 ***	T1
/ p - v /	94.2 ***	T1
/ p - tʃ /	92.6 ***	T2
/ θ - dʒ /	92.3 ***	T1
	72.2 **	T2
/ θ - v /	88.9 ***	T2
/ p - ʃ /	86.5 ***	T1
	86.5 ***	T1
/ ʃ - θ /	84.6 ***	T1
/ tʃ - θ /	78.8 ***	T1
	70.4 **	T2
/ p - dʒ /	64.8 *	T2

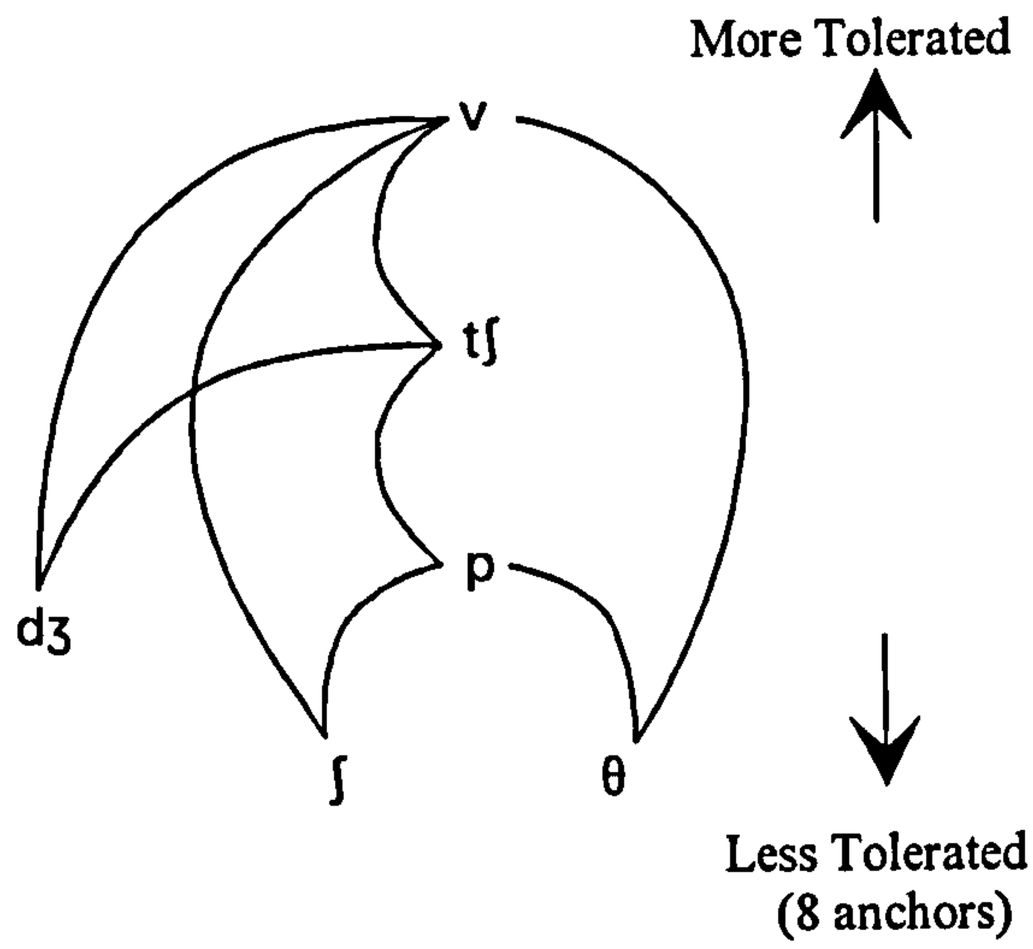


Notes: For T1, $N = 52$; for T2, $N = 54$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: / tʃ - dʒ / 96.3 *** T2
/ dʒ - tʃ / 87.0 *** T2

The tolerance hierarchy structure for the Finnish listeners (Tape 1 $N = 99$; Tape 2 $N = 97$) is not as simple as for the English-speaking listeners, there being some tied ('equal tolerance') positions and neither the Binomial test nor Chi Square analysis revealing any significant preference one way or the other for several contrasting pairs. See Table 7.5 and Table 7.6, P. 135. There are only eight significant relationships, or 'anchors', for the female voices (Table 7.5). However, the most striking feature here is the acceptance of mispronounced / v /: whereas English-speaking listeners overwhelmingly rejected mispronounced / v / in the speech of women, Finnish-speaking listeners tolerated it very easily, preferring it to all the other mispronounced phonemes except mispronounced / p / (the / v - p / contrast showed no significant preference either way: 'equal tolerance'). Where the men's voices are concerned (Table 7.6), mispronounced / p / and / v / are again tied equally. However, the position of / dʒ / relative to / p / and / θ / is puzzling: since mispronounced / p / is *more* tolerated than mispronounced / θ / ($p < .001$), how can mispronounced / dʒ / be *more* tolerated than mispronounced / p / and *less* tolerated than mispronounced / θ /? I have no explanation for this. I have checked the recordings, and checked the test tape, and I can find nothing to suggest the contrasting pairs of sentences are faulty or anything other than intended.

Table 7.5 Finnish-speaking listeners listening to female voices

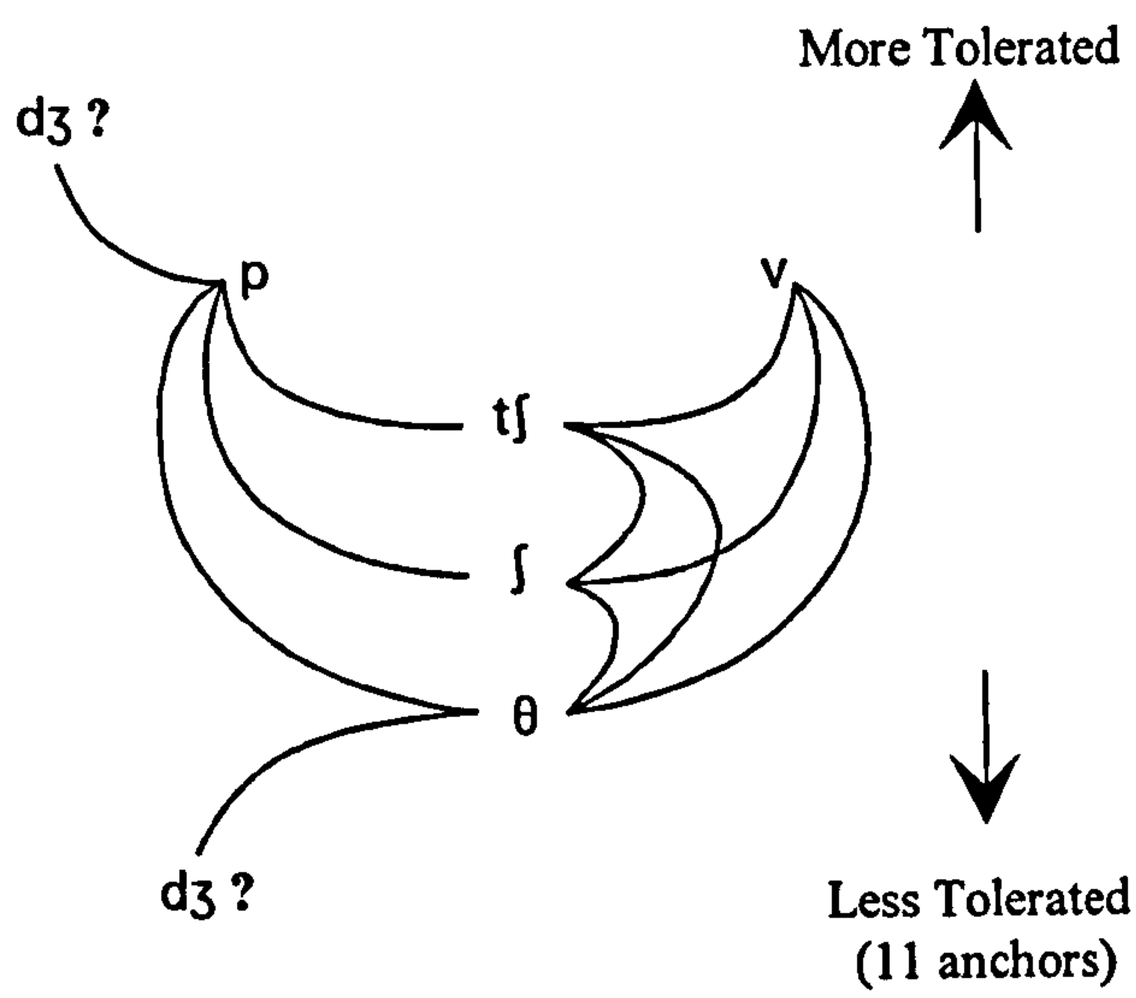
Contrast	%	Tape
/ tʃ - dʒ /	73.7 ***	T1
	93.9 ***	T1
/ v - θ /	88.9 ***	T1
/ p - θ /	88.9 ***	T1
	86.6 ***	T2
/ p - ʃ /	83.5 ***	T2
	75.3 ***	T2
/ v - tʃ /	82.8 ***	T1
	66.7 ***	T1
/ tʃ - p /	73.7 ***	T1
/ v - ʃ /	73.2 ***	T2
/ v - dʒ /	61.6 *	T1



Notes: For T1, $N = 99$; for T2, $N = 97$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: / θ - dʒ / 64.9 ** T2
/ dʒ - θ / 61.6 * T1

Table 7.6 Finnish-speaking listeners listening to male voices

Contrast	%	Tape
/ p - tʃ /	83.5 ***	T2
/ θ - dʒ /	80.8 ***	T1
	71.1 ***	T2
/ tʃ - θ /	76.8 ***	T1
	66.0 **	T2
/ p - θ /	75.8 ***	T1
	71.1 ***	T2
/ v - θ /	74.2 ***	T2
/ dʒ - p /	70.1 ***	T2
/ v - ʃ /	67.7 ***	T1
/ v - tʃ /	63.9 **	T2
/ p - ʃ /	62.6 *	T1
	62.6 *	T1
/ tʃ - ʃ /	61.6 *	T1
/ ʃ - θ /	60.6 *	T1



Notes: For T1, $N = 99$; for T2, $N = 97$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

The results for the Swedish-speaking (Finnish) listeners also differ from the tolerance hierarchy of the native English listeners, with mispronounced / v / and / p / again together at the more tolerated end of the spectrum: Table 7.7 and Table 7.8, P. 137. These two tables also show two distinct pairs of equally tolerated mispronunciations for both the women's voices and the men's voices, and this results in only four 'levels' for each table. Notice that both tables put mispronounced / ʃ / and / dʒ / together at the bottom of the toleration league.

The contrast in reactions to mispronounced / v / between the English-speaking informants on the one hand and the Finnish-speaking and Swedish-speaking (Finnish) informants on the other is underlined by the fact that for *all* the mispronunciation contrasts involving mispronounced / v / (including error-free pronunciation v. mispronounced / v /), for both female voices and male voices, there are significant correlations for mother-tongue of speaker (English v. Finnish, English v. Swedish; each mother-tongue informant group taken all together, without subdividing for men and women, or young and old). All the values for Cramér's *V* are significant at the $p < .001$ level; nearly all the values are over .40, several are over .90. See the GB/SF and GB/SW columns in Table 7.9, P. 138. All the high values for Cramér's *V* are of course a direct result of counteropposing preferences.

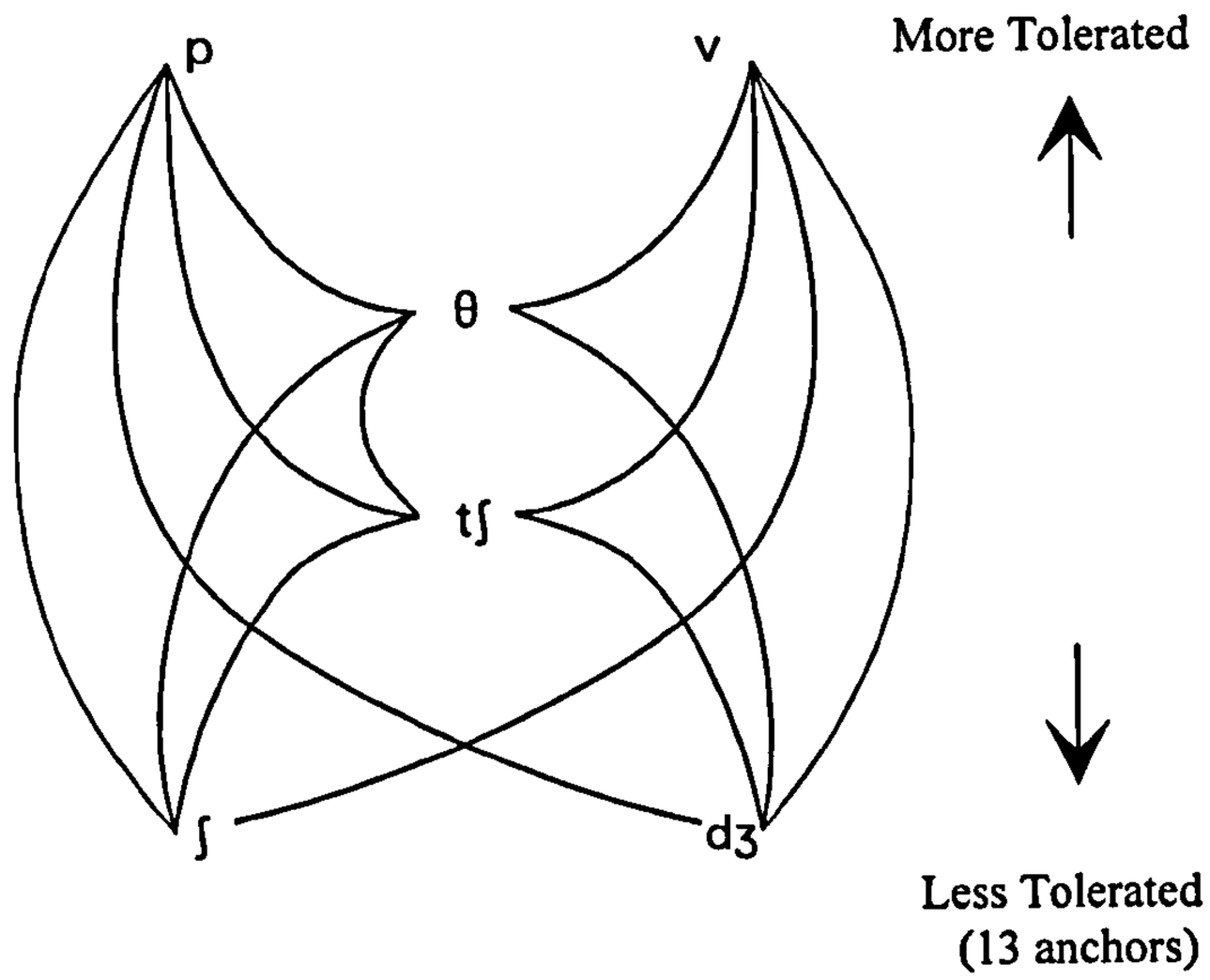
There are two clearly opposing camps here as regard mispronounced / v /: the GB listeners on the one side and the Finnish-speaking and Swedish-speaking listeners on the other. However, the Finnish-speaking and Swedish-speaking members of the latter camp are also distinguished one from the other, and by observing closely the values for Cramér's *V* in the GB/SF and GB/SW columns, one can predict more or less where those distinctions lie. Thus where in Table 7.9 the GB/SF value for Cramér's *V* is about the same as the GB/SW value (e.g. in the first row the / v - θ / values for the female voice are respectively .76 and .83) then the SF/SW value is obviously not going to be statistically significant (here the value for Cramér's *V* turns out to be .10). However, where the values are noticeably further apart (e.g. in the second row the / v - ʃ / values for the male voice are .37 and .91) then the SF/SW value is going to be high (here it is .52, $p < .001$).

Notice that the Finnish-speaking native listeners and the Swedish-speaking native listeners do not differ in their evaluation of the / v - θ / mispronunciation contrast: Cramér's $V = .10$ for the female voice, .09 for the male voice. Both listener groups are essentially the same in their attitude towards these mispronunciations.

The Finnish-speaking listeners and the Swedish-speaking listeners are also similar in their responses to the / v - p / contrast: the former make no significant distinction between the two mispronunciations for male or female voice (no anchor binds the mispronunciations), nor do the latter for the female voice though they do for the male voice. This closeness yields the expected low value for Cramér's *V* (.06) for the female voice, and a mildly significant value ($p < .05$) for the male voice (Cramér's $V = .21$). This result suggests that these two

Table 7.7 Swedish-speaking listeners listening to female voices

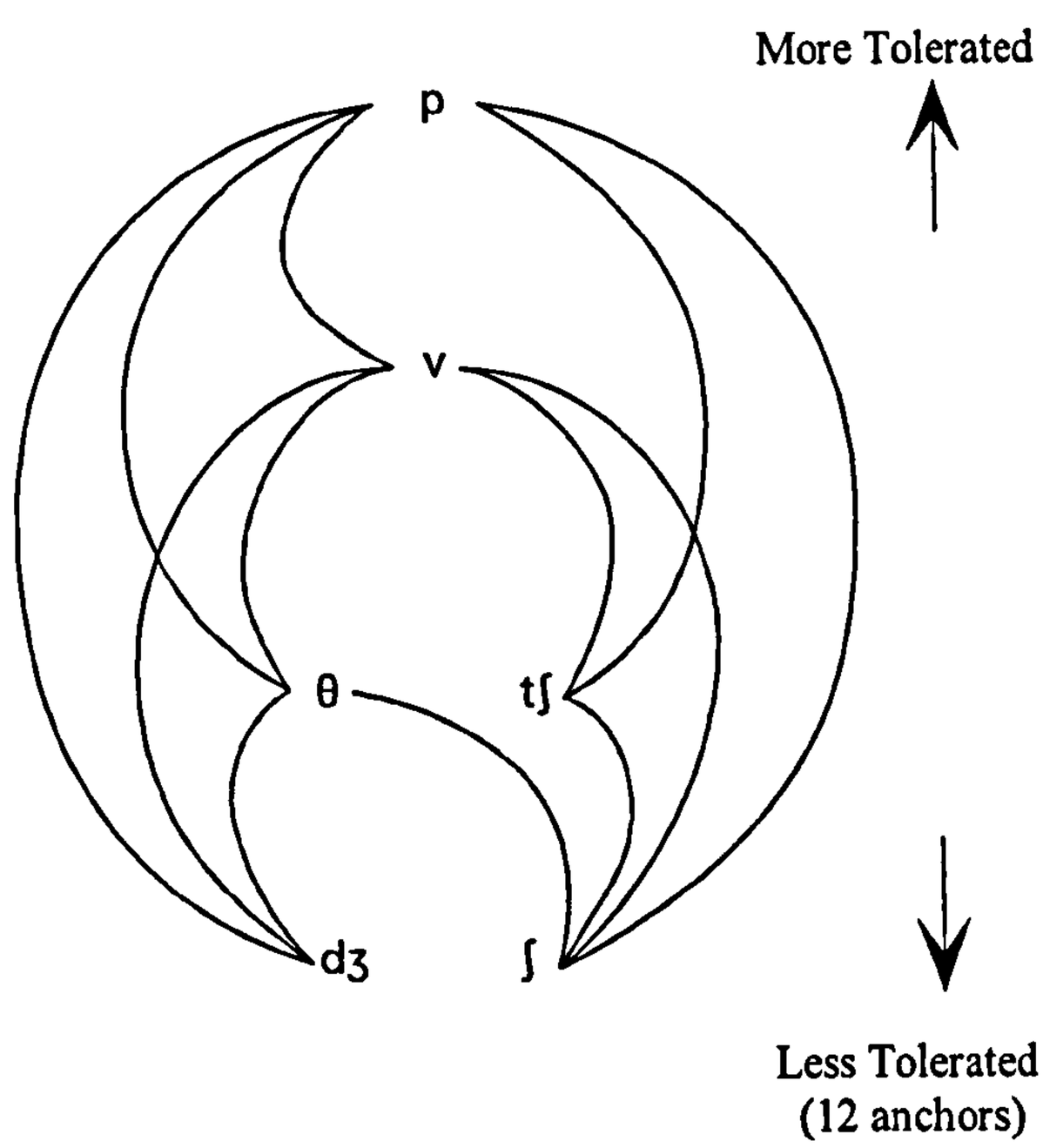
Contrast	%	Tape
/v - dʒ/	100.0 ***	T1
/p - ʃ/	97.5 ***	T2
	97.5 ***	T2
/v - ʃ/	97.5 ***	T2
	97.5 ***	T2
/θ - ʃ/	97.5 ***	T2
/v - tʃ/	97.6 ***	T1
	95.1 ***	T1
/p - dʒ/	95.1 ***	T1
/v - θ/	95.1 ***	T1
/θ - dʒ/	92.7 ***	T1
	92.5 ***	T2
/tʃ - dʒ/	82.9 ***	T1
	85.4 ***	T1
/p - θ/	85.0 ***	T2
	80.5 ***	T1
/p - tʃ/	82.9 ***	T1
/tʃ - ʃ/	82.5 ***	T2
/θ - tʃ/	80.5 ***	T1
	72.5 **	T2



Notes: For T1, $N = 41$; for T2, $N = 40$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.8 Swedish-speaking listeners listening to male voices

Contrast	%	Tape
/v - tʃ/	100.0 ***	T2
	100.0 ***	T2
/θ - dʒ/	100.0 ***	T2
	92.7 ***	T1
/v - ʃ/	97.6 ***	T1
	92.7 ***	T1
/p - ʃ/	97.6 ***	T1
	92.7 ***	T1
/p - tʃ/	95.0 ***	T2
	95.0 ***	T2
/v - dʒ/	87.5 ***	T2
/v - θ/	82.5 ***	T2
/θ - ʃ/	73.2 **	T1
/p - θ/	70.7 **	T1
/p - dʒ/	70.0 *	T2
/p - v/	68.3 *	T1
/tʃ - ʃ/	65.9 *	T1



Notes: For T1, $N = 41$; for T2, $N = 40$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: /tʃ - dʒ/ 97.5 *** T2
/dʒ - tʃ/ 77.5 *** T2

Table 7.9 Correlations (Cramér's *V*) of mispronunciation preferences involving mispronounced / v / to mother-tongue of informants

	Female voices				Male voices			
Contrast	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape
/ v - θ /	.76 ***	.83 ***	.10	T1	.61 ***	.72 ***	.09	T2
/ v - ʃ /	.51 ***	.94 ***	.41 ***	T2	.37 ***	.91 ***	.52 ***	T1
	.41 ***	.67 ***	.28 **	T2	.61 ***	.89 ***	.26 **	T1
/ v - p /	.50 ***	.64 ***	.06	T2	.48 ***	.34 ***	.21 *	T1
/ v - tʃ /	.74 ***	.91 ***	.20 *	T1	.60 ***	.98 ***	.38 ***	T2
	.56 ***	.87 ***	.30 ***	T1	.46 ***	.98 ***	.51 ***	T2
/ v - dʒ /	.58 ***	.98 ***	.39 ***	T1	.49 ***	.87 ***	.35 ***	T2

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB/SF T1: $N = 52/99$; GB/SW T1: $N = 52/41$; SF/SW T1: $N = 99/41$.
GB/SF T2: $N = 54/97$; GB/SW T2: $N = 54/40$; SF/SW T2: $N = 97/40$.

*** $p < .001$; ** $p < .01$; * $p < .05$.

informant groups are essentially similar in their attitude towards this / v - p / mispronunciation contrast, whatever the respective diagrammatic tolerance hierarchies I have proposed. In other words although mispronounced / v / appears two levels *above* mispronounced / p / in Table 7.5 (Finnish-speaking listeners, female voice) and *on the same* level in Table 7.7 (Swedish-speaking listeners, female voice), the responses are fundamentally the same: no anchor binds the sounds into the hierarchy.

For the / v - tʃ / contrast, the diametrically opposed preferences of the English-speaking and Swedish-speaking informants is reflected in the extremely high values for Cramér's *V* in the GB/SW columns of Table 7.9 (the antepenultimate and penultimate lines). The similar contrast between the English-speaking informants and the Finnish-speaking informants is not so striking, though still highly significant ($p < .001$); the relatively low value for Cramér's *V* (.46 ***) for one of the male voices results from the fact that the Finnish-speaking listeners displayed no preference one way or the other here between mispronounced / tʃ / and mispronounced / v /, but the English-speaking listeners were clear-cut in rejecting mispronounced / v /. A counter-opposite response is also seen in the responses to / v - dʒ /, the English-speaking listeners and the Swedish-speaking listeners again almost diametrically opposed (Cramér's *V* = .98/.87), but the English-speaking listeners and Finnish-speaking listeners only moderately opposed (Cramér's *V* = .58/.49).

The ten extremely high values for Cramér's *V* (over .80) in the two GB/SW columns (reflecting contradicting preferences) are confirmed (not surprisingly) by equally high significant values for Lambda. In other words, by knowing the mother tongue of the informant (English or Swedish), one can reduce the error in predicting the response by P% (P being the value of Lambda). The important parts of the two GB/SW columns are presented again in Table 7.10, with the values for Cramér's *V* and Lambda λ alongside each other.

Table 7.10 Correlations (Cramér's *V*, Lambda λ) for English v. Swedish mother-tongue of informants

	Female voices			Male voices		
Contrast	(*V*)	(λ)	Tape	(*V*)	(λ)	Tape
/ v - θ /	.83 ***	.82 ***	T1			
/ v - ʃ /	.94 ***	.93 ***	T2	.91 ***	.91 ***	T1
				.89 ***	.88 ***	T1
/ v - tʃ /	.91 ***	.91 ***	T1	.98 ***	.98 ***	T2
	.87 ***	.86 ***	T1	.98 ***	.98 ***	T2
/ v - dʒ /	.98 ***	.98 ***	T1	.87 ***	.83 ***	T2

Notes: For GB/SW T1: *N* = 52/41; for GB/SW T2: *N* = 54/40.
*** $p < .001$.

In an attempt to explain the discrepancy in tolerance for mispronounced / v / I should point out that Finnish 'v' is produced sometimes as a voiced labiodental

fricative [v] and sometimes as a voiced labiodental frictionless continuant [ʋ]. For example, in the Finnish word *sauva* (= 'stick') it is possible to use either. There is free variation. Thus I could propose that Finns carry this tolerance across to their evaluation of Finnish-accented English. Even mispronounced / v / when contrasted with error-free pronunciation (rejected almost unanimously by English native listeners—unsurprisingly) is tolerated to a remarkable level by certain groups of Finnish-speaking listeners: Table 7.11.

Table 7.11 Rejection (Yes/No) of mispronounced / v / in the / ∅ - v / contrast in the sentence:

/ ðə ˈvæn hæz ˈpɑːst ðə ˈvɪlɪdʒ ˈpʌb /
/ ðə ˈʋæn hæz ˈpɑːst ðə ˈʋɪlɪdʒ ˈpʌb /

Finnish-speaking listeners listening to:

(a) Female voices [a]

		ALL Yes *** (73%, *N* = 97)	
		MEN	**WOMEN**
		Yes ** (72%, *N* = 43)	Yes *** (74%, *N* = 54)
YOUNGER (<= 22)	Yes *** (83%, *N* = 54)	Yes ** (85%, *N* = 20)	Yes *** (82%, *N* = 34)
OLDER (>= 23)	No (61%, *N* = 43)	No (61%, *N* = 23)	No (60%, *N* = 20)

(b) Male voices [b]

		ALL Yes * (63%, *N* = 99)	
		MEN	**WOMEN**
		No (61%, *N* = 44)	Yes * (64%, *N* = 55)
YOUNGER (<= 22)	Yes ** (69%, *N* = 48)	No (68%, *N* = 19)	Yes * (69%, *N* = 29)
OLDER (>= 23)	No (57%, *N* = 51)	No (56%, *N* = 25)	No (58%, *N* = 26)

[a] Crosstabulation Younger v. Older: Cramér's $V = .25$ *
[b] Crosstabulations give no significant values for Cramér's V
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

When listening to a *woman*'s voice, younger Finnish-speaking listeners (everyone all together; and younger men alone; and younger women alone) rejected mispronounced / v / when contrasted with an error-free pronunciation, but older Finnish-speaking listeners (everyone; and older men alone; and older women alone) clearly tolerated the mispronunciation. Cross-tabulation on the younger v. older listeners showed a significant correlation for age: Cramér's

$V = .25$, $p < .05$. It is possible that this age discrepancy is an indication of a greater tolerance (or deference) towards women among older people than among younger people.

When a *man* was talking, only all younger people as a group, and all women as a group and younger women as a group (but not older women) rejected the mispronunciation of / v /; all older listeners, and men of any age, were tolerant. However, the values for Cramér's *V* revealed no significant correlations for age or gender, or age and gender together.

Swedish Swedish has no variation in its / v /: the sound is always a fricative, though the perceptual prominence of the friction noise may vary of course. Swedish-speaking Finns, however, are 'contaminated', dare I say, to a certain extent, by their familiarity with Finnish (and by their encounters with English-speaking Finns), and some people may tend towards a non-fricative articulation for / v / in their own Swedish. This would explain, I would surmise, why Swedish-speaking Finns are also prepared to accept [ʋ] for / v /. A group of Swedish Swedish informants would be most welcome here to discover whether *their* tolerance is markedly different to that of Finnish Swedish informants.

The fact that British listeners preferred mispronounced / tʃ, p, θ, dʒ, ʃ / to mispronounced / v / is perplexing. (Would an American informant group react in the same way?) For most English ears [ʋ] sounds like the labial-velar central gliding approximant / w /, and yet there are few minimal pairs in English (like *verse/worse*, *vine/wine*) involving / v / and / w /. The situation is much more critical for Finnish-accented / tʃ, p, θ, dʒ, ʃ / with lots of possibilities for confusion and misunderstanding: e.g. *pitch* sounds like 'pits'; *pat* like 'bat'; *thank* like 'tank'; *hedge* like 'heads'; *ship* like 'sip'. A minimum level of general intelligibility is easier to achieve with / v / than for the other phonemes here. So why this rejection of mispronounced / v /? Does the loss of friction noise for the / v / suggest a weak personality of some sort? A wimpish speaker? Possibly. (Could this help explain why Finnish women seem to dislike mispronounced English / v / in the speech of their male compatriots?) Does the frictionless [ʋ] imply some kind of a speech defect? Feasibly. Does the position of a mispronounced / v / in a word have a bearing on the reactions of listeners? Conceivably. (All my examples were word initial in a stressed syllable: *van*, *village*, *vicar*, *vulgar*. None were intervocalic.) Is the [ʋ] sound itself, rather than its use as a substitute for / v /, the source of irritation, for British listeners are known to have mocked the marked U-RP substitution of [ʋ] for / r / in the accent affected by Roy Jenkins in the '70s and '80s ('Incwease gwowth and weduce VAT': see Honey 1989)? Doubtful. Do other mispronunciations of / v / (e.g. a voiceless labiodental articulation [f], or a bilabial articulation [β]) in other foreign accents of English earn a similar low toleration mark? Perhaps (though Dretzke, 1987:515, found that / v / replaced by this same [ʋ] in German-accented English is considered as only a minor mistake by English listeners). And do other phonetic realisations of / v / in other varieties of native English (such as the same [ʋ] articulation found in the speech of some Indians) prompt the same reactions in British listeners? Perhaps not if the reaction is influenced by the company that

phonemes keep, by their phonetic realisations and by the total inventory of phonemes used.

Viewed from a historical perspective, the attitude of the English-speaking informants towards the mispronunciation of / v / is extremely curious. There was interchange of / v / and / w / in some dialects of English before the late 19th century (particularly in Cockney[2]), the more common substitution being / w / for / v / (e.g. *wicious*, *wittles*, *werry*, *wex*). Several novelists exploited this accent feature for comic effect in their depiction of lower social class characters. For example, Charles Dickens, well-known for his orthographic licence and inventiveness for character portrayal, occasionally focuses his readers' attention on the social origins of a character by respelling a word that one expects to have the letter 'v' (= phoneme / v /) with the letter 'w':

> 'Why, this here candle, now!' said Mr Peggotty, gleefully holding out his hand towards it. 'I know *wery* well that arter she's married and gone, I shall put that candle theer, just that same as now. I know *wery* well that when I'm here o' nights. . .' (Charles Dickens, 1850, *David Copperfield*, quoted by Mugglestone 1995:242. My emphasis.)

> Accordin' as the world went round, which round it did *rewolve* undoubtedly, even the best of gentlemen must take his turn of standing with his ed upside down and all his air a flying the wrong way. (Charles Dickens, 1857, *Little Dorrit*, quoted by Mugglestone 1995:221. My emphasis.)

One should be careful, however, Matthews warns (1938:160), not to take Dickensian orthography as a true representation of 19th century Cockney pronunciation: Dickens never gives an exact transliteration of an accent. He presents merely enough for the reader to identify the accent intended, i.e. he uses a literary convention. And the Cockney substitution of / w / for / v / was already a thing of the past by the mid 19th century.

Lewis Carroll also transforms the letter 'v' (phoneme / v /) into 'w' (/ w /). In *Through the Looking Glass* (published 1871), Alice knocks at the arched door with a bell handle on either side, and the Frog enquires:

> 'What did it ask you?'
> 'Nothing,' Alice said impatiently. 'I've been knocking at it!'
> 'Shouldn't do that—shouldn't do that—' the Frog muttered. '*Wexes* it, you know.'
> (See Gardner's edition 1970:328–9. My emphasis.)

There is, in other words, a historical precedent for the / v / to / w / substitution, and of course a real stigma was once attached to it, though how many people today are consciously aware of these facts is another matter. Is there perhaps an atavistic memory that remains? If there is, then it is perhaps this that underlies the present firm rejection of mispronounced non-fricative / v / (i.e. [υ]) in Finnish-accented English.

More interesting perhaps in the present Finnish context is Waller's (1987:22) quotation from the writings of a Samuel Bamford. Talking of a prison encounter

---

[2] See Matthews (1938:180–1), who claims that the replacement of / v / with / w / was characteristic of Elizabethan Cockney, though Cercignani (1981:328) states that Shakespeare's works provide little evidence for such an assertion.

in 1817, Bamford is confounded by the v's and w's (and rhyming slang) of two Londoners; they in turn

> stood wide agape when I opened upon them in broad Lancashire; the old man put on his spectacles, and peered at me as if I were uttering barbarous Tuss, or a Lappish incantation. (Bamford S. 1844. *Passages in the Life of a Radical.* Harvester Press edition 1967, Pp. 110–11, quoted by Waller *loc. cit.*)

Obviously 'Lappish' here could refer to Swedish or Norwegian or Finnish Lapland. And possibly 'Lappish' is being used imprecisely to conjure up vague images of the far European (Scandinavian) north. However, is it possible that, even in the 19th century, the confusion of / v / and / w / was recognised (at least by some knowledgeable people) as a Scandinavian characteristic?

7.2.2. Absence of plosive aspiration

Having examined the hierarchies of tolerance from the point of view of mispronounced / v /, it is now possible to examine other details more easily. Let me start with / p / since Cruttenden claims that despite the relatively low frequency of / p / in 'general texts' (1.78%), aspiration of / p, t, k / is an essential both of 'high acceptability in RP' and of 'minimum general intelligibility' (1994:196, 278, 285). Moreover, the phoneme / p /, whether aspirated or not, is an extremely common phoneme in the inventory of total words in the English language: according to Higgins (1996) the frequency of occurrence of my six consonant phonemes in the words listed in the *Oxford Advanced Learners' Dictionary of Current English* (*OALDCE*, 1974, electronic edition with Roger Mitton's additions) is as shown in Table 7.12.

Table 7.12 Frequency of particular consonants in the *OALDCE*

Phoneme	Appearance in X% of words in *OALDCE*	Consonant frequency rank
/ p /	20.62%	9
/ ʃ /	8.55%	15
/ v /	8.29%	16
/ dʒ /	5.38%	18
/ tʃ /	3.74%	21
/ θ /	2.25%	22

If we re-examine the responses of the English-speaking listeners to the *male* voices (Table 7.4, P. 134), we see that unaspirated / p / is firmly anchored (having the maximum 5 anchors) as the most acceptable mispronunciation. As for reactions to *female* voices (Table 7.3, P. 133), the native listeners of English place mispronounced / tʃ / as most acceptable, just pipping mispronounced / p /: both mispronounced / tʃ / and / p /, however, are similar in having 4 anchors placing them preferentially 'above' the other mispronunciations. These results would seem to run completely counter to Cruttenden's opinion about the importance of getting the aspiration of / p / right. The wealth of minimal pairs like *pat/bat*, *Peter/beater*, etc. that certainly exist in English may have led Gimson, Cruttenden and possibly others to believe that the absence of aspiration for / p /

is of the utmost importance: without the expected aspiration confusion would reign.[3] This may of course be true for a longer 'text,' for an extended conversation, though I remain highly dubious, believing that listeners are easily able to tune into an accent, adjust their sub-phonemic expectations and disambiguate meaning. Be that as it may, in the present experiment, with short simple sentences, and with attention directed away from meaning, unaspirated / p / appears to worry English native listeners hardly at all. Theoretical and potential confusions of / p / and / b / weigh lightly in comparison to the real mispronunciations of / θ, v, ʃ, tʃ, dʒ /.

The results of the Finnish-speaking listeners for unaspirated / p / are confusing. Where evaluation of the *male* voices is concerned, mispronounced / p / would seem to be equally tolerated at the top of the hierarchy together with mispronounced / v / (Table 7.6, P. 135), but the position of mispronounced / p / relative to mispronounced / dʒ / is still bewildering and apparently illogical. As for evaluation of the *female* voices (Table 7.5, P. 135) it would seem that the position of unaspirated / p / in the hierarchy is *below* that of / v / and / tʃ / even though no anchor binds it directly to / v /. Mispronounced / p / is indeed relatively insecure here, there being only three anchors holding it to its place in the hierarchy (admittedly / dʒ, ʃ, θ / all have only *two* anchors). Overall the picture of Finnish-speaking listeners' preferences for mispronunciation in female voices would seem to indicate either relative indifference to the seriousness of mispronunciations, or uncertainty as to how to rank the mispronunciations.

In Finnish the frequently occurring / p / is invariably unaspirated—as in *pankki* (bank), *posti* (post-office), *pussi* (bag), *piispa* (bishop)—and the voiced bilabial plosive / b /, word-initial and prevocalic, is relatively rare, e.g. *bussi* (bus), *baari* (bar), *booli* (punch = drink), *bileet* (party), *betoni* (concrete). Minimal pairs are few and far between: *pussi/bussi*; *baarit/paarit*. Hence, while a non-Finn's (e.g. an English person's) aspiration of / p / in Finnish sounds really odd (i.e. hilariously foreign-accented), the likelihood of confusion or misunderstanding in Finnish as a result of unnecessary aspiration is all but non-existent. This adequately explains, I would contend, the tolerance of the Finnish-speaking informants towards unaspirated / p / in English: the absence of aspiration in Finnish, together with its unimportance, is transferred to evaluations of unaspirated / p / in Finnish-accented English.

The data concerning unaspirated / p / provided by the Swedish-speaking informants is clear-cut. For the *male* voices (Table 7.8, P. 137) the mispronounced / p / is the most acceptable mispronunciation (as it was for the English-speaking informants), with the maximum 5 anchors placing it preferentially at the top of the hierarchy. For the *female* voices (Table 7.7, P. 137) the situation is almost identical, though no preference is shown between mispronounced / p / and mispronounced / v /: both are equally tolerated as most acceptable (similar to the

[3] According to John Higgins's analysis of occurrence of minimal pairs for all combinations of consonants and vowels (Higgins 1996), the / p - b / contrast in fact furnishes only a medium number of minimal pairs (424). Compare with the / t - s / contrast with 1258 minimal pairs.

situation revealed by the Finnish-speaking listeners for male voices, Table 7.6, P. 135).

When crosstabulations were performed on the mother-tongue of the informants in order to obtain values for Cramér's *V*, the three mother-tongue groups clearly distinguished themselves. See Table 7.13, P. 146. (Details of the / p - v / contrast are included for completeness but have already been discussed.)

First for the two sentences illustrating the / p - θ / contrast, all three mother-tongue language groups place mispronounced / p / above mispronounced / θ /. However, while the responses are almost identical for one sentence, as revealed by the low values for Cramér's *V*, the English-speaking listeners are distinct from both the Finnish-speaking listeners and the Swedish-speaking listeners (who in turn are almost identical in their responses) for the other sentence (line 3).

Second the Finnish-speaking informants set themselves apart for the / p - ʃ / contrast (both male and female voices, lines 4 and 5) and for the / p - dʒ / contrast (both male and female voices, last line): the values for Cramér's *V* are low in the GB/SW columns (i.e. the English-speaking listeners and Swedish-speaking listeners are similar in their evaluations, preferring both / p / to / ʃ / and / p / to / dʒ /), but the values are correspondingly higher in the GB/SF and SF/SW columns. Thus the attitudes here of the English-speaking and Swedish-speaking informants are very similar, and the Finnish-speaking listeners are distinct from both of them.

Third the English-speaking and Finnish-speaking listeners share identical preferences for / tʃ / over / p / for a *female* voice but for / p / over / tʃ / for a *male* voice, and this is confirmed by the low values for Cramér's *V* in the respective GB/SF columns of Table 7.13. The Swedish-speaking listeners' distinctive preference for / p / over / tʃ / for a *female* voice is clearly confirmed in the GB/SW and SF/SW columns with the modestly high values for Cramér's *V*; by contrast their preference for / p / over / tʃ / for a male voice mirrors the preference of both the English-speaking and Finnish-speaking listeners, and hence the low non-significant values of Cramér's *V* for one sentence here (.05/.15). The highly significant values for the other sentence (.47/.49) is caused by the fact that neither the English-speaking listeners nor the Finnish-speaking listeners have a significant preference one way or the other for the sentence.

One irregularity remains annoyingly unexplainable: / p - θ /, rows 2 and 3. The results along row 2 are broadly uniform, though the first column (GB/SF) does have a value for Cramér's *V* (.20) which is significant at the $p < .05$ level; the responses of the mother-tongue groups are broadly identical. The frustration felt here stems from the fact that the data from the second sentence used to test the / p - θ / contrast (row 3) apparently contradicts the first finding. Various explanations for this situation are possible: e.g. the recordings are somehow deficient; or the position of the mispronounced phoneme in the syllable and word structure is a confounding issue. In any event more investigations are necessary in order to clarify the matter.

Table 7.13 Correlations (Cramér's *V*) of mispronunciation preferences involving mispronounced / p / to mother-tongue of informants

	Female voices				Male voices			
Contrast	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape
/ p - v /	.50 ***	.64 ***	.06	T2	.48 ***	.34 ***	.21 *	T1
/ p - θ /	.20 *	.09	.11	T1	.01	.08	.06	T2
	.38 ***	.35 ***	.02	T2	.26 **	.35 ***	.05	T1
/ p - ʃ /	.22 **	.02	.19 *	T2	.25 **	.10	.30 ***	T1
	.30 ***	.02	.26 **	T2	.25 **	.20	.36 ***	T1
/ p - tʃ /	.03	.29 **	.24 **	T1	.09	.47 ***	.49 ***	T2
	.02	.39 ***	.34 ***	T1	.13	.05	.15	T2
/ p - dʒ /	.48 ***	.08	.42 ***	T1	.34 ***	.05	.37 ***	T2

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB/SF T1: $N = 52/99$; GB/SW T1: $N = 52/41$; SF/SW T1: $N = 99/41$.
GB/SF T2: $N = 54/97$; GB/SW T2: $N = 54/40$; SF/SW T2: $N = 97/40$.

*** $p < .001$; ** $p < .01$; * $p < .05$.

Another noteworthy feature is the fact that none of the various contrasts involving mispronounced / p / receives the same evaluation across the board: mispronounced / p / is neither equally preferred nor equally rejected by *everyone*, by all three mother-tongue groups of informants. This lack of congruity of response for mispronounced / p / would seem to suggest that evaluation is tempered more by mother-tongue of informant than by some aesthetic or other quality of the sound *per se*. Such a claim, even though it must remain highly tentative because of the lack of strong evidence here, probably seems intuitively unsurprising. On the other hand the implications for learners, teachers and classroom practice are enormous.

The present experiment included only one of the three fortis plosives of English, the alveolar and velar plosives being excluded. Further tests are needed to discover whether the patterns of acceptance and rejection found for non-aspirated / p / are replicated with non-aspirated / t / and / k /: is the tolerance for non-aspiration similar for all three RP fortis plosives?

The problems some non-native speakers of English have with the / p - b / contrast in English is well-known and has been mocked and mimicked for centuries. For example, Shakespeare, in *Henry V*, exploiting pronunciation (in addition to grammar and vocabulary) to indicate the origin of three non-English army captains (Fluellen, Welsh; MacMorris, Irish; and Jamy, Scots), characterised Fluellen's Welsh speech primarily with 'b' becoming 'p' (my *emphasis* in all the following Shakespeare quotations):

> To the mines! . . . I think 'a will *plow* up all. (III.ii:54)
>
> Kill the *poys* and the luggage. (IV.vii:1)
>
> Ay, he was *porn* at Monmouth, Captain Gower. What call you the town's name where Alexander the *Pig* was born? (IV.vii:11)
>
> . . . it is out of my *prains* what is the name of the other river. (IV.vii:27)
>
> . . . and your great-uncle Edward the *Plack* Prince of Wales . . . fought a most *prave pattle* here in France. (IV.vii:90–1)
>
> All the water in Wye cannot wash your Majesty's Welsh *plood* out of your *pody* . . . God *pless* it and preserve it . . . (IV.vii:103–5)

Shakespeare repeats the same 'Welsh-accent' feature with the Welsh parson, Sir Hugh Evans, in *The Merry Wives of Windsor*:

> It is *petter* that friends is the sword, and end it; and there is also another device in my *prain*, which peradventure *prings* goot discretions with it. (I.i:39–40)
>
> I will *peat* the door for Master Page. [Knocks] What hoa? Got *pless* your house here! (I.i:64–5)

Presumably this Welsh-accented 'b' is devoiced, or rather produced with a delayed voice onset time (VOT), though whether on the one hand it be produced fortis together with concomitant aspiration or with following approximant devoicing, or whether on the other hand it be produced 'weak' with no aspiration and with no following approximant devoicing, is a matter for debate.

MacMorris's Irish accent is indicated mainly by the lenis alveolar sibilant being articulated as a fortis palatoalveolar sibilant (e.g. 'The work *ish* ill done; it *ish* give over.' III.ii:84). Today this retracted 's' is perhaps more commonly

associated with drunkenness, or, as a result of Sean Connery's (alias 007's) way of articulating his 's' sounds, with a Scottish accent, a feature mimicked by Irvine Welsh in *Trainspotting* (e.g. 'the unsheen ashashin').

By contrast, Jamy's Scottish accent is conveyed chiefly through the vowels and diphthongs: e.g. in Act III scene ii: *gud* 78, *bath* 96, *grund* 109, *wad* 113, *tway* 114. Like Fluellen he also says *sall* for *shall.*

Since Shakespeare's intentions here were presumably to support the main themes in his play, the Welsh accent with its 'funny' plosives is *positive*, for Henry V, the ideal warrior king, is proud of his Welsh connections. The Scottish accent, on the other hand, conveyed in Jamy's speech through the altered vowels and diphthongs, is presumably *negative*, for the 'weasel' Scots are slighted, denigrated and mistrusted.

Is it just possible—despite all the years and the social/cultural/national upheavals intervening between Shakespeare and today, despite the differences between reactions to regional/national accents of English and reactions to foreign-accented English, despite the aesthetic can of worms my question is about to open—is it possible that mispronounced sibilants, vowels and diphthongs are widely perceived as being major distortions, whereas mispronounced plosive sounds are generally perceived as being intrinsically of overall minor importance, are indeed widely recognised as being more acceptable articulatory deviations than others, are even somehow inherently pleasing? How else can one interpret the violent reaction against mispronounced / v / in the present study and the mild reaction against mispronounced / p /? Gimson's and Cruttenden's apparent reliance on phoneme counts and minimal pair counts to decide what is or is not of higher and lesser importance would seem to receive little support from the present results.

7.2.3. Mispronounced affricates

The position of mispronounced / tʃ / in the preference hierarchies would seem to vary according to the informant group. The English-speaking listeners (Table 7.3, P. 133 and Table 7.4, P.134) place it at the top end of the toleration scale: when *female* voices are being evaluated, mispronounced / tʃ / *is* the most tolerated mispronunciation (top level of 6 levels; 4 anchors), and when *male* voices are being evaluated, it is in joint second position (level 2 of 5 levels; 3 anchors). The Finnish-speaking listeners place the mispronounced / tʃ / on level 2 of 4 levels for both female and male voices (3 anchors and 4 anchors respectively; Table 7.5 and Table 7.6, P. 135), and Swedish-speaking listeners place it on level 3 of 4 levels (5 anchors and 3 anchors respectively; Table 7.7 and Table 7.8, P. 137).

The situation for mispronounced / dʒ / also seems to be different for all three informant groups. The English-speaking informants place mispronounced / dʒ / more or less in the middle of the hierarchy, but on the less tolerated side of the middle (Table 7.3 and Table 7.4).The Finnish-speaking informants would seem to place the mispronounced / dʒ / in the middle of their hierarchy for female voices though there are only two anchors holding the sound in that position (Table 7.5), but for the male voices the picture is—I repeat—extremely confused

(Table 7.6): is mispronounced / dʒ / the most preferred or the least preferred mispronunciation?

Finally the Swedish-speaking informants clearly place the mispronounced / dʒ / together with mispronounced / ʃ / at the less tolerated end of the scale for both female and male voices (Table 7.7 and Table 7.8).

This disparate situation is reflected in the values for Cramér's *V* in the cross-tabulations performed on the mother-tongue of the informants (Table 7.14, P. 150. N.B. The contrasts of the affricates with mispronounced / v / and / p / are omitted as they have already been discussed.).

The reaction of the Finnish-speaking listeners and the English-speaking listeners towards the / tʃ - θ / contrast is the same for the *male* voice (low non-significant values for Cramér's *V*: .02 and .04), but significantly different for the female voice (Cramér's *V* = .27 ***/.31 ***). The Swedish-speaking listeners and the English-speaking listeners noticeably agree in their evaluation of one of the mispronounced / tʃ - θ / contrasts for female voices but disagree on the other (Cramér's *V* = .04/.53 ***): this is a direct result of the self-contradictory results found in the data for the English-speaking listeners alone (though I can offer no explanation for this anomaly).

With a *female* voice the / tʃ - ʃ / provokes no preference from the Finnish-speaking listeners whereas it does from both the English-speaking and Swedish-speaking listeners. However, this difference is only significant between the Finnish-speaking and Swedish-speaking listeners (Cramér's *V* = .25 ***).

With a *male* voice it is the English-speaking listeners who betray no preference one way or the other on the / tʃ - ʃ / contrast, and this leads to a mildly significant difference with the other two mother-tongue groups (Cramér's *V* = .20 */.25 *).

As for the / dʒ - ʃ / contrast it is noteworthy that only the English-speaking informants listening to a female voice show any preference between the two mispronunciations, / dʒ / preferred to / ʃ / (Table 7.3, P. 133): apart from this exception mispronounced / dʒ / and mispronounced / ʃ / are equally acceptable, equally tolerated. However, the non-significant preference *tendencies* do not always correspond. Hence, for the female voices, the significant values for Cramér's *V* in the GB/SF and GB/SW columns (.28 *** and .33 ** respectively) arise from the fact that the preferences of the Finnish-speaking and Swedish-speaking listeners, although not reaching statistical significance, go in the *opposite* direction to those of the English-speaking listeners (mispronounced / ʃ / is preferred to mispronounced / dʒ /, 58.8% and 62.5% respectively, $p > .05$).

Similarly but inversely, for the male voices, the English-speaking listeners lean towards a preference for / ʃ / over / dʒ / (61.5%, $p > .05$) whereas Finnish-speaking and Swedish-speaking listeners lean towards / dʒ / over / ʃ / (59.6% and 61.0% respectively, $p > .05$). This discrepancy prompts the low significant values for Cramér's *V* (.20 * and .22 *) in the GB/SF and GB/SW columns.

For the female voices all three listener groups prefer mispronounced / tʃ / to mispronounced / dʒ /, but the English-speaking listeners are so different to the

Table 7.14 Correlations (Cramér's *V*) of mispronunciation preferences involving mispronounced / tʃ / and / dʒ / to mother-tongue of informants

	Female voices				Male voices			
Contrast	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape
/ tʃ - θ /	.27 ***	.04	.29 ***	T2	.02	.38 ***	.34 ***	T1
	.31 ***	.53 ***	.20 *	T1	.04	.30 **	.24 **	T2
/ tʃ - ʃ /	.14	.14	.25 **	T2	.20 *	.25 *	.04	T1
/ dʒ - θ /	.15	.01	.15	T1	.01	.38 ***	.33 ***	T2
	.06	.51 ***	.50 ***	T1	.26 **	.06	.28 ***	T2
/ dʒ - ʃ /	.28 ***	.33 **	.03	T2	.20 *	.22 *	.01	T1
/ tʃ - dʒ /	.37 ***	.20	.17	T1	.35 ***	.13	.23 ***	T2
	.27 ***	.19	.13	T1	.41 ***	.03	.39 ***	T2

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB/SF T1: $N = 52/99$; GB/SW T1: $N = 52/41$; SF/SW T1: $N = 99/41$.
GB/SF T2: $N = 54/97$; GB/SW T2: $N = 54/40$; SF/SW T2: $N = 97/40$.

*** $p < .001$; ** $p < .01$; * $p < .05$.

Finnish-speaking listeners in their percentage response (65.4% and 96.2% to 93.9% and 73.7% respectively) that this leads to a highly significant value for Cramér's V (.37/.27, $p < .001$): Table 7.14, P. 150, 1st column.

For the male voices the patterns of evaluation for the / tʃ - dʒ / contrast are not immediately apparent from the preference hierarchies presented in Tables 7.4, 7.6 and 7.8. The English-speaking and Swedish-speaking listeners have self-contradictory results, and hence their evaluations are excluded from the hierarchies, and the Finnish-speaking listeners have no statistically significant preferences one way or the other (Table 7.15).

Table 7.15 Preferences of informants listening to male voices

Contrast	Tape	GB %	SF %	SW %
/ tʃ - dʒ /	T2	96.3 ***	57.7	97.5 ***
	T2	87.0 (dʒ) ***	52.6 (dʒ)	77.5 ***(dʒ)

Notes: For GB T2 $N = 52$; for SF T2 $N = 97$; for SW T2 $N = 40$.
Chi-Square analysis: *** $p < .001$

The low percentages in the SF column explain the pattern of significant values for Cramér's V for the male voices in Table 7.14, P. 150.

In an attempt to understand how the curious situation of mispronounced / dʒ / in the hierarchy of the Finnish-speaking listeners arises, one needs to remember that apart from a very limited set of words using the voiceless denti-alveolar affricate [t̪s] (e.g. *tsaari, tsemppiä, tsekata*), affricates don't normally appear in Finnish. For Finns affricates constitute a 'strange' manner of articulation, and palatoalveolar a 'strange' place of articulation. In this light the Finnish listener's ambivalence towards the Finnish-accented mispronunciations is perhaps not so surprising: 'the speaker's accent sounds Finnish, yes, does not sound truly English, but what should the sound *really* be since it more or less fits the norms of the Finnish language?'

Notice that Cruttenden (1994:278–80, 285–6), in his sections on the essentials in RP for high acceptability and minimum general intelligibility, says nothing at all about the affricates. This is a serious omission, I feel, since I believe English native listeners are well aware that Finnish-accented mispronounced / tʃ / and / dʒ /, with their fronted post-alveolar (possibly alveolar) articulations, are a source of very real confusion: *watch/what's, beach/beats, witch/wits, age/aids, hedge/heads*, etc.

7.2.4. Mispronounced sibilants

The Swedish-speaking informants placed mispronounced / ʃ / at the bottom of their tolerance hierarchies for female and male voices, together with / dʒ /. The result is clear: mispronounced / ʃ / and / dʒ / are not liked (Table 7.7 and Table 7.8, P. 137).

The Finnish-speaking informants also seem to place / ʃ / at the lower end of the scale, the low position being clearer when a female voice is involved but less obvious when a male voice is involved (Table 7.5 and Table 7.6, P. 135). While the one voiceless sibilant sound in Finnish is theoretically amenable to considerable variation in place of articulation (there are no other fricatives 'near' it, speaking in articulatory terms), the amount of permissible and tolerated movement away from its postalveolar place of articulation is very small: school children are regularly scolded by teachers for any suspected heinous *s-vika* (= 's-mistake') and even packed off to speech therapists to eradicate it. On the other hand, regional accents of Finnish (particularly Helsinki Finnish) are often noted for their 's'. Speculation as to why there is this intolerance towards mispronounced English / ʃ /, therefore, remains for the moment purely that: speculation.

The English-speaking informants are the odd ones out here. They seem to react differently to mispronounced / ʃ / depending on whether the mispronunciation is produced by a woman or a man. When a woman's voice is being evaluated (Table 7.3, P. 133), mispronounced / ʃ / is placed just above mispronounced / v / at the *less* tolerated end of the scale, but when a man's voice is being evaluated, the mispronounced / ʃ / is placed in joint second place at the *more* tolerated end of the scale (Table 7.4, P. 134). The clear suspicion of a differentiation based on the gender of the voice here is supported by the values for Cramér's *V* given in crosstabulations focusing on gender of voice (see Table 7.16).

Table 7.16 Preferences of English-speaking informants for mispronounced / ʃ /, and correlations for gender of voice

Contrast	% preference when Female voice ($N = 54$)	% preference when Male voice ($N = 52$)	Cramér's *V* for gender of voice (V)
/ θ - ʃ /[a]	100.0 ***	84.6 *** (ʃ)	.86 ***
/ ʃ - v /	68.5 **	96.2 ***	.36 ***
	96.3 ***	94.2 ***	.05
/ dʒ - ʃ /	70.4 **	61.5 (ʃ)	.32 ***
/ tʃ - ʃ /	70.4 **	59.6 (ʃ)	.30 **
/ p - ʃ /	98.1 ***	86.5 ***	.22 *
	98.1 ***	86.5 ***	.22 *

[a] N.B. the first phoneme presented in the contrast column is the preferred mispronunciation: thus for the / θ - ʃ / contrast in the first row, mispronounced / θ / is preferred to mispronounced / ʃ /. However, for the male voices the preferred mispronunciation is unexpectedly reversed and this is indicated by the new preferred mispronunciation being shown in brackets: 84.6 *** (ʃ).

Chi-Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Unfortunately—and I want to stress this forcefully—this difference in evaluation is affected not only by the gender of the voice, but also by voice quality,

pitch, tempo and stress patterning.[4] The difference between the hierarchies for female voices and male voices discovered here, and especially the values for Cramér's *V* given in Table 7.16, are merely suggestive: they prompt one to musings of differentiation based on gender of voice, but one cannot proceed further, not with the present material. Indeed no experiment can satisfactorily isolate gender of voice as the sole influencing variable in listener evaluation: it's an impossibility. A speaker may talk quickly or slowly; their voice may be high-pitched or low-pitched, harsh or breathy, nasal or adenoidal; they may misarticulate dental fricatives or not; and so on; but their voice is *always* either masculine or feminine. Thus the inferences to be drawn here are necessarily highly speculative: what role the gender of the voice plays *on its own* is impossible to evaluate. Nevertheless, it is clear from Table 7.16 that three mispronunciation contrasts (/ θ - ʃ /, / ʃ - v / and / dʒ - ʃ /) yield highly significant values for Cramér's *V* based on gender of voice, one value (.86) indicating an almost diametrically opposed preference.

The differences in evaluations arising from mother-tongue of informant group can be seen in the values for Cramér's *V* presented in Table 7.17, P. 154. All the values are repeated even though all except those in line 1 have been discussed already.

For female voices the Finnish-speaking listeners are indifferent to the / θ - ʃ / contrast, but the other two groups prefer mispronounced / θ / to / ʃ /. This explains the pattern of values for Cramér's *V* in the respective GB/SF, GB/SW and SF/SW columns of Table 7.17.

For male voices the preferences of the Finnish-speaking and English-speaking listeners go in the same direction (mispronounced / ʃ / preferred to mispronounced / θ /), but are not of the same strength (60.6%, $p < .001$, and 84.6%, $p < .05$, respectively). By contrast the Swedish-speaking listeners prefer mispronounced / θ / (73.2%, $p < .01$). This pattern results in the remaining respective values for Cramér's *V* in Table 7.17.

When discussing the sibilant articulations, Cruttenden merely claims (1994: 279) that for high acceptability 'alveolar / s, z / must remain clearly distinct from . . . palatoalveolar / ʃ, ʒ /.' This exhortation is fine as far as it goes, but the tables and lists on which Cruttenden bases his advice are sterile, untouched by users and usage. If voice-gender differentiation *is* present in reactions to mispronounced sibilants (and the present data would seem to suggest that it is), then Cruttenden walks right past the issue, blithely oblivious of it. It is of course possible that the English-speaking listeners' dislike for mispronounced / ʃ / in male voices is influenced by a feeling that the mispronounced / ʃ / sounds lispy (possibly effeminate?) or else drunken. A qualitative follow-up study might elucidate this point.

---

[4] N.B. No attempts were made during the recordings for this experiment to control the three variables which might be amenable to control, viz. pitch contouring, tempo and stress patterning. However, efforts *were* made to control them during the recordings for Experiment 2 (Speaker evaluation).

Table 7.17 Correlations (Cramér's *V*) of mispronunciation preferences involving mispronounced / ʃ / to mother-tongue of informants

	Female voices				Male voices			
Contrast	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape
/ ʃ - θ /	.56 ***	.12	.49 ***	T2	.25 **	.58 ***	.31 ***	T1
/ ʃ - v /	.51 ***	.94 ***	.41 ***	T2	.37 ***	.91 ***	.52 ***	T1
	.41 ***	.67 ***	.28 **	T2	.61 ***	.89 ***	.26 **	T1
/ ʃ - dʒ /	.28 ***	.33 **	.03	T2	.20 *	.22 *	.01	T1
/ ʃ - tʃ /	.14	.14	.25 ***	T2	.20 *	.25 *	.04	T1
/ ʃ - p /	.22 **	.02	.19 *	T2	.25 **	.10	.30 ***	T1
	.30 ***	.02	.26 **	T2	.25 **	.20	.36 ***	T1

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB/SF T1: $N = 52/99$; GB/SW T1: $N = 52/41$; SF/SW T1: $N = 99/41$.
GB/SF T2: $N = 54/97$; GB/SW T2: $N = 54/40$; SF/SW T2: $N = 97/40$.

*** $p < .001$; ** $p < .01$; * $p < .05$.

7.2.5. Mispronounced fortis dental fricative / θ /

Whereas the English-speaking and Swedish-speaking informants place mispronounced fortis dental / θ / in (or towards) the centre of their tolerance hierarchies (Table 7.3, Table 7.4, Table 7.7, Table 7.8), the Finnish-speaking informants would seem to tolerate the mispronunciation least of all, or joint least of all (Table 7.5 and Table 7.6). Since there are no dental fricatives in Finnish, the sound is notoriously difficult for Finnish native speakers, and beginners invariably misarticulate the sound, typically producing a dental plosive. Presumably the sound is so well recognised as a shibboleth by Finns themselves that they quickly spot the mispronunciation and refuse to tolerate it. This intolerance is somewhat poignant and ironic when one recalls that Irish English (particularly southern Irish) regularly articulates RP / θ, ð / as [t̪ , d̪] (and sometimes there is also affrication: [t̪^{θ} , d̪^{ð}]) (Wells 1982:429). It is highly probable, I suspect, that a group of Irish informants would refuse to recognise the Finnish accented mispronunciation here as mispronunciation at all. More research in this area would also be most welcome.

Crosstabulations performed on mispronounced / θ / for mother-tongue of informant are presented in Table 7.18, P. 156, though all the details have been discussed already.

Cruttenden only mentions / θ / under the general heading of 'high acceptability' in order to say (somewhat lamely) that it must remain distinct from alveolar / s / (1994:279). On the other hand, he does claim (1994:286) that for minimum general intelligibility 'it is not acceptable to replace / θ, ð / by dental aspirated [t̪^{h}, d̪^{h}] (*sic*).'

7.3. Influence of gender of informants

The data was analysed to test whether the relationship between alternative mispronunciations and preference is moderated by the gender of the informants. Schematically the situation being tested may be presented as in Figure 7.1, the horizontal arrow (⟶) showing 'independent variable → dependent variable,' and the vertical arrow (↑) indicating the moderating variable.

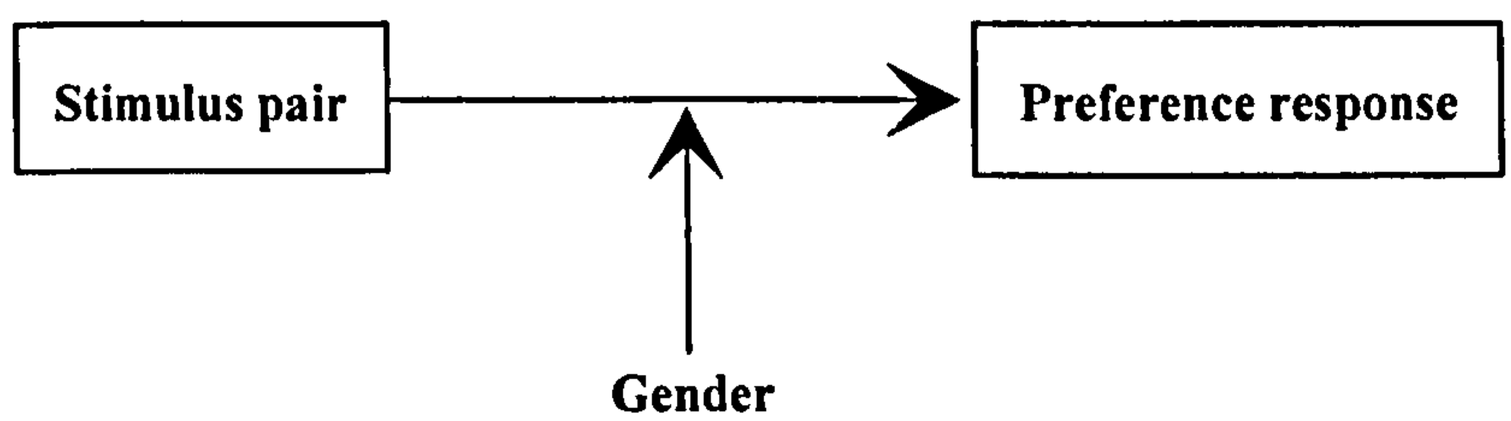


Figure 7.1 Influence of informant gender on response

The aim was to ascertain whether male and female informants respond differently to mispronunciations produced by male and female speakers, and whether they reveal own-gender solidarity (more support/tolerance) or own-gender betrayal (more rejection/intolerance). The analyses took the null hypothesis as their starting point, and used 2-tailed tests. The Swedish-speaking informants were not

Table 7.18 Correlations (Cramér's *V*) of mispronunciation preferences involving mispronounced / θ / to mother-tongue of informants

	Female voices				Male voices			
Contrast	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape	GB/SF (*V*)	GB/SW (*V*)	SF/SW (*V*)	Tape
/ θ - v /	.76 ***	.83 ***	.10	T1	.61 ***	.72 ***	.09	T2
/ θ - dʒ /	.15	.01	.15	T1	.01	.38 ***	.33 ***	T2
	.26 **	.06	.28 ***	T2	.06	.51 ***	.50 ***	T1
/ θ - tʃ /	.27 ***	.04	.29 ***	T2	.02	.38 ***	.34 ***	T1
	.31 ***	.53 ***	.20 *	T1	.04	.30 **	.24 **	T2
/ θ - ʃ /	.56 ***	.12	.49 ***	T2	.25 **	.58 ***	.31 ***	T1
/ θ - p /	.20 *	.09	.11	T1	.01	.08	.06	T2
	.38 ***	.35 ***	.02	T2	.26 **	.35 ***	.05	T1

Notes: Informant groups: GB = English-speaking informants; SF = Finnish-speaking informants; SW = Swedish-speaking informants.

Group sizes: GB/SF T1: $N = 52/99$; GB/SW T1: $N = 52/41$; SF/SW T1: $N = 99/41$.
GB/SF T2: $N = 54/97$; GB/SW T2: $N = 54/40$; SF/SW T2: $N = 97/40$.

*** $p < .001$; ** $p < .01$; * $p < .05$.

included in this part of the study since the sample size was too small for any confident analyses to be performed.

7.3.1. English-speaking informants

The hierarchy tables established from the Chi-Square analysis of the data for English-speaking men and English-speaking women listening to both the female voices and the male voices were broadly consistent—not surprisingly—with those for all the English-speaking listeners taken as one group all together.

With the English-speaking men listening to the *female* voices (Table 7.19, P. 158), the top three places seen in the general hierarchy with everyone evaluating the mispronunciations is levelled so that mispronounced / θ, tʃ, p / are in joint first place. Men on their own have a truncated hierarchy, with only four 'vertical' levels rather than six on the scale for everyone. Notice, however, that mispronounced / ʃ /, placed just 'above' mispronounced / v /, is not well tolerated.

With the English-speaking men listening to *male* voices, the mispronounced / tʃ / is displaced in the hierarchy and seems more unstable and poorly anchored than when everyone is evaluating the mispronunciations (Table 7.20, P. 158). Here there are five levels, just as there are in the hierarchy established with everyone all together. Notice that here the mispronounced / ʃ / is relatively well tolerated.

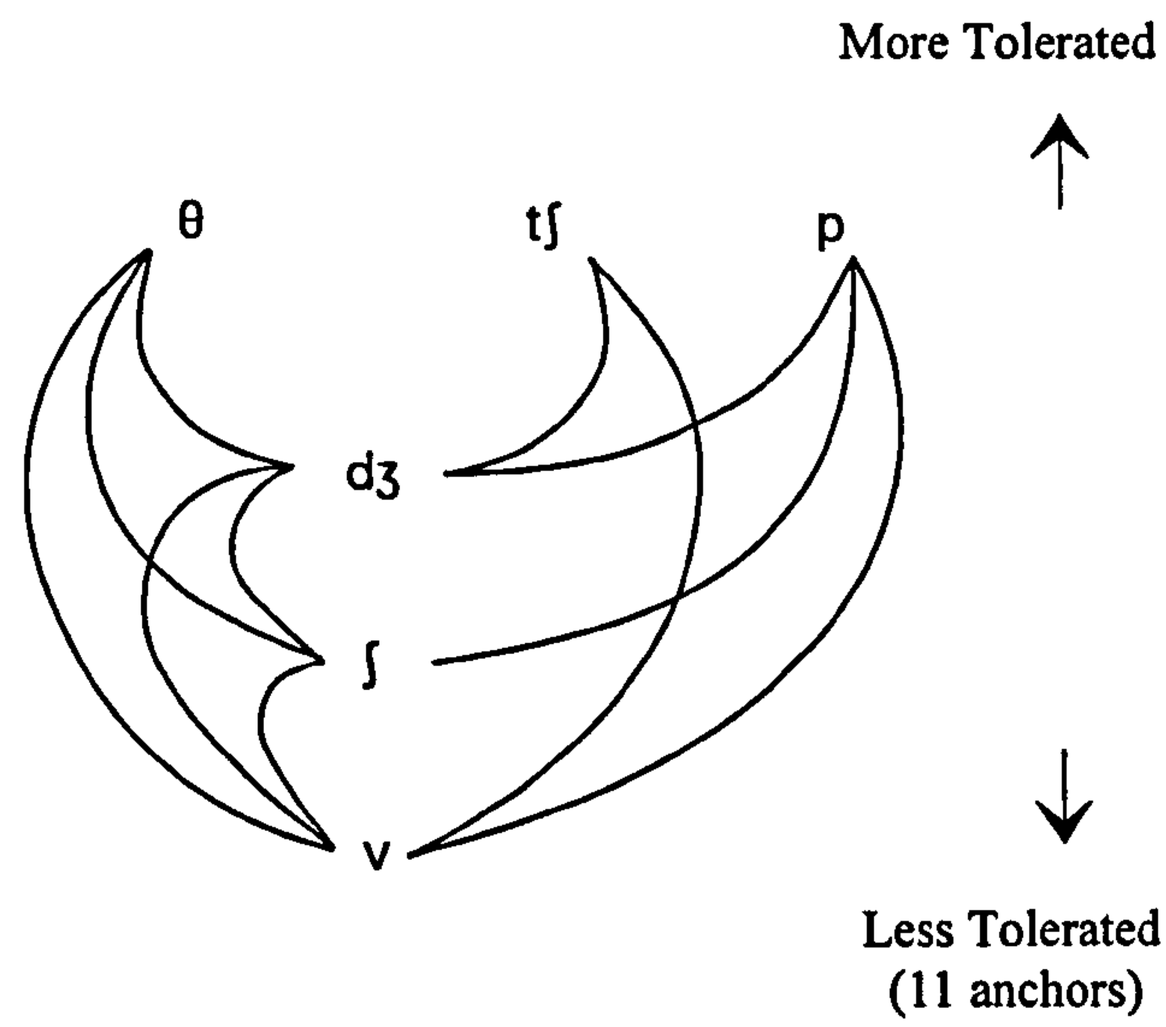
The hierarchy for the English-speaking women listening to *female* voices is firmly anchored (14 anchors) and is practically the same as that for everyone all together, except that mispronounced / dʒ / and / ʃ / are level pegging (see Table 7.21, P. 159). There are five levels here. Mispronounced / ʃ / is again not well tolerated.

The English-speaking women listening to the *male* voices produced a hierarchy identical to that for everyone all together (see Table 7.22, P. 159). Mispronounced / ʃ / is equal with mispronounced / tʃ / on the second to top level.

Overall the hierarchies of preferences as shown by the English-speaking men and the English-speaking women separately reveal nothing surprising apart from the truncated hierarchy of the men. Even the discrepancy found here in evaluations of mispronounced / ʃ / in female voices v. male voices is a repetition of that found in evaluations by English-speaking men and women all together (see Table 7.3 and Table 7.4): both the men and the women separately downgrade mispronounced / ʃ / in a female voice, but upgrade it in a male voice. There is no crossover in preferences here, and hence no evidence for suggesting that the informants exhibit own-gender solidarity or own-gender betrayal. However, since overall the men have only a four-level hierarchy when listening to female voices, there is perhaps some reason for tentatively speculating that when evaluating female voices men are less critical, less categorical, overall, than women: men would seem to be reluctant to criticise the fairer sex. Women, on the other hand, are just as sharp with male voices as with female voices (both five-level hierarchies).

Table 7.19 English-speaking men listening to *female voices*

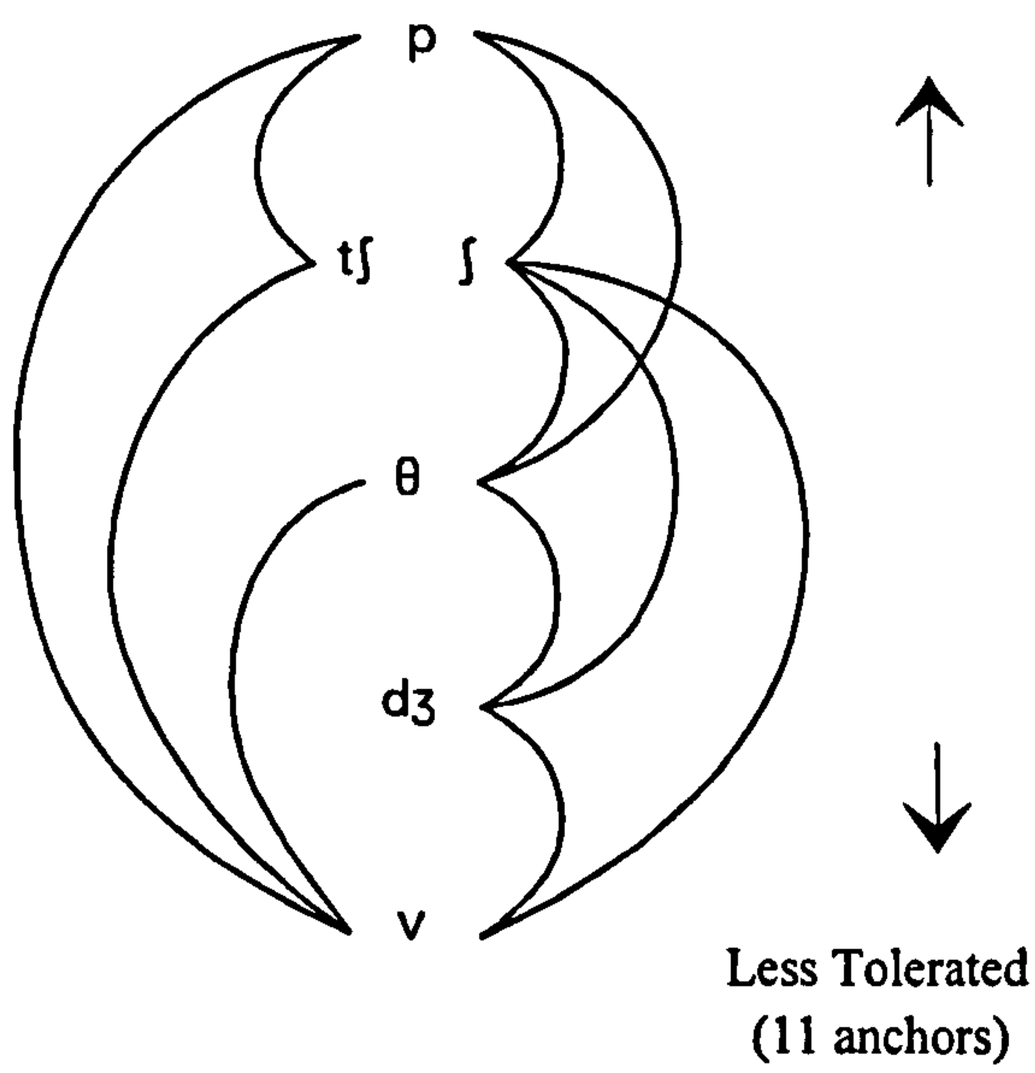
Contrast	%	Tape
/θ - ʃ/	100.0 ***	T2
/dʒ - v/	100.0 ***	T1
/p - v/	100.0 ***	T2
/p - dʒ/	95.0 ***	T1
/tʃ - v/	95.0 ***	T1
	90.0 ***	T1
/ʃ - v/	94.4 ***	T2
/tʃ - dʒ/	90.0 ***	T1
/θ - dʒ/	88.9 ***	T2
/θ - v/	80.0 **	T1
/p - ʃ/	94.4 ***	T2
	94.4 ***	T2
/dʒ - ʃ/	83.3 **	T2



Notes: For T1, $N = 20$; for T2, $N = 18$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: /tʃ - θ/ 80.0 ** T1
/θ - tʃ/ 77.8 * T2

Table 7.20 English-speaking men listening to *male voices*

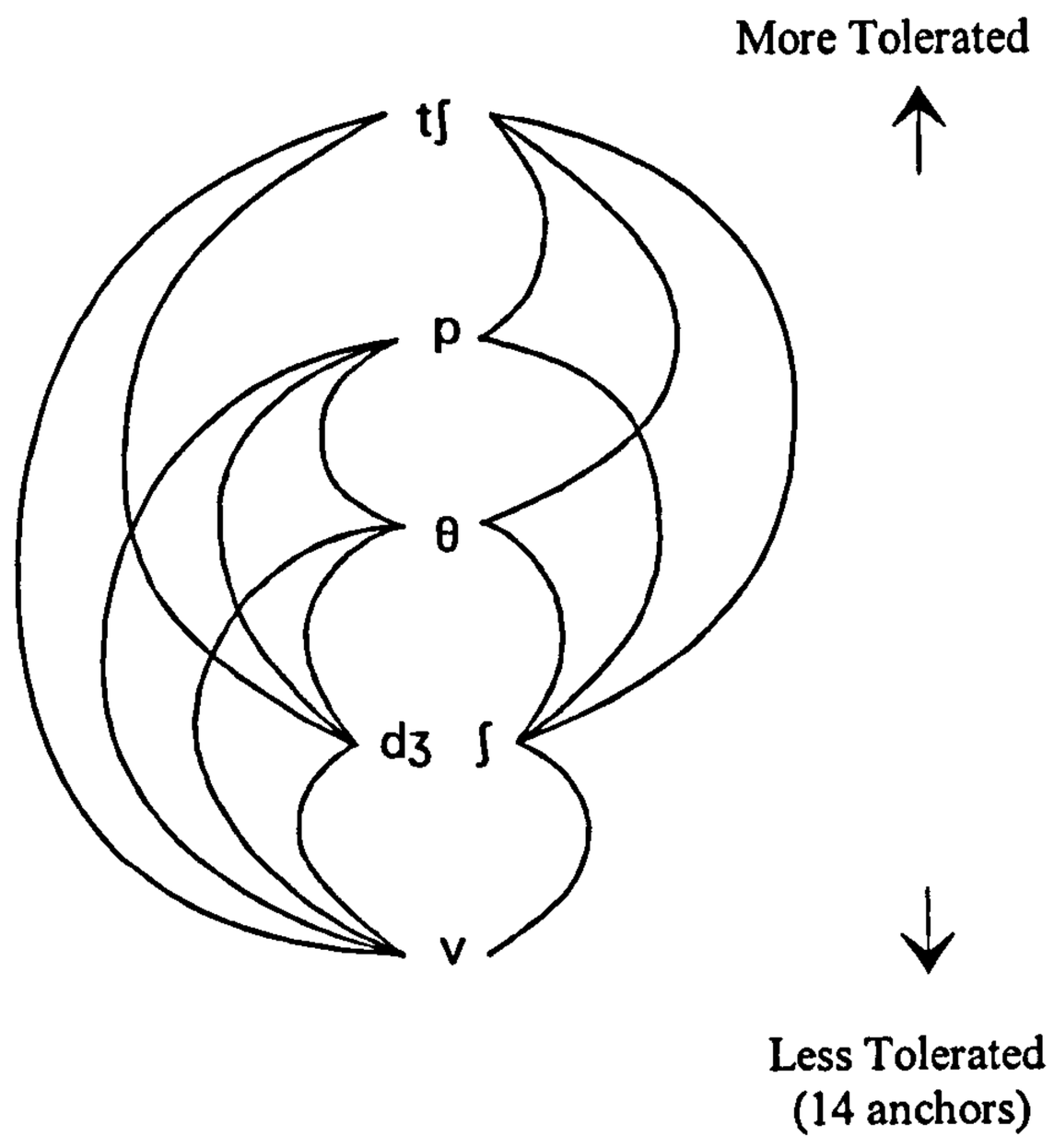
Contrast	%	Tape
/tʃ - v/	100.0 ***	T2
	100.0 ***	T2
/p - θ/	95.0 ***	T1
/θ - dʒ/	95.0 ***	T1
/p - v/	95.0 ***	T1
/ʃ - v/	95.0 ***	T1
	90.0 ***	T1
/dʒ - v/	94.4 ***	T2
/p - tʃ/	94.4 ***	T2
/ʃ - θ/	90.0 ***	T1
/p - ʃ/	90.0 ***	T1
	90.0 ***	T1
/θ - v/	88.9 ***	T2
/ʃ - dʒ/	75.0 *	T1



Notes: For T1, $N = 20$; for T2, $N = 18$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.21 English-speaking women listening to *female voices*

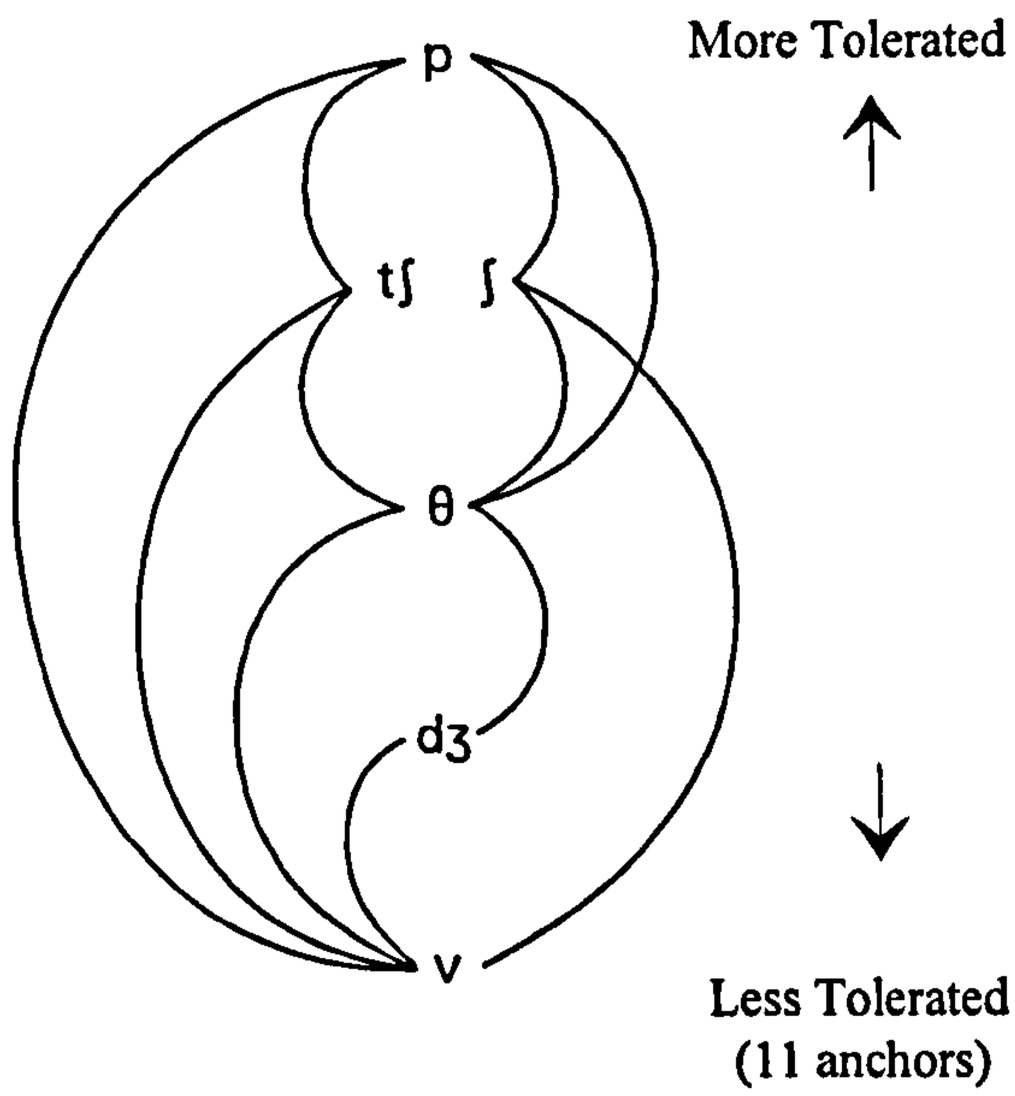
Contrast	%	Tape
/p - ʃ/	100.0 ***	T2
	100.0 ***	T2
/θ - ʃ/	100.0 ***	T2
/p - dʒ/	100.0 ***	T1
/tʃ - dʒ/	100.0 ***	T1
/p - v/	100.0 ***	T2
/ʃ - v/	97.2 ***	T2
	75.0 **	T2
/dʒ - v/	96.9 ***	T1
/tʃ - v/	96.9 ***	T1
	90.6 ***	T1
/θ - v/	93.8 ***	T1
/θ - dʒ/	88.9 ***	T2
/tʃ - p/	81.3 ***	T1
/p - θ/	78.1 **	T1
/tʃ - ʃ/	75.0 **	T2
/tʃ - θ/	68.8 *	T1



Notes: For T1, $N = 32$; for T2, $N = 36$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.22 English-speaking women listening to *male voices*

Contrast	%	Tape
/dʒ - v/	100.0 ***	T2
/p - tʃ/	91.7 ***	T2
/p - ʃ/	84.4 ***	T1
	84.4 ***	T1
/tʃ - θ/	84.4 ***	T1
	75.0 **	T2
/ʃ - θ/	81.3 ***	T1
/ʃ - v/	96.9 ***	T1
	96.9 ***	T1
/p - v/	93.8 ***	T1
/p - θ/	96.9 ***	T1
	77.8 ***	T2
/tʃ - v/	97.2 ***	T2
	97.2 ***	T2
/θ - dʒ/	90.6 ***	T1
	75.0 **	T2
/θ - v/	88.9 ***	T2



Notes: For T1, $N = 32$; for T2, $N = 36$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

However, crosstabulations using Cramér's *V* were performed on informant gender for the various individual mispronunciation contrasts and *no* statistically significant values were found, i.e. the preferences of both the men and the women, whether listening to a female voice or a male voice, were the same. Although this is a 'non-result', the uniformity of responses is important: the gender of the English-speaking informants was not a moderating variable for the individual contrasts.

7.3.2. Finnish-speaking informants

The hierarchy tables constructed on the basis of the Chi Square analyses performed on the data from the Finnish-speaking men and the Finnish-speaking women listening to both the female voices and the male voices are broadly similar—as expected—to those for all Finnish-speaking listeners together (see Table 7.5 and Table 7.6), with the exception of the Finnish-speaking men listening to female voices.

With the Finnish-speaking men listening to the *female* voices, only three vertical levels and eight anchors are to be found (Table 7.23, P. 161). In other words the Finnish-speaking men were relatively uncritical as regards the hierarchy for the female voices: a case of lenience (or deference?) towards the opposite gender; everything is almost equally good (or equally bad). The position of mispronounced / θ / relative to mispronounced / dʒ /, / tʃ / and / ʃ / is noticeably unclear: / θ / has no link to any of these three mispronunciations. Is / θ / above, below or level with them in the hierarchy?

With the Finnish-speaking men listening to the *male* voices, the hierarchy is extremely unstable structurally, though there are again eight anchors (Table 7.24, P. 161). There appears to be four levels with a problematic fifth one at the bottom, the position of mispronounced / p / being ambiguous and confusing (and I can find no mistake in the analysis to explain this). Moreover mispronounced / v / only has one anchor to hold it in place (as also / dʒ / if one ignores the dubious / p / anchor). The overall picture seems weak and fuzzy, suggesting a considerable degree of uncertainty among the men as to how to evaluate own-gender voices.

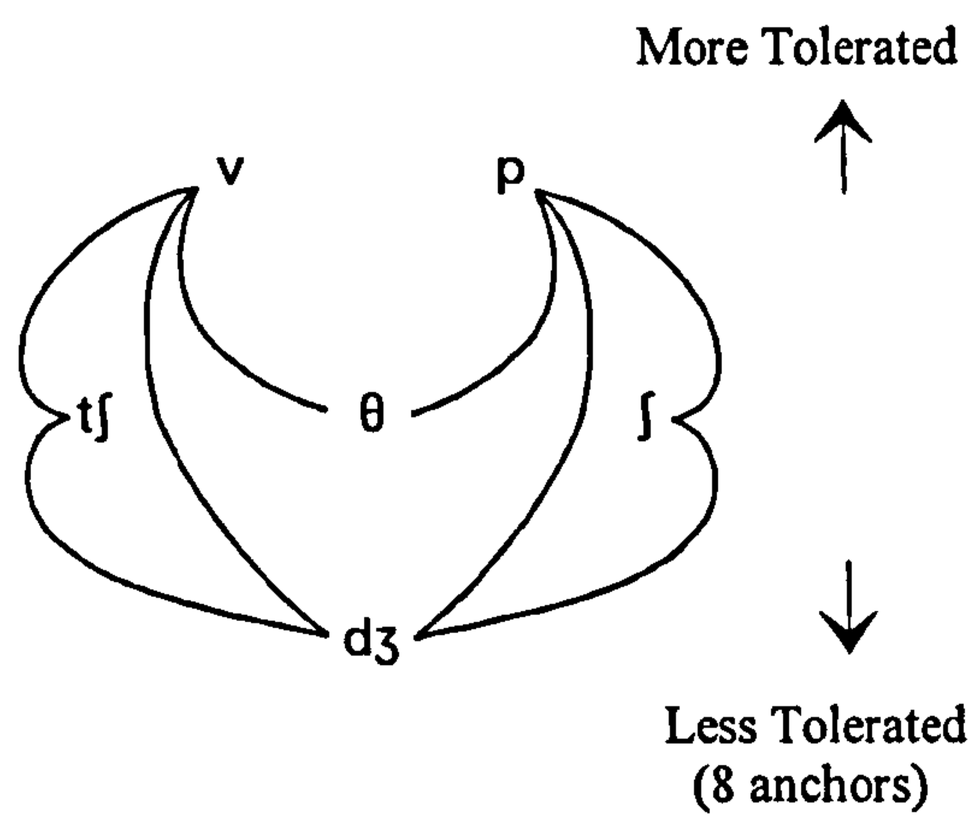
When the Finnish-speaking women listened to *female* voices, there was a relatively stable structure: nine anchors and four levels (Table 7.25, P. 162). However, mispronounced / dʒ / is particularly unstable since it has only one anchor (N.B. there were two contradictory results for / dʒ / v. / θ /, one of the mispronunciations being upgraded for one of the female voices, but being downgraded for the other).

When the Finnish-speaking women listened to *male* voices, there were eight anchors and four levels, plus a problematic fifth level involving / ʃ / (Table 7.26, P. 162). The position of / dʒ / in the hierarchy is again unstable, for it only has one anchor if one ignores the dubious / ʃ /. The distinct separation of mispronounced / v, p, tʃ, ʃ / onto two levels is belied by two contradictory pairs of results: / v - ʃ / and / p - tʃ /: the four sounds could be viewed as being very much of a muchness, with no marked distinction between them.

The difference between men and women in their reactions to *female* voices mispronouncing / ʃ / and / p /, which is witnessed by the positions of these two sounds in the hierarchies for men and women (Table 7.23 and Table 7.25), is confirmed to some extent by the values for Cramér's *V* based on gender of informant (Table 7.27, P. 163). The men and women react to these two mispronounced phonemes in mildly different ways when they are set against mispronounced / tʃ /: men have no significant preference for either sound, but women prefer / tʃ / both times.

Table 7.23 Finnish-speaking men listening to *female voices*

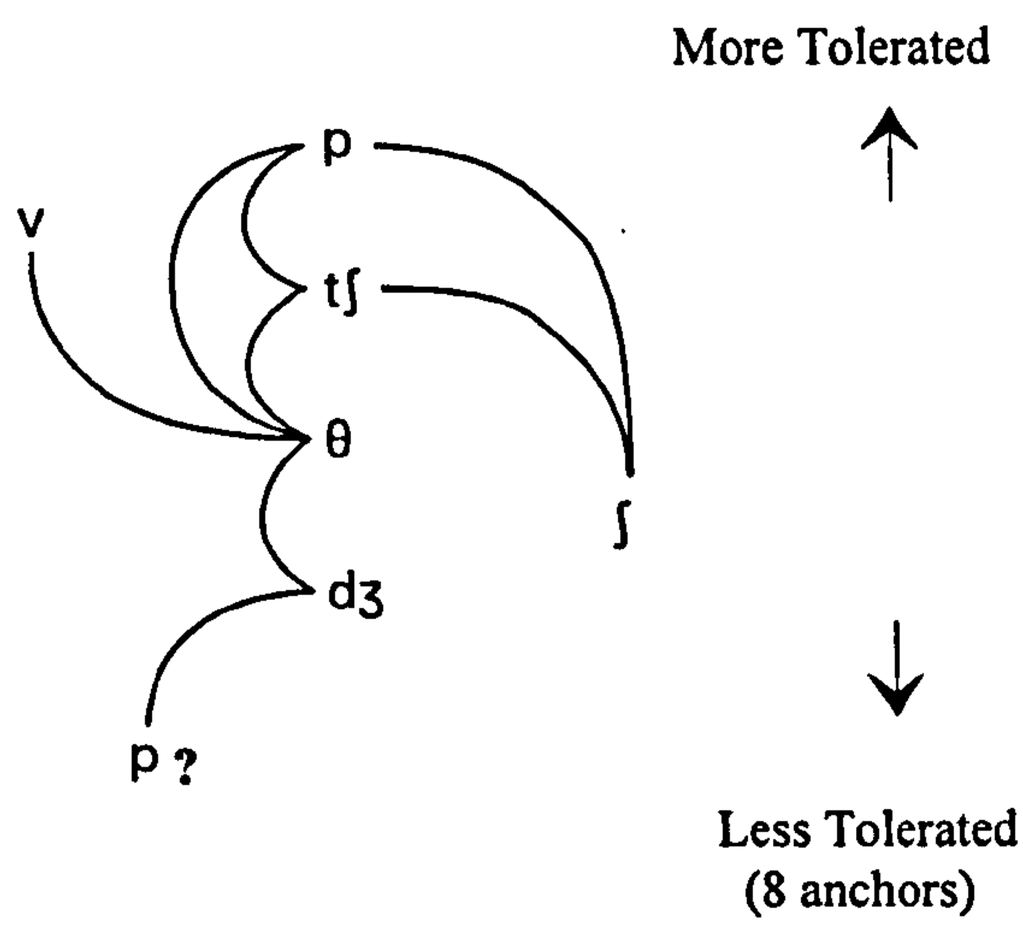
Contrast	%	Tape
/ tʃ - dʒ /	93.2 ***	T1
	65.9 *	T1
/ v - θ /	90.9 ***	T1
/ p - θ /	84.1 ***	T1
	79.1 ***	T2
/ v - tʃ /	84.1 ***	T1
	65.9 *	T1
/ p - ʃ /	79.1 ***	T2
	69.8 **	T2
/ v - ʃ /	72.1 **	T2
/ v - dʒ /	65.9 *	T1
/ ʃ - dʒ /	65.1 *	T2



Notes: For T1, $N = 44$; for T2, $N = 43$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.24 Finnish-speaking men listening to *male voices*

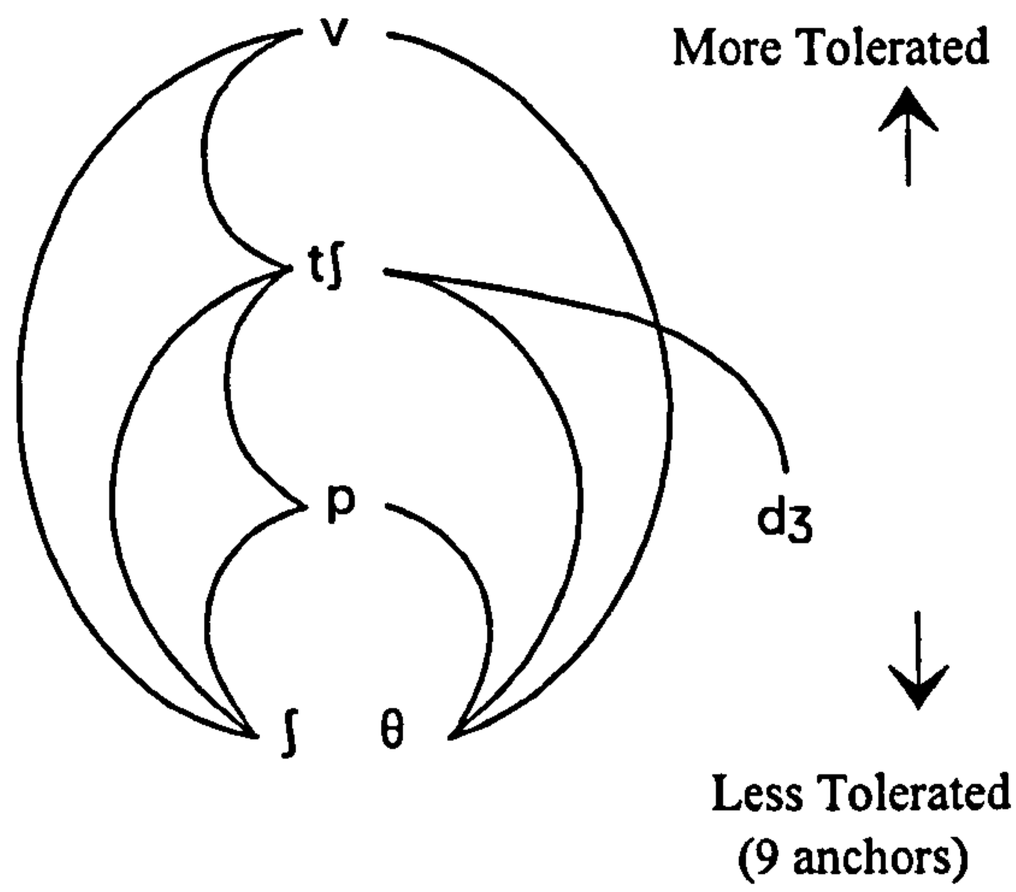
Contrast	%	Tape
/ θ - dʒ /	79.5 ***	T1
	69.8 **	T2
/ tʃ - θ /	77.3 ***	T1
/ p - tʃ /	76.7 ***	T2
/ tʃ - ʃ /	75.0 ***	T1
/ dʒ - p /	74.4 **	T2
/ v - θ /	69.8 **	T2
/ p - ʃ /	68.2 *	T1
/ p - θ /	68.2 *	T1



Notes: For T1, $N = 44$; for T2, $N = 43$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.25 Finnish-speaking women listening to *female voices*

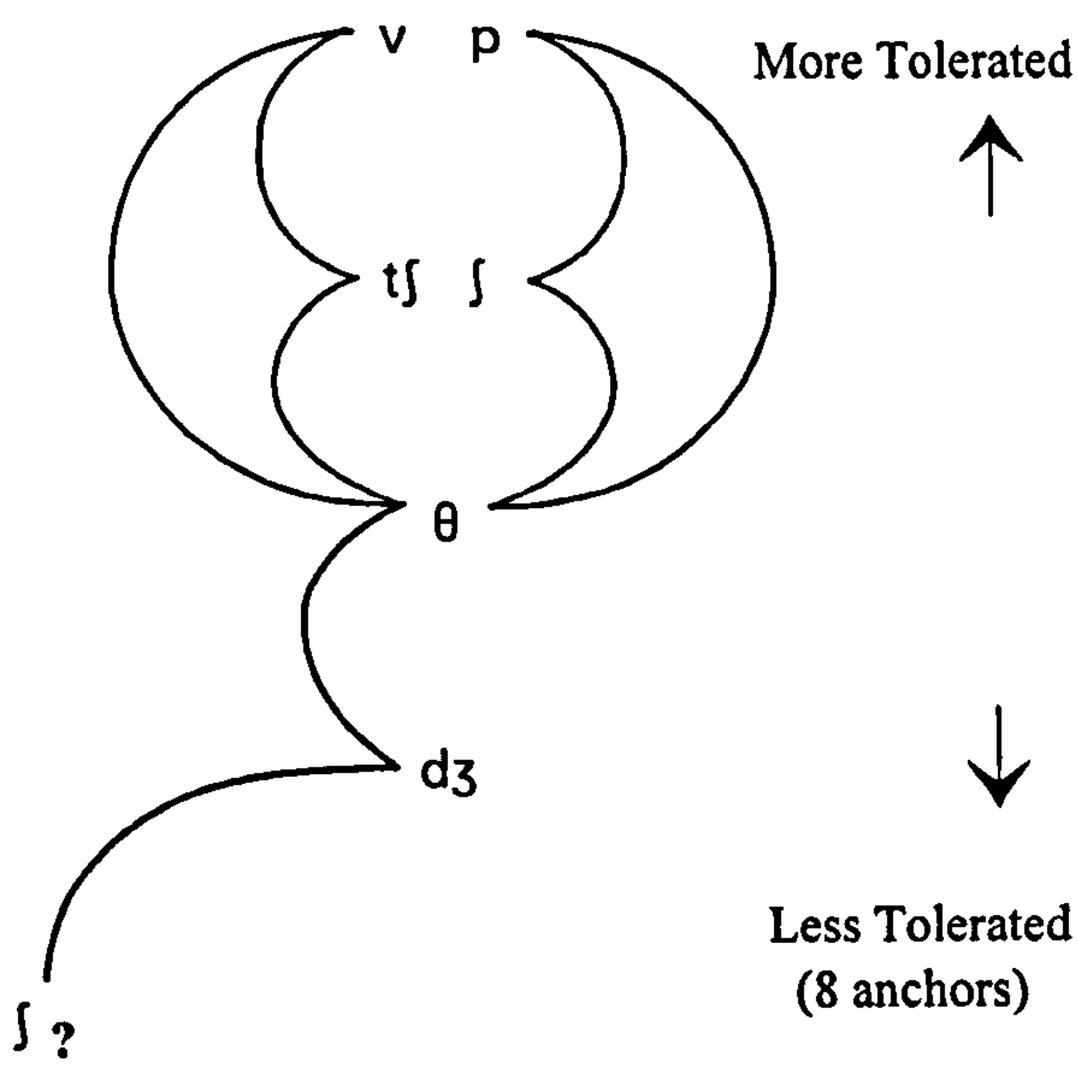
Contrast	%	Tape
/ tʃ - dʒ /	94.5 ***	T1
	80.0 ***	T1
/ p - θ /	92.7 ***	T1
	92.6 ***	T2
/ v - θ /	87.3 ***	T1
/ p - ʃ /	87.0 ***	T2
	79.6 ***	T2
/ tʃ - p /	81.8 ***	T1
/ v - tʃ /	81.8 ***	T1
	67.3 *	T1
/ v - ʃ /	74.1 ***	T2
/ tʃ - ʃ /	68.5 **	T2
/ tʃ - θ /	64.8 *	T2



Notes: For T1, $N = 55$; for T2, $N = 54$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: / dʒ - θ / 67.3 * T1
/ θ - dʒ / 70.4 ** T2

Table 7.26 Finnish-speaking women listening to *male voices*

Contrast	%	Tape
/ p - θ /	81.8 ***	T1
	77.8 ***	T1
/ θ - dʒ /	81.8 ***	T1
	72.2 **	T2
/ v - θ /	77.8 ***	T2
/ tʃ - θ /	76.4 ***	T1
	72.2 **	T2
/ p - ʃ /	69.1 **	T1
/ dʒ - ʃ /	67.3 **	T1
/ ʃ - θ /	65.5 *	T1
/ v - tʃ /	64.8 *	T2



Notes: For T1, $N = 55$; for T2, $N = 54$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
+ 2 contradictory results: / v - ʃ / 70.9 ** T1
/ ʃ - v / 63.6 * T1
/ p - tʃ / 88.9 *** T2
/ tʃ - p / 64.8 * T2

Table 7.27 Crosstabulation on gender of Finnish-speaking informants

		Men	Women	Cramér's *V*
Female voices				
/ tʃ - ʃ /	T2	60.5% (ʃ) [a]	68.5% **	.29 **
/ ∅ - ʃ /	T2	62.8%	87.0% ***	.28 **
/ tʃ - p /	T1	63.6%	81.8% ***	.21 *
Male voices				
/ ∅ - dʒ /	T2	83.7% ***	98.1% ***	.26 * [b]
/ tʃ - ʃ /	T1	75.0% ***	50.9%	.25 *
/ ∅ - tʃ /	T2	60.5%	79.6% ***	.21 *

Notes: T2 (Tape 2): men $N = 43$, women $N = 54$.
T1 (Tape 1): men $N = 44$, women $N = 55$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

[a] My convention in the contrast column is to present the preferred mispronunciation as the first phoneme: thus for the / tʃ - ʃ / contrast in this row, mispronounced / tʃ / is preferred to mispronounced / ʃ /. However, the preferred mispronunciation is reversed in this cell, and this is indicated by the new preferred mispronunciation being shown in brackets: 60.5% (ʃ).

[b] This value for *V* is unreliable since 1 of the 4 cells in the crosstabulation had a value falling below 5.

The difference between men and women in their evaluations of the mispronunciation contrasts with *male* voices is restricted to the / tʃ - ʃ / contrast (Table 7.27). The men clearly prefer mispronounced / tʃ / to mispronounced / ʃ /, the women by contrast being split almost fifty-fifty. Curiously the contrasts involving mispronounced / p / furnished no values for Cramér's *V* that were statistically significant.

Overall the hierarchies of preferences as shown by the Finnish-speaking men are characterised by more fuzziness and lack of sharpness compared to those shown by the Finnish-speaking women.

Unlike the situation for the English-speaking listeners, the crosstabulations using Cramér's *V* performed on gender for the various separate mispronunciations revealed a few gender contrasts: see Table 7.27.

Five of the six results here (all except the first row) are a matter of degree: the direction of preference is the same but the percentage of votes cast is different. However, the Finnish-speaking men, though not the women, were noncommittal (i.e. their results were not statistically significant) on four of the contrasts: the four involve only / ʃ, tʃ / and / p /. Three of these four evaluations were with a female voice, so one might be tempted to suggest that the men are displaying a certain hesitation or tolerance with the female voice.

A clear crossover phenomenon is seen with the / tʃ - ʃ / contrast for with a *female* voice the men are neutral while the women prefer mispronounced / tʃ /, but with a *male* voice the women are neutral and the men prefer mispronounced / tʃ /. For each case separately, gender of informant is acting as a moderating variable, the value for Cramér's *V* indicating this. The crossover of results,

however, would also seem to imply own-gender intolerance (or tolerance/indifference towards other-gender mispronunciation), but I must stress yet again that gender of voice is not an isolatable control variable. The result is thought-provoking, but factors other than gender of voice may be operating here and distorting the picture.

7.4. Influence of age of informants

Just as the data was analysed to test for the influence of informant gender on responses, so the data was analysed to test for the influence of informant age. Does the age of informants lead to a moderated relationship? (Figure 7.1)

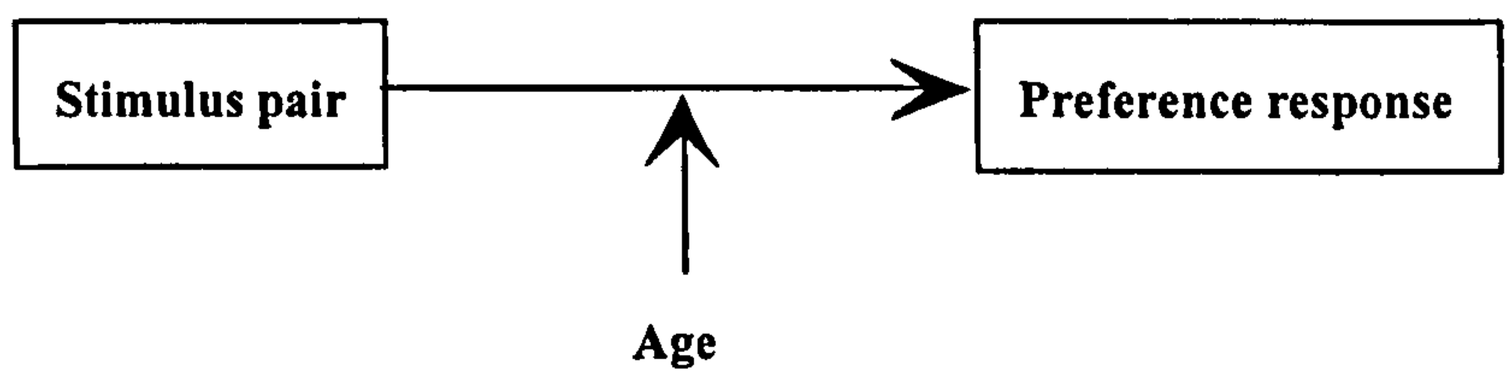


Figure 7.1 Influence of informant age on response

The statistics for the ages of the informants who participated in the experiment can be seen in Table 7.28.

Table 7.28 Ages of informants in Experiment 1

	N	Mean age	Min–Max	*sd*
Finnish-speaking listeners	196	23.71	19–47	5.11
– women	109	23.15	19–43	4.59
– men	87	24.41	19–47	5.63
English-speaking listeners	106	27.05	17–76	11.50
– women	68	25.71	17–76	9.95
– men	38	29.45	18–76	13.65

An arbitrary dividing point was chosen for 'younger' and 'older' and set at 22/23 years old (see section 6.2.3. above). The Swedish-speaking informants were again excluded from this part of the study because the sample size was too small for any confident analyses to be performed.

7.4.1. English-speaking informants

The age distribution of the English-speaking listeners is given in Table 7.29.

Table 7.29 Age of English-speaking informants

	Informants *N*	<=22 years of age *N* (%)	>=23 years of age *N* (%)
Tape 1	52	30 (57.7)	22 (42.3)
Tape 2	54	28 (51.9)	26 (48.1)

Crosstabulations using Cramér's *V* were performed on informant age for the various contrasts, and only two contrasts—both pronounced with a male voice—yielded values for *V* that are statistically significant (Table 7.30). For both contrasts here there is a 14% difference in the proportion of informants preferring one mispronunciation to the other; hence the mildly significant value of *V*. However, since both groups of informants (both younger and older) prefer the *same* mispronunciations for both contrasts, and prefer them at the same $p < .001$ level, this finding is not very revealing. The results of the whole crosstabulation are another 'non-result', but the striking thing is that the results strongly suggest that informant age (with the arbitrary division set at 22/23 years of age) is *not* a moderating variable.

Table 7.30 Crosstabulation on age of English-speaking informants

		<=22 years of age	>=23 years of age	Cramér's *V*
Male voices				
/ v - p /	T1	100.0% ***	86.5% ***	.29 *
/ ∅ - θ /	T2	85.7% ***	100.0% ***	.27 *

Notes: T1: younger $N = 30$; older $N = 22$.
T2: younger $N = 28$; older $N = 26$.
Chi Square analysis: *** $p < .001$; * $p < .05$.

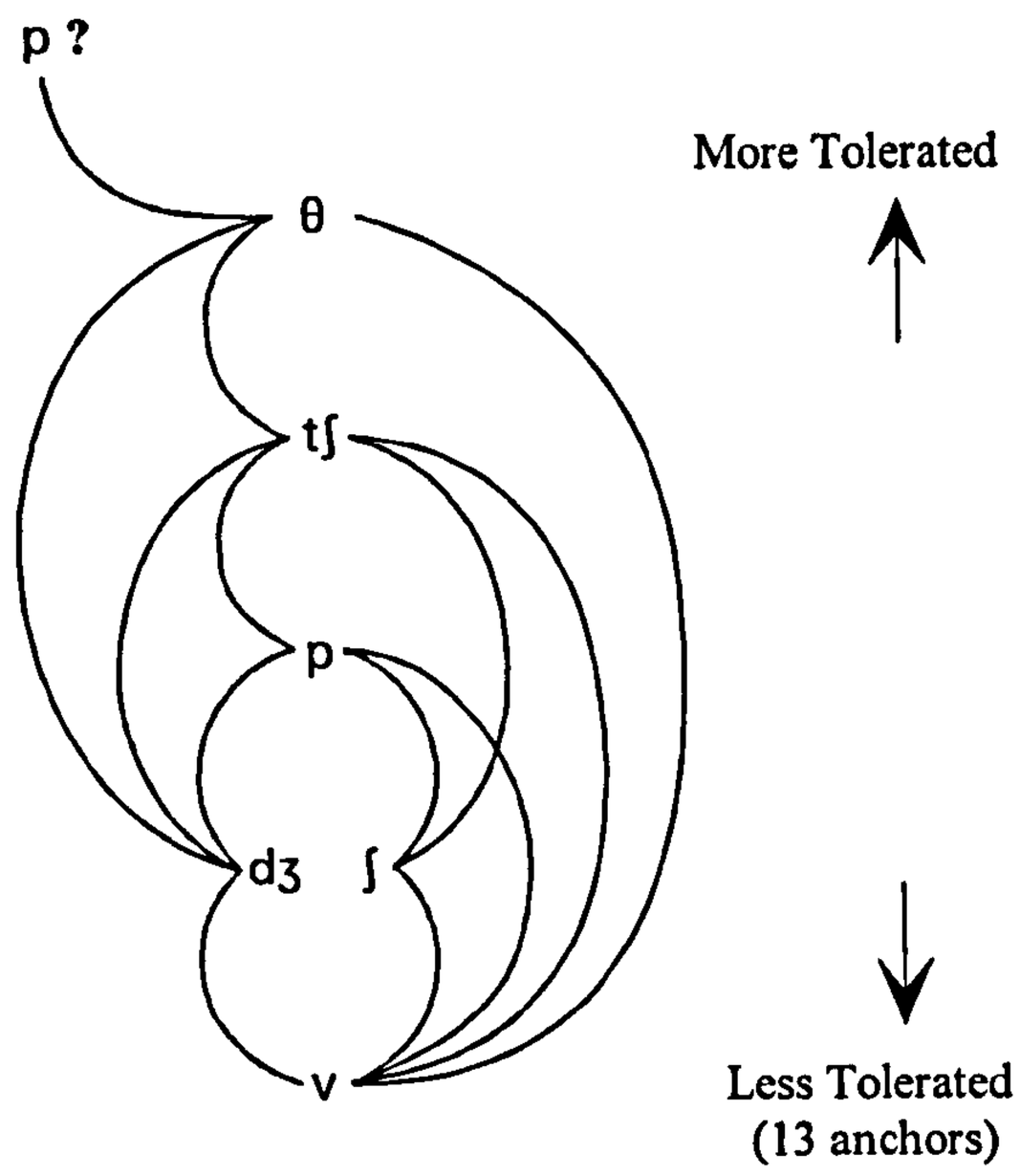
Despite this fact the hierarchy tables based on the Chi Square analysis of the data from the two age groups of English-speaking listeners are interesting because of the problems they pose.

The data for the *younger* English-speaking informants listening to *female* voices produced results which are difficult to put into a hierarchy table. There are thirteen anchors (i.e. thirteen significant contrasts in the contrast column of Table 7.31, P. 166), but the relative positions of / θ , tʃ / and / p / are problematic: / θ / is preferred to / tʃ /, and / tʃ / is preferred to / p /; so how is it that / p / is also preferred to / θ /? Since the / p - θ / contrast is significant at only the $p < .05$ level, one could put this contrast temporarily to one side and form a hierarchy with just twelve anchors: this produces the basic hierarchy found in Table 7.31 (a), and then the thirteenth anchor with / p / (at the top) has the question mark. Alternatively, rather than momentarily side-stepping the / p - θ / contrast, one could interpret the structure as in hierarchies (b) or (c), with question marks against / tʃ / and / θ / respectively. Notice that / p, θ, tʃ / are precisely those sounds which are found together in joint top place when English-speaking *men* are listening to *female* voices (Table 7.19). It would seem therefore that young English-speaking people as a group, and all English-speaking men as a group, are relatively unstable in their opinions when choosing between alternative mispronunciations of / p, θ / and / tʃ / as voiced by women.

Table 7.31 Younger English-speaking informants (<=22) listening to *female voices*

Contrast	%	Tape
/p - v/	100.0 ***	T2
/θ - ʃ/	100.0 ***	T2
/p - dʒ/	96.7 ***	T1
/p - ʃ/	96.4 ***	T2
	96.4 ***	T2
/dʒ - v/	96.7 ***	T1
/tʃ - dʒ/	93.3 ***	T1
/tʃ - v/	93.3 ***	T1
	86.7 ***	T1
/ʃ - v/	92.9 ***	T2
	75.0 **	T2
/θ - v/	90.0 ***	T1
/θ - dʒ/	89.3 ***	T2
/θ - tʃ/	78.6 **	T2
/tʃ - p/	76.7 **	T1
/p - θ/	73.3 *	T1

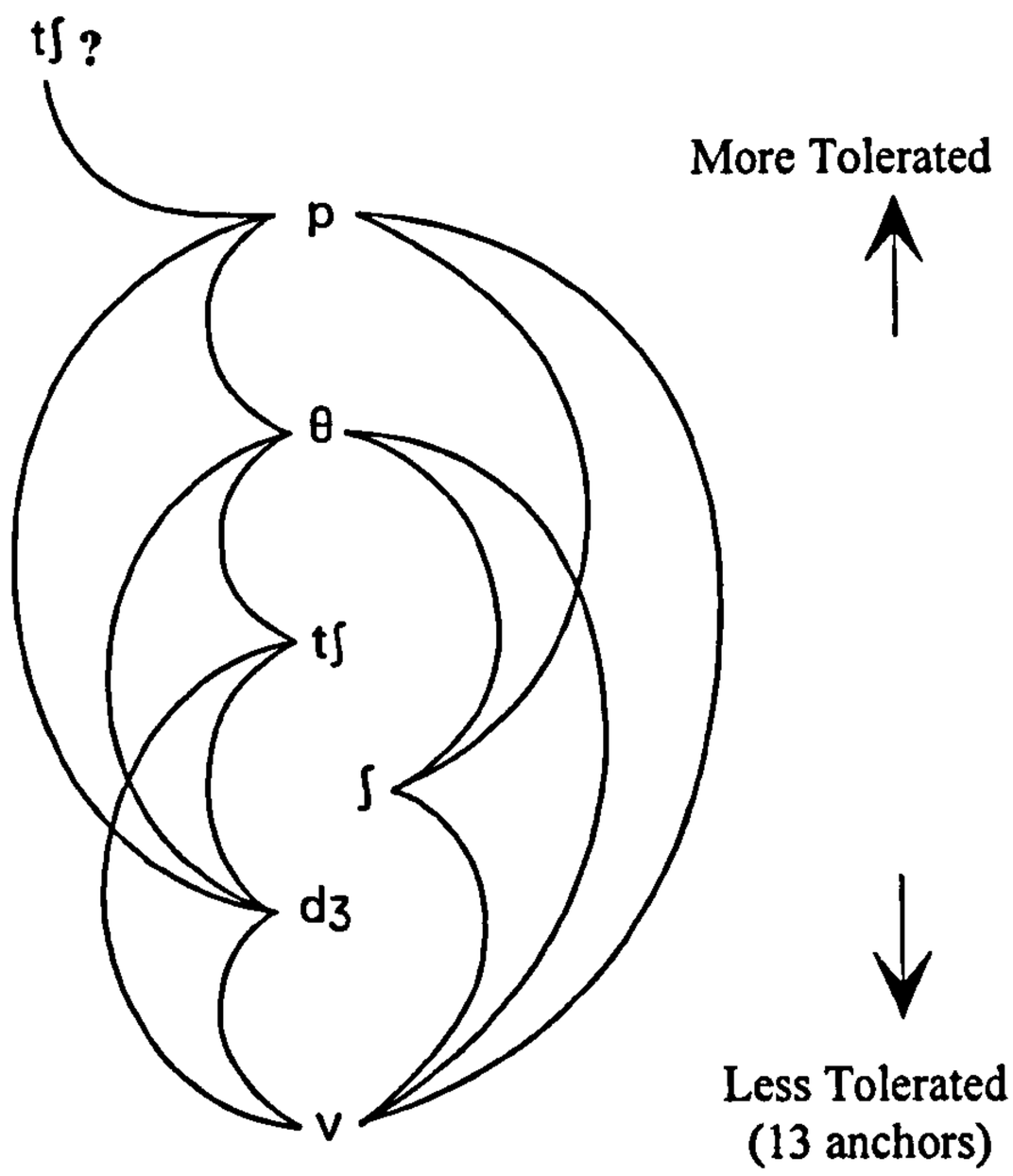
(a)



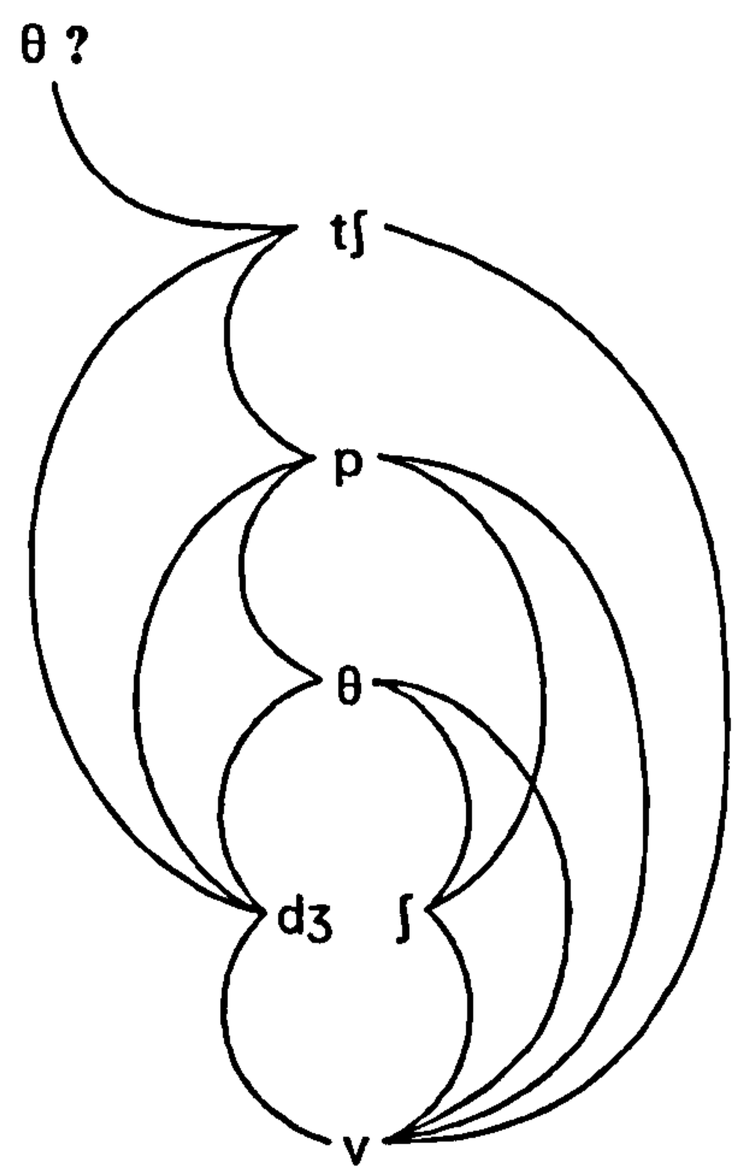
Notes: For T1, $N = 30$; for T2, $N = 28$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Alternatively:

(b)



(c)



By contrast—and startlingly so—the hierarchy for the *older* English-speaking informants listening to *female* voices (Table 7.33, P. 168) is extremely secure, offering the clearest hierarchical picture possible for mispronunciation preferences: six levels (the maximum possible) and fifteen anchors (the maximum possible), with all mispronounced phonemes having five anchors. The preferences displayed are extremely confident and consistent. (The next clearest hierarchical picture is that of English-speaking *women* also listening to *female* voices: 5 levels and 14 anchors. Table 7.21.) Older people presumably are more experienced in their interactions with women, and are more confident in their opinions.

The hierarchical picture for the *older* English-speaking informants listening to *male* voices (Table 7.34, P. 169) is not quite so clear-cut: five levels and eleven anchors. However, the younger and older English-speaking informants display identical hierarchies for the *male* voices (Table 7.32, P. 168, and Table 7.34), though the younger informants have 12 anchors and the older ones 11. The identicality with the situation for men and women all together (Table 7.4), and for women on their own (Table 7.22), when also listening to *male* voices, is striking.

In brief, in the English-speaking listeners' evaluations of Finnish-accented mispronunciations in *male* voices, there is noticeable congruity and uniformity across gender and age.

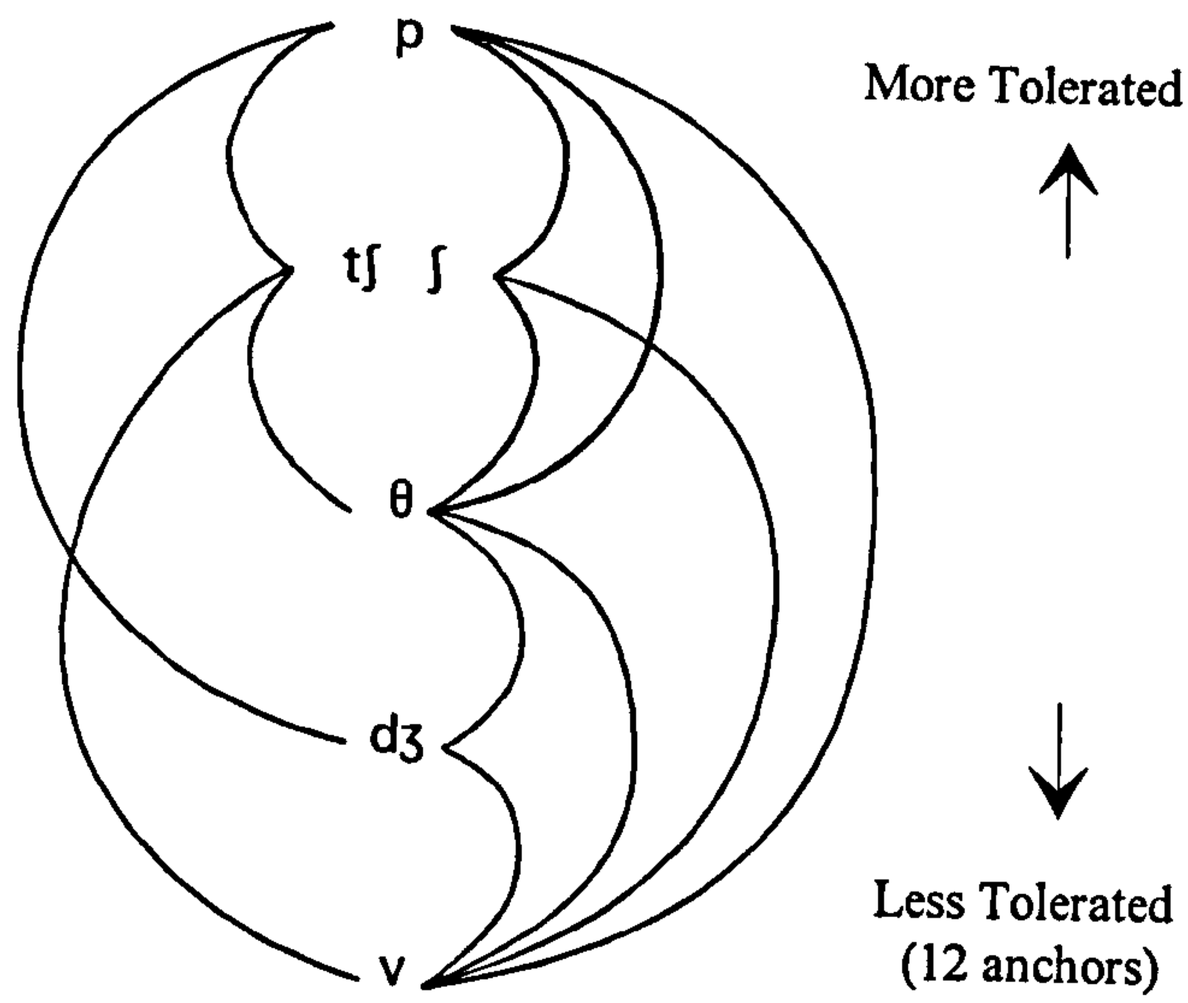
It is in evaluations of mispronunciations in *female* voices where there is diversity and divergence, though the variation focuses on the top of the scale with the three most tolerated mispronunciations / tʃ, p, θ /. Depending on the informant group—everyone all together (Table 7.3), men alone (Table 7.19), women alone (Table 7.21), younger people (Table 7.31) or older people (Table 7.33)—these three fortis sounds align themselves in various ways, the young people being particularly precarious in their evaluations of these mispronounced sounds.

The three other mispronunciations, viz. / dʒ, ʃ, v /, are either in that decreasing order (with mispronounced / v / firmly at the bottom) or else / dʒ / and / ʃ / are level pegging (with / v / again at the bottom).

One is forced to the conclusion that English-speaking listeners (men and women, younger and older) are more confident (less deferent) in making decisions with a male voice than with a female voice, though why this should be so is a matter for speculation. Is the *aural image* of what a foreign-accented male voice should sound like clearer than that for a female voice? Are the expectations made of a male voice more congruent and widely recognised? Is the relative hesitancy shown towards a foreign female voice an indication that a foreign-accented female voice is given more 'leeway' in variation (i.e. is shown more tolerance) than a male voice? A qualitative study might shed some light on this subject.

Table 7.32 Younger English-speaking informants (<=22) listening to *male voices*

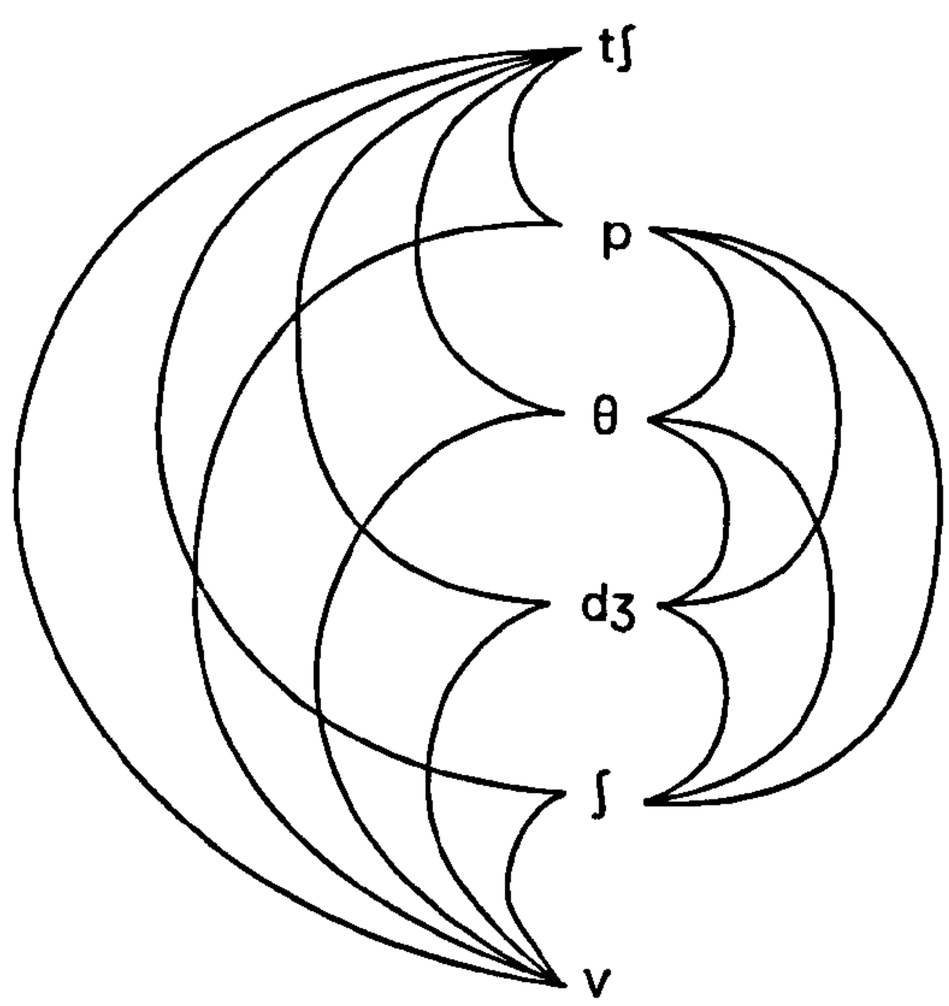
Contrast	%	Tape
/p - v/	100.0 ***	T1
/dʒ - v/	100.0 ***	T2
/tʃ - v/	100.0 ***	T2
	96.4 ***	T2
/ʃ - v/	100.0 ***	T1
	93.3 ***	T1
/p - θ/	100.0 ***	T1
	78.6 **	T2
/p - tʃ/	92.9 ***	T2
/θ - dʒ/	90.0 ***	T1
	78.6 **	T2
/θ - v/	89.3 ***	T2
/p - ʃ/	86.7 ***	T1
	83.3 ***	T1
/tʃ - θ/	83.3 ***	T1
/ʃ - θ/	76.7 **	T1
/p - dʒ/	71.4 *	T2



Notes: For T1, $N = 30$; for T2, $N = 28$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: /dʒ - tʃ/ 99.3 *** T2
/tʃ - dʒ/ 96.4 *** T2

Table 7.33 Older English-speaking informants (>=23) listening to *female voices*

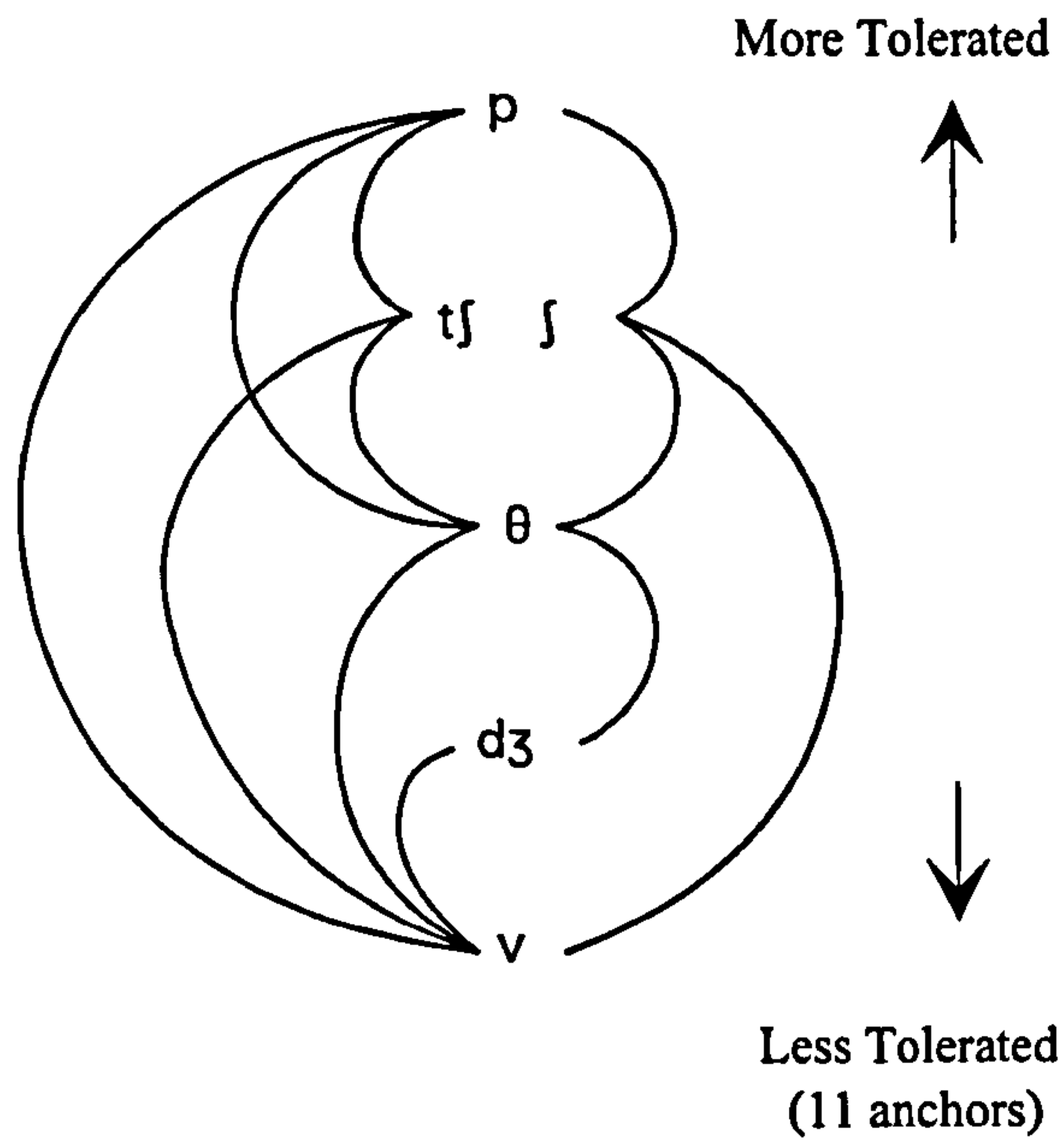
Contrast	%	Tape
/p - v/	100.0 ***	T2
/ʃ - v/	100.0 ***	T2
/dʒ - v/	100.0 ***	T1
/θ - ʃ/	100.0 ***	T2
/p - dʒ/	100.0 ***	T1
/tʃ - dʒ/	100.0 ***	T1
/p - ʃ/	100.0 ***	T2
	100.0 ***	T2
/tʃ - v/	100.0 ***	T1
	95.5 ***	T1
/θ - dʒ/	88.5 ***	T2
/θ - v/	86.4 ***	T1
/tʃ - θ/	86.4 ***	T1
/dʒ - ʃ/	80.8 **	T2
/tʃ - ʃ/	76.9 **	T2
/tʃ - p/	77.3 *	T1
/p - θ/	72.7 *	T1



Notes: For T1, $N = 22$; for T2, $N = 26$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.34 Older English-speaking informants (>=23) listening to *male voices*

Contrast	%	Tape
/ tʃ - v /	100.0 ***	T2
	96.2 ***	T2
/ dʒ - v /	96.2 ***	T2
/ ʃ - θ /	95.5 ***	T1
/ ʃ - v /	95.5 ***	T1
	90.0 ***	T1
/ θ - dʒ /	95.5 ***	T1
/ p - tʃ /	92.3 ***	T2
/ p - θ /	90.9 ***	T1
/ p - ʃ /	90.9 ***	T1
	86.4 ***	T1
/ θ - v /	88.5 ***	T2
/ p - v /	86.4 ***	T1
/ tʃ - θ /	76.9 **	T2
	72.7 *	T1



Notes: For T1, $N = 22$; for T2, $N = 26$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: / tʃ - dʒ / 96.2 *** T2
/ dʒ - tʃ / 84.6 *** T2

7.4.2. Finnish-speaking informants

The age distribution of the Finnish-speaking listeners is given in Table 7.35.

Table 7.35 Age of Finnish-speaking informants

	Informants *N*	<=22 years of age *N* (%)	>=23 years of age *N* (%)
Tape 1	99	48 (48.5)	51 (51.5)
Tape 2	97	54 (55.7)	43 (44.3)

Crosstabulations using Cramér's *V* were performed on informant age for the various contrasts, and nine significant values were found (Table 7.36, P. 170).

The two contrasts which involve a correct pronunciation v. a mispronunciation (/ ∅ - v / and / ∅ - tʃ /) reveal that the older Finnish-speaking listeners were noticeably weaker than the younger ones in recognising mispronounced / v / and / tʃ / as wrong, hence the relatively high values for Cramér's *V*, .25 and .30 respectively. The older listeners were also unable to decide on the / ∅ - v / contrast when pronounced by a man (56.9%, $p > .05$, not presented in Table 7.36), but the discrepancy with the younger listeners, who preferred the correct articulation (68.8%, $p < .01$), is not large enough to produce a statistically significant value for Cramér's *V* (.12, $p > .05$). Is this a reflection of the intolerance of youth, and/or the tolerance of experience and maturity?

Table 7.36 Crosstabulation on age of Finnish-speaking informants

		<=22 years of age	>=23 years of age	Cramér's *V*
Female voices				
/ tʃ - ʃ /	T2	70.4% **	62.8% (ʃ)	.33 **
/ p - θ /	T1	97.9% ***	80.4% ***	.28 **
/ ∅ - v /	T2	83.3% ***	60.5%	.25 *
/ ʃ - θ /	T2	55.6% (θ)	69.8% **	.25 *
/ ʃ - dʒ /	T2	68.5% **	53.5% (dʒ)	.22 *
/ θ - dʒ /	T2	74.1% ***	53.5%	.21 *
Male voices				
/ ∅ - tʃ /	T2	83.3% ***	55.8%	.30 **
/ v - p /	T1	56.3% (p)	64.7% *	.21 *
/ tʃ - dʒ /	T2	66.7% *	53.5% (dʒ)	.20 *

Notes: T1: younger $N = 48$; older $N = 51$.
T2: younger $N = 54$; older $N = 43$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

One contrast, / p - θ / with female voices, produced a significant difference in preference for one mispronunciation over the other of 17 percentage points, giving a value of Cramér's *V* of .28, but the preference expressed by both groups is in the same direction.

The remaining six contrasts in Table 7.36 are interesting in that in each case only one of the two informant groups (more often the younger listeners than the older ones) produces a result which is statistically significant. Moreover, for five of these six contrasts the direction of preference for the non-significant result of one group would seem to be in the opposite direction to the significant result of the other.

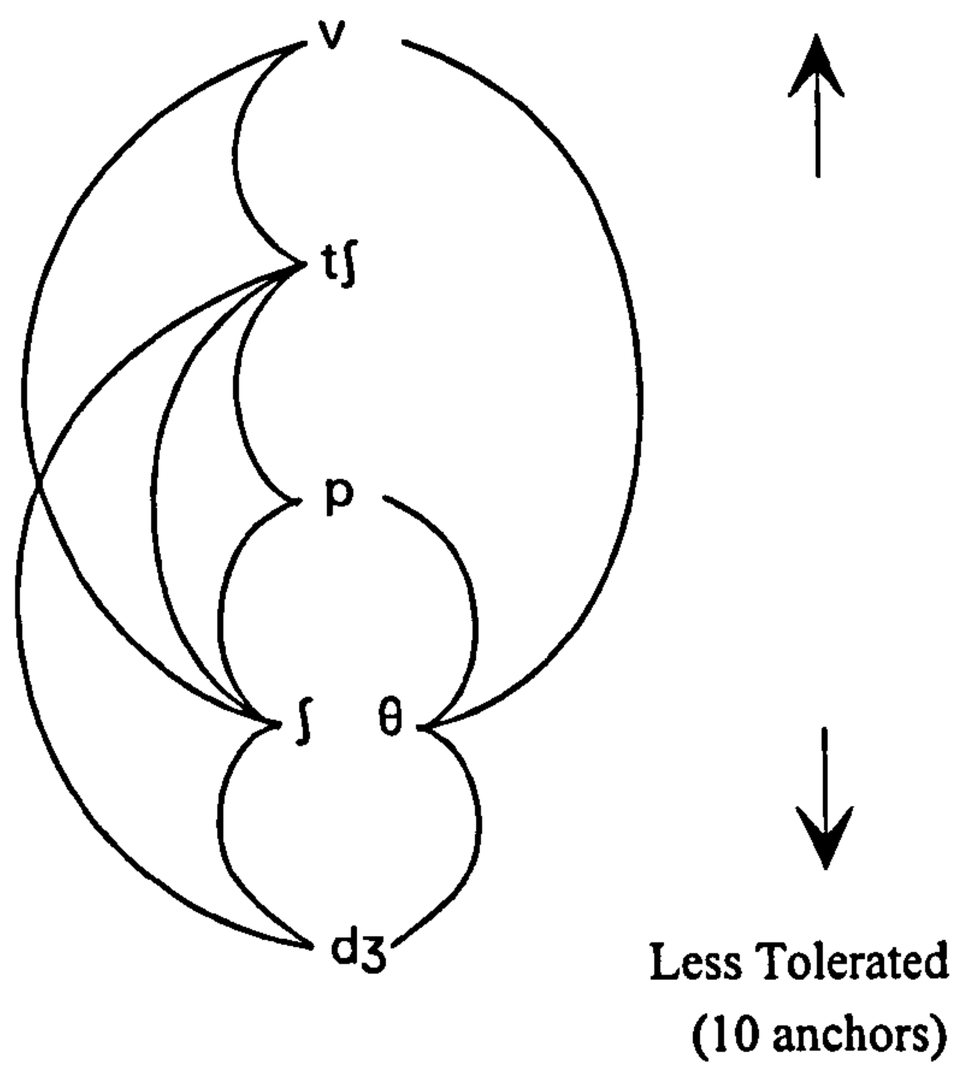
In brief, the cross-tabulation for Cramér's *V* gives the impression that the younger informants are more confident in making a decision than the older informants, are less tolerant of mispronunciation than the older informants.

As for the hierarchy tables derived from the Chi Square analysis of the data from the younger and older Finnish-speaking listeners, they would seem to be less secure than those revealed by the English-speaking informants: two of the four tables here have 10 anchors (Table 7.37, P. 171, and Table 7.40, P. 172), and two have 9 (Table 7.38, P. 171, and Table 7.39, P. 172). Nevertheless, despite a certain lack of clarity for the position of some mispronunciations, there is some measure of uniformity across the four hierarchies in that mispronounced / v, tʃ, p / would appear to be rated above mispronounced / ʃ, dʒ, θ /.

The *younger* Finnish-speaking informants listening to *female* voices (Table 7.37, P. 171) have a five-level hierarchy with mispronounced / ʃ / and / θ / sharing the same level. By contrast, the *older* Finnish-speaking informants listening to *female* voices (Table 7.39, P. 172) have a five-level hierarchy with mispronounced / ʃ / anchored above mispronounced / θ /, and mispronounced / dʒ / poorly anchored. It is mispronounced / dʒ /, therefore, which seems to cause

Table 7.37 Younger Finnish-speaking informants (<=22) listening to *female voices*

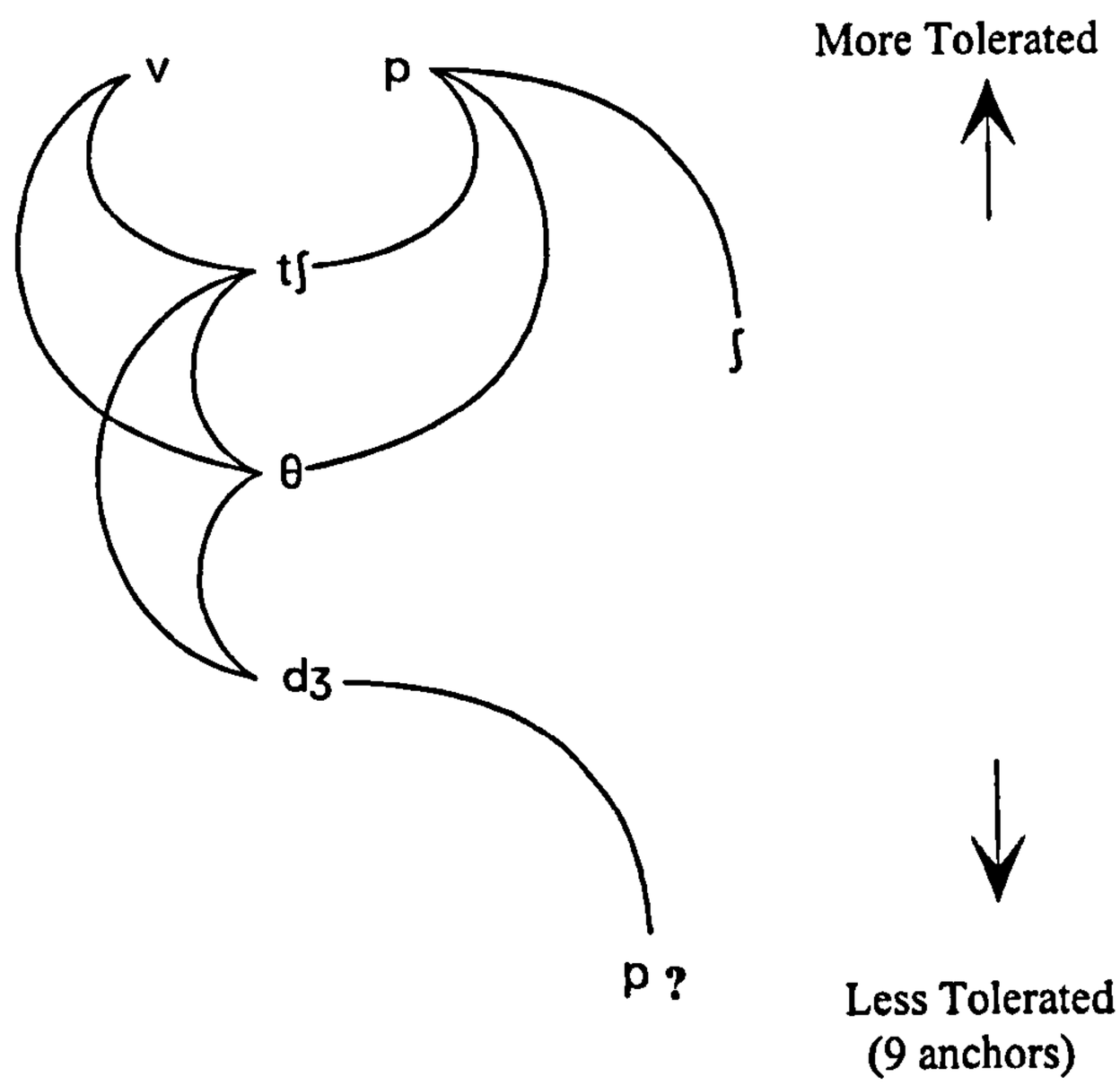
Contrast	%	Tape
/p - θ/	97.9 ***	T1
	88.9 ***	T2
/tʃ - dʒ/	95.8 ***	T1
	75.0 ***	T1
/p - ʃ/	88.9 ***	T2
	81.5 ***	T2
/v - θ/	87.5 ***	T1
/v - tʃ/	83.3 ***	T1
	64.6 *	T1
/v - ʃ/	74.1 ***	T2
/θ - dʒ/	74.1 ***	T2
/tʃ - p/	72.9 ***	T1
/tʃ - ʃ/	70.4 **	T2
/ʃ - dʒ/	68.5 **	T2



Notes: For T1, $N = 48$; for T2, $N = 54$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.38 Younger Finnish-speaking informants (<=22) listening to *male voices*

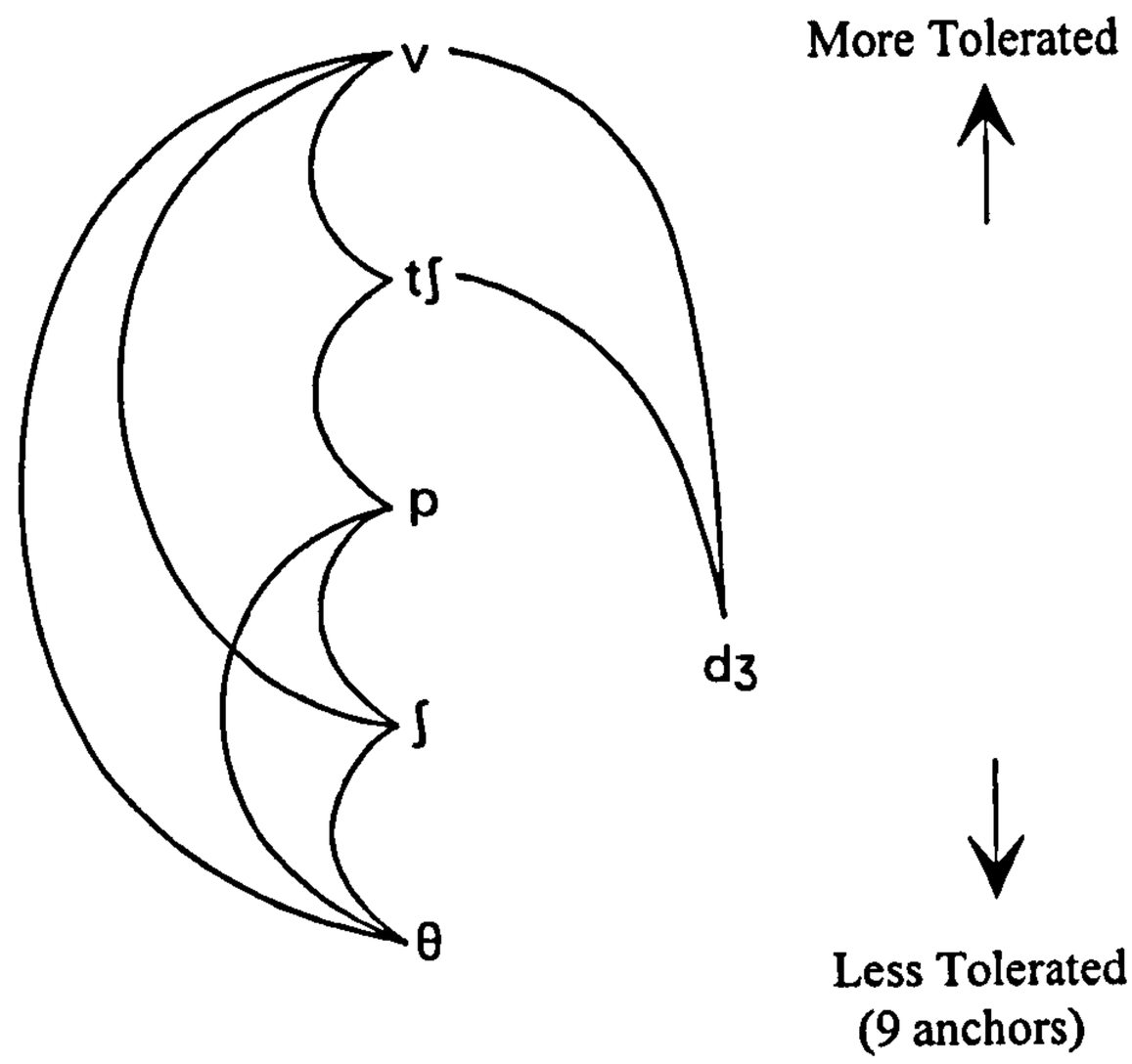
Contrast	%	Tape
/p - tʃ/	87.0 ***	T2
/p - θ/	83.3 ***	T1
	74.1 ***	T2
/θ - dʒ/	79.2 ***	T1
	75.9 ***	T2
/v - θ/	75.9 ***	T2
/tʃ - θ/	83.3 ***	T1
	64.8 *	T2
/v - tʃ/	68.5 **	T2
/dʒ - p/	68.5 **	T2
/tʃ - dʒ/	66.7 *	T2
/p - ʃ/	66.7 *	T1



Notes: For T1, $N = 48$; for T2, $N = 54$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.
One self-contradictory result: /ʃ -v/ 66.7 * T1
/v - ʃ/ 64.6 * T1

Table 7.39 Older Finnish-speaking informants (>=23) listening to *female voices*

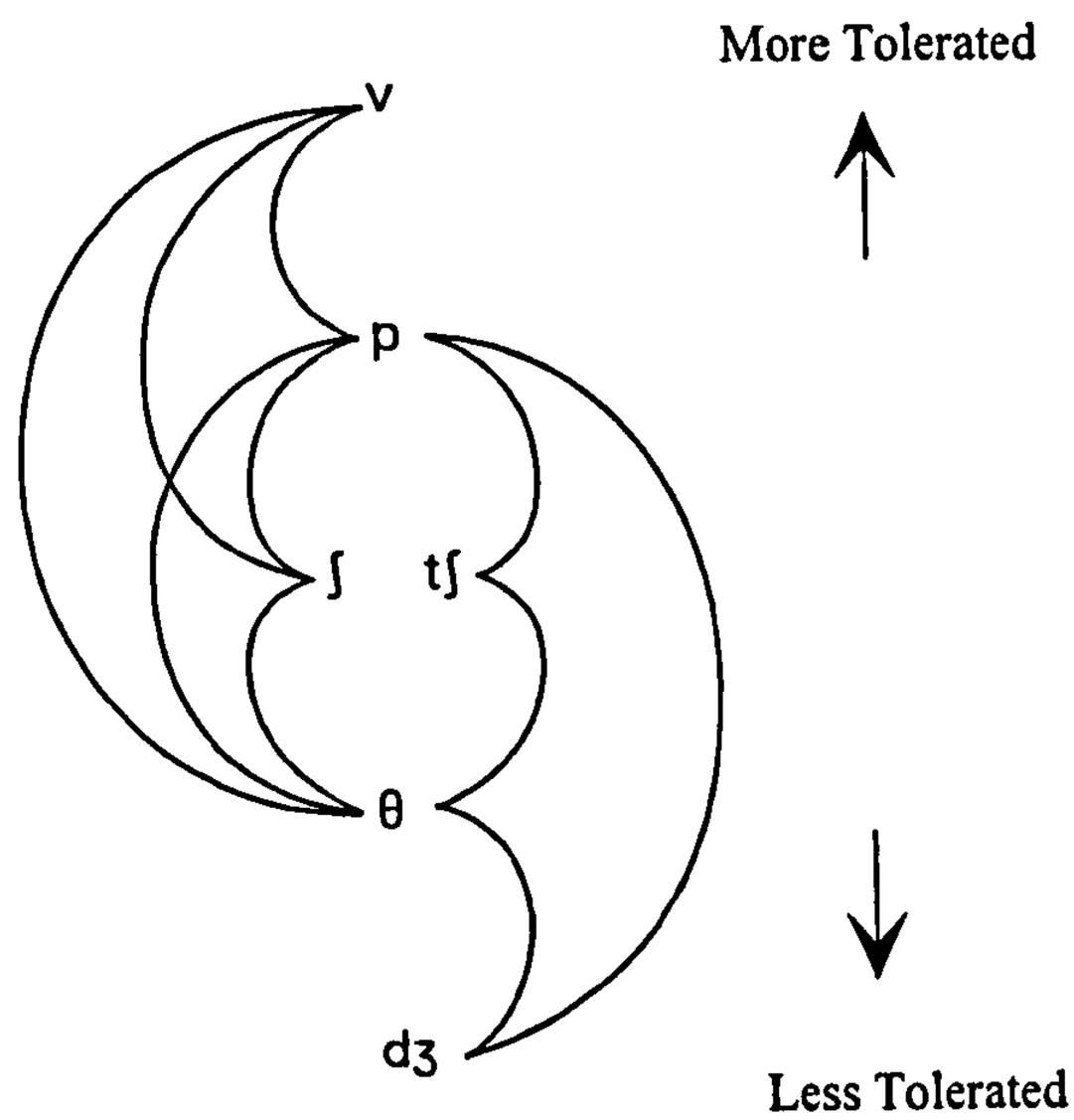
Contrast	%	Tape
/tʃ - dʒ/	92.2 ***	T1
	72.5 **	T1
/v - θ/	90.2 ***	T1
/p - θ/	83.7 ***	T2
	80.4 ***	T1
/v - tʃ/	82.4 ***	T1
	68.6 **	T1
/p - ʃ/	76.7 ***	T2
	67.4 *	T2
/tʃ - p/	74.5 ***	T1
/v - ʃ/	72.1 **	T2
/ʃ - θ/	69.8 **	T2
/v- dʒ/	64.7 *	T1



Notes: For T1, $N = 51$; for T2, $N = 43$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 7.40 Older Finnish-speaking informants (>=23) listening to *male voices*

Contrast	%	Tape
/θ - dʒ/	82.4 ***	T1
	65.1 *	T2
/p - tʃ/	79.1 ***	T2
/p - dʒ/	72.1 **	T2
/v - θ/	72.1 **	T2
/v - ʃ/	70.6 **	T1
/p - ʃ/	70.6 **	T1
/tʃ - θ/	70.6 **	T1
	67.4 *	T2
/ʃ - θ/	68.6 **	T1
/p - θ/	68.6 **	T1
	67.4 **	T2
/v - p/	64.7 *	T1



Notes: For T1, $N = 51$; for T2, $N = 43$.
Chi Square analysis: *** $p < .001$; ** $p < .01$; * $p < .05$.

uncertainty in evaluation, as was already evident from the situation for all Finnish-speaking listeners listening to female voices (Table 7.5). However, it would appear to be the older listeners who are less categorical in their preferences, who are more tolerant.

The data for the *younger* Finnish-speaking informants listening to *male* voices (Table 7.38) is problematic: / p / is preferred to / tʃ /, and / tʃ / is preferred to / dʒ /, but then / dʒ / is puzzlingly preferred to / p /. Remember that Finnish-speaking men had similar difficulties with / p / (Table 7.24), though their hierarchy was more unstable than this one here. (Cf. the puzzling situation of younger English-speaking informants listening to female voices: Table 7.31, P. 166). In addition mispronounced / ʃ / has only one anchor which attaches it below mispronounced / p /.

The data for the *older* Finnish-speaking informants listening to *male* voices yields a hierarchy (Table 7.40) which is noticeably more stable than that of the younger Finnish-speaking listeners listening to *male* voices (Table 7.38): all the mispronunciations have at least two anchors locking them into the pattern.

Overall the younger Finnish-speaking listeners listening to *male* voices (Table 7.38) would seem to be unsure and/or tolerant. Thus, curiously, while the younger English-speaking listeners (Table 7.31)—and to a lesser degree perhaps the male English-speaking listeners (Table 7.19)—are unsure and confused in their reactions to *female* voices, here the younger Finnish-speaking listeners are confused (and confusing) in their reactions to *male* voices. Notice too that the young Finnish-speaking listeners are particularly indecisive about the relationships between / v, p / and / ʃ / in *male* voices, but the young Anglophones are more indecisive about the relationships between / θ, tʃ / and / p / in *female* voices.

8. RESULTS AND ANALYSIS – EXPERIMENT 2 : SPEAKER EVALUATION

'We call it Horse,' said Mr. Podsnap, with forbearance. 'In England, Angleterre, England, We Aspirate the "H," and We Say "Horse." Only our Lower Classes Say "Orse!"'

'Pardon,' said the foreign gentleman. 'I am alwiz wrong!'

—Dickens, *Our Mutual Friend*

The results and analyses of the second experiment, speaker evaluation, are presented in three sections according to the three main questions asked: estimated age; estimated excellence of pronunciation; and perception of the fourteen traits.

8.1. Estimated age of speakers

The first question asked about each of the five speakers concerned age (Figure 8.1). Notice that the measuring tool presented to informants here is visibly nominal/ordinal and not interval.

How old do you think the speaker is?

☐ 14–15 ☐ 16–17 ☐ 18–19 ☐ 20–21 ☐ 22–23 ☐ 24–25 ☐ 26–27

☐ 28–29 ☐ 30–31 ☐ 32–33 ☐ 34–35 ☐ 36–37 ☐ 38–39 ☐ 40–41

Figure 8.1

My intention with this first question had been to help informants to start creating a mental picture of the speaker, a gentle introduction to the task of trying to visualise the speaker before tackling the bi-polar descriptive adjectives. I had naively assumed that the determination of age from a voice would be easy. After all, it is widely believed that the estimation of age for speakers of one's own language is easy (e.g. 'A person . . . who cannot tell an old from a young speaker is not a competent user of her language.' Cameron 1985:24). Why should it be any more difficult for a compatriot speaking a foreign language, even speaking a language one does not understand? However, not only was the wide range of answers given here startling, but so also was the discovery that the Finnish-speaking informants judged four of the five speakers to be significantly older when they spoke English with *less* of a Finnish accent and significantly younger when they spoke with a *stronger* Finnish accent.

First, a recap of the details of the speakers and tapes as presented in section 6.1.2.2 is given in Table 8.1 and Table 8.2.

Table 8.1 The speakers and the mispronunciation versions

Speaker			Tape 1		Tape 2	
No.	Gender	Age	Version & label	Phonemic mistakes	Version & label	Phonemic mistakes
1	Woman	23.00	1 – Excellent	0	3 – Fair	31
2	Man	26.1	3 – Fair	31	1 – Excellent	0
3	Man	26.5	2 – Good	11	5 – Poor	47
4	Woman	20.1	5 – Poor	47	2 – Good	11
5	Woman	20.75	1 – Excellent	0	4 – Fair	33

Table 8.2 Age of the speakers

	Age	
	Mean	sd
Overall	23.3	3.0
Women	21.3	1.5
Men	26.2	0.3

Speaker 1

When Finnish-speaking informants listened to Speaker 1, a 23-year-old woman, the estimation of age was mildly affected by the pronunciation (see Table 8.3 and Figure 8.2, P. 176. N.B. The mean shown in this and subsequent tables is mathematically an unsatisfactory measure for this question with its two-year answer boxes, but it is 'visually' useful and will be used in the graphic presentation of Figure 8.12, P. 187. See section 6.1.2.3, P.118).

The mode for both Tapes is 24-25 years of age. However, for the Excellent Pronunciation, the distribution leans towards the older end of the spectrum such that the median is 26-27, whereas for the Fair Pronunciation, the median remains at 24-25. In other words, in addition to the fact that the mode for both versions slightly overshoots the real age of exactly 23 years, the better pronunciation would seem to make the speaker seem slightly older than she really is, whereas the worse pronunciation would seem to have relatively little effect on the listeners' perception of her real age. The Mann-Whitney *U* test and the Median test for two independent groups reveal that this difference in the evaluations of the two tapes is significant ($p < .001$).[1]

[1] A more powerful statistical test to use for this question concerning age might be the Kolmogorov-Smirnov test for two independent groups. It does indeed confirm the significance values of the Mann-Whitney *U* test for the data concerning the native *Finnish* listeners' estimation of the age of all five speakers. However, the number of *English*-speaking informants relative to the 14 boxes for age on the scale is regularly too small, with the result that the SPSS program gives a warning about it. Consequently I shall refrain from using the Kolmogorov-Smirnov test and use only the Mann-Whitney *U* test and the Median test for both the Finnish-speaking and English-speaking informants. (See section 2.2.4.2.)

While this finding may not be so surprising if—as seems likely—Finnish-speaking informants consider proficiency to be the result of age and experience, it is nevertheless a finding which it may be advisable *not* to advertise in the classroom: young students in their early twenties may not wish to appear to be in their late twenties or, worse still, in their thirties. Sounding youthful may be more important to some students than sounding almost like a native speaker.

Table 8.3 Speaker 1: 23-year-old woman

Tape		**Mode**	**Median**	**Mean**
1: 0 mistakes	–'Excellent'	24-25	26-27	27.2
2: 31 mistakes	– 'Fair'	24-25	24-25	24.1

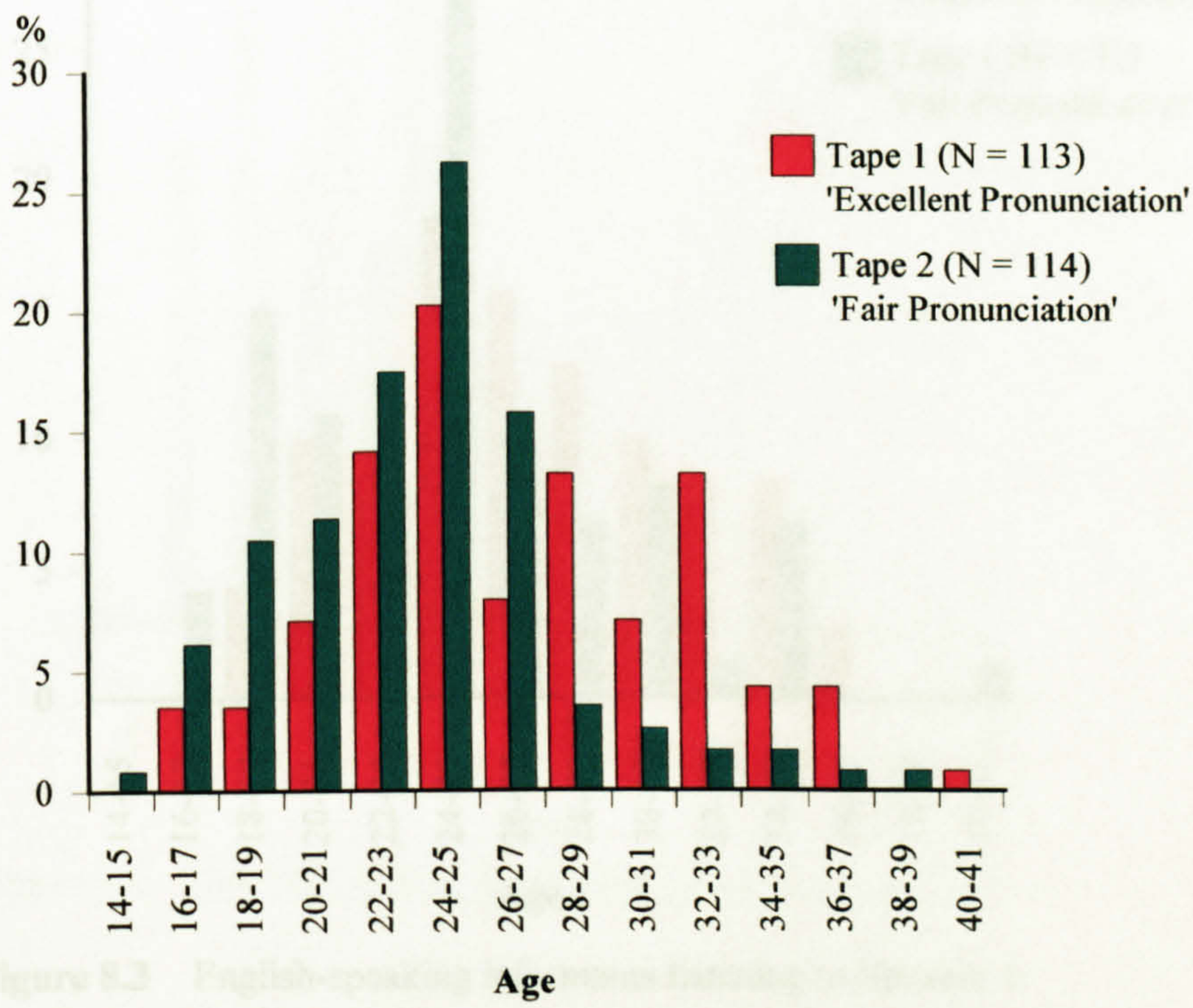


Figure 8.2 Finnish-speaking informants listening to Speaker 1

The other astounding result here is the range of answers given: for the Excellent Pronunciation from 16-17 to 40-41, for the Fair Pronunciation from 14-15 to 38-39. Many native Finnish listeners are obviously not able to pinpoint the age of a Finn speaking English with very much accuracy, though whether they are any more accurate when a Finn speaks Finnish is unfortunately unanswerable within the limits of the present experiment.

The data from the English-speaking informants look much the same as those from the Finnish-speaking informants (see Table 8.4 and Figure 8.3, P. 177).

The mode for both versions (24-25) places the speaker as slightly older than she actually is. The median for the Excellent Pronunciation is 26-27, whereas the median for Tape 2, the Fair Pronunciation, is 24-25. Thus the English-speaking informants also see the speaker as slightly older than she actually is, but as even

older (slightly so) when she speaks with no phonemic mistakes. There is a significant difference between the evaluations of the two tapes: both Mann-Whitney *U* test and Median test, $p < .01$.

Table 8.4 Speaker 1: 23-year-old woman

Tape		**Mode**	**Median**	**Mean**
1: 0 mistakes	– 'Excellent'	24-25	26-27	27.5
2: 31 mistakes	– 'Fair	24-25	24-25	25.1

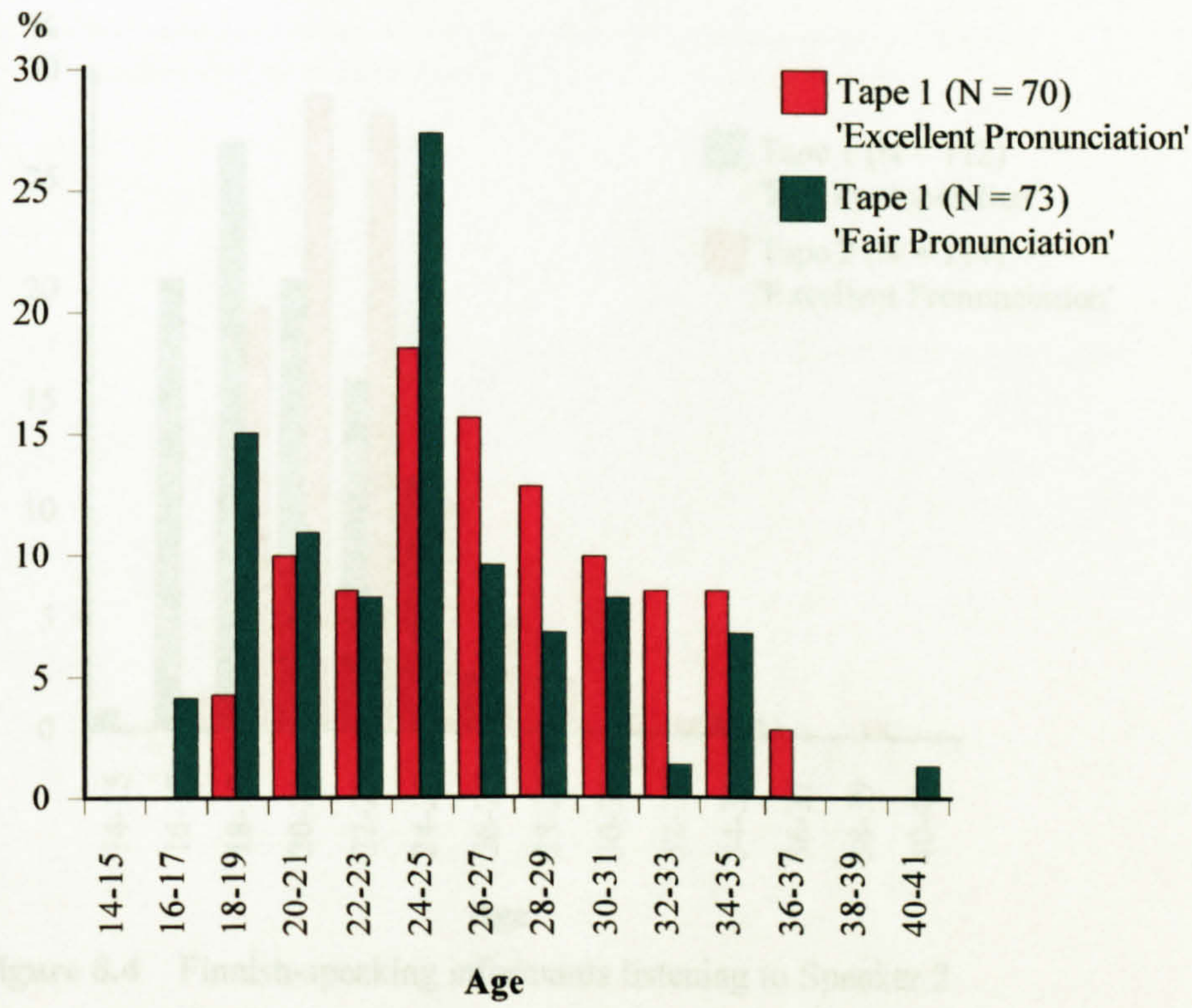


Figure 8.3 English-speaking informants listening to Speaker 1

The range of ages proposed for this speaker is again surprisingly wide: native English listeners also have difficulties guessing the age of this foreign woman speaking English, though the range is not quite as wide as that found with the native Finnish listeners.

The Mann-Whitney *U* test and the Median test also show that there is no significant difference in the evaluations of the speaker between the Finnish-speaking informants and the English-speaking informants. Both informant groups see the speaker as being equally young/old for the Excellent Pronunciation, and equally young/old for the Fair Pronunciation: $p > .05$ for both versions. Contrast Figure 8.2 and Figure 8.3.

Speaker 2

When Finnish-speaking informants listened to Speaker 2, a 26.1-year-old man, the estimation of age was noticeably inaccurate with a striking under-estimation of

the real age. In addition, once again, the assessment was affected by the degree of phonemic accuracy (see Table 8.5 and Figure 8.4).

Table 8.5 Speaker 2: 26.1-year-old man

Tape		Mode	Median	Mean
1: 31 mistakes	– 'Fair'	18-19	20-21	20.7
2: 0 mistakes	– 'Excellent'	20-21	21-22[a]	22.5

[a] The median for Tape 2 falls between the boxes '20-21' and '22-23.'

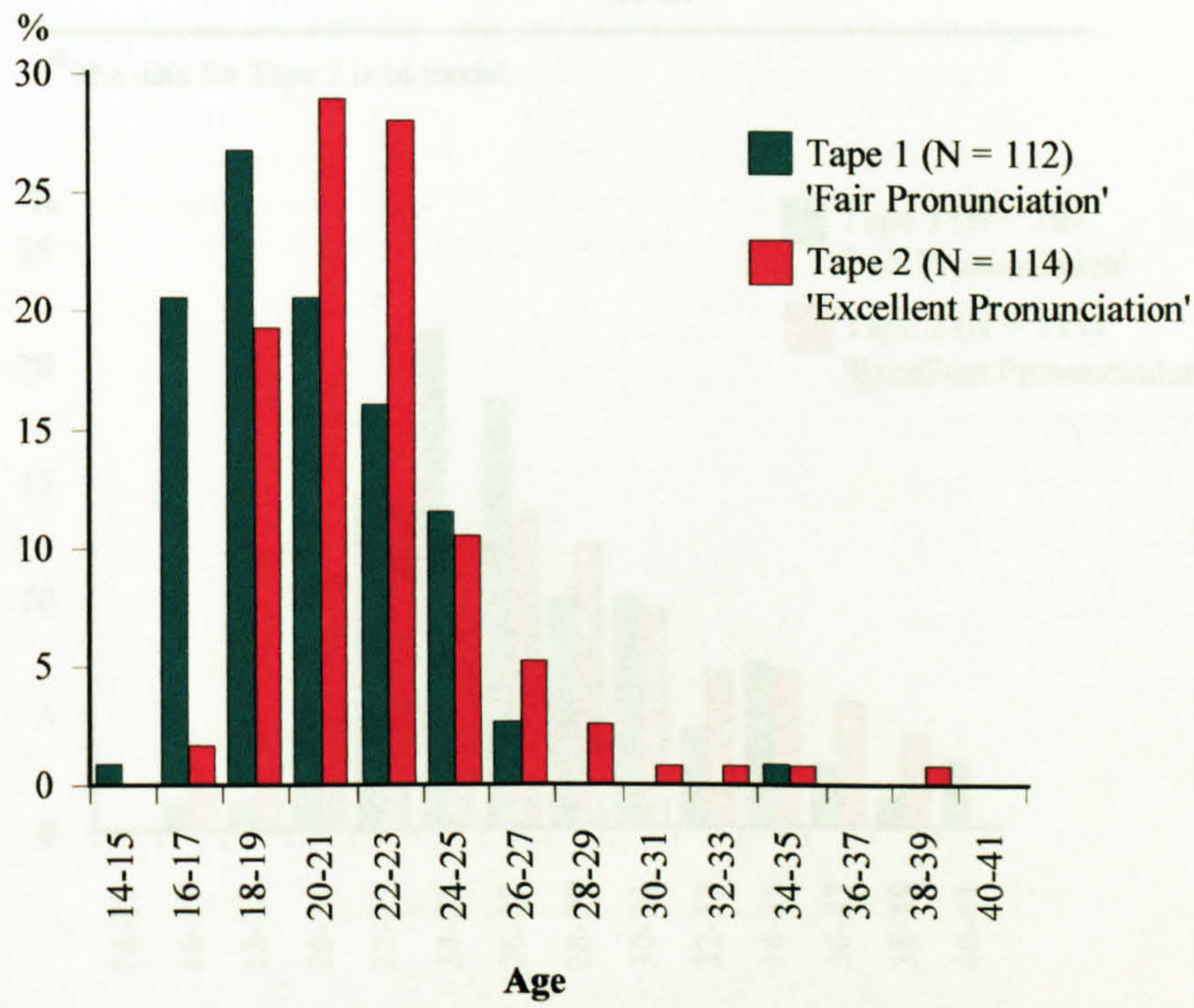


Figure 8.4 Finnish-speaking informants listening to Speaker 2

The values of both mode and median for both versions (18-19 and 20-21 respectively for the Fair Pronunciation; 20-21 and 21-22 for the Excellent Pronunciation) were clearly below the real age of 26.1 years (Table 8.5). However, both the Mann-Whitney *U* test and the Median test reveal a significant difference in the evaluations of the two tapes ($p < .001$ and $p < .01$ respectively). Thus the fact that overall the estimation of age was lower for the Fair Pronunciation (31 mispronounced phonemes) than for the Excellent Pronunciation (no phonemic mistakes) confirms the finding with Speaker 1 that for native Finnish listeners a better pronunciation would seem to make a speaker seem older and a worse pronunciation younger.

It is worth noting that with this male speaker the distribution of proposed ages falls into a much narrower band than that for Speaker 1. Is the age of a male voice easier for native Finnish listeners to estimate, and/or are some intervening variables (such as voice quality, voice pitch, pitch range) making the task easier?

When evaluating Speaker 2, the English-speaking informants are reasonably accurate—at least when judging by central tendency—in their estimation of age, regardless of which pronunciation they hear, Fair or Excellent (see Table 8.6 and Figure 8.5).

Table 8.6 Speaker 2: 26.1-year-old man

Tape		**Mode**	**Median**	**Mean**
1: 31 mistakes	– 'Fair'	24-25	26-27	27.6
2: 0 mistakes	– 'Excellent'	22-23 26-27[a]	26-27	27.8

[a] The data for Tape 2 is bi-modal.

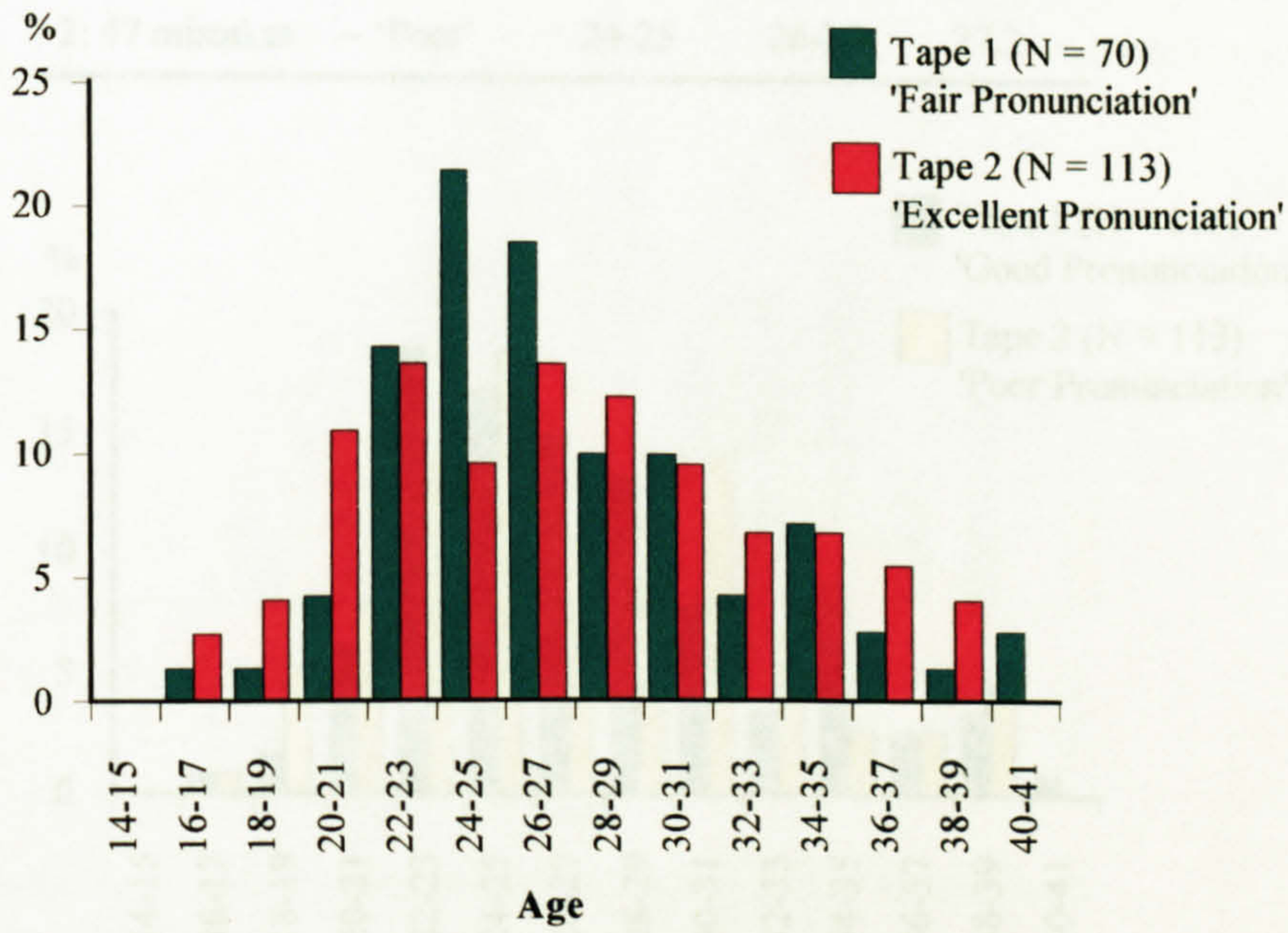


Figure 8.5 English-speaking informants listening to Speaker 2

Interestingly the estimations of the age of this male speaker are basically the same for the two versions: neither the Mann-Whitney *U* test nor the Median test show any significant difference in the evaluations of the two tapes ($p > .05$). One is tempted to conclude—very tentatively at this stage—that when there is only a medium difference in the phonemic accuracy of a speaker (perfect versus 31 mistakes), then native English listeners do *not* 'over-age' or 'juvenate' a *male* speaker in the same way as they do a *female* speaker (Speaker 1).

However, it is obvious that the native English listeners have problems estimating the age of this male speaker, as many problems as they had with Speaker 1, a woman: the range and distribution of proposed ages is incredibly wide.

For the evaluations of Tape 1 and Tape 2, it is clear that the Finnish-speaking informants and the English-speaking informants do *not* have the same reactions towards this male speaker (although they *did* for the female speaker, Speaker 1). The native English listeners clearly feel this man is older than what the Finnish-

speaking informants think he is, regardless of which pronunciation they hear, Fair or Excellent. The Mann-Whitney *U* test and the Median test show that the difference between the two informant groups for the two tapes separately is highly significant: $p < .001$ for both tapes. Contrast Figure 8.4 and Figure 8.5.

Speaker 3

The two groups of Finnish-speaking informants who listened to the separate versions of Speaker 3, a man of 26.5 years of age, were fairly similar in their estimation of age (see Table 8.7 and Figure 8.6).

Table 8.7 Speaker 3: 26.5-year-old man

Tape		**Mode**	**Median**	**Mean**
1: 11 mistakes	– 'Good'	22-23	26-27	27.6
2: 47 mistakes	– 'Poor'	24-25	26-27	27.3

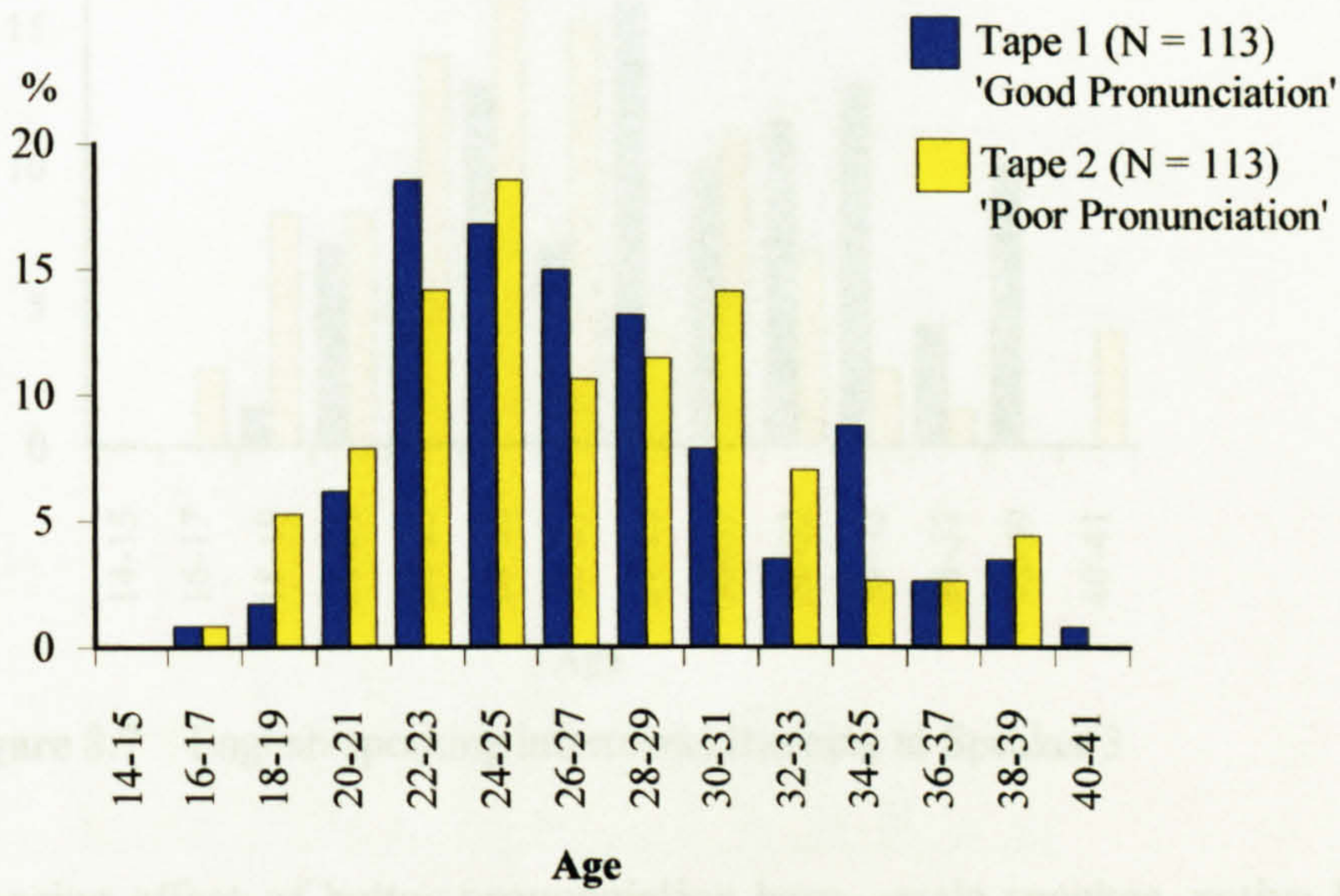


Figure 8.6. Finnish-speaking informants listening to Speaker 3

The mode values (22-23 for the Good Pronunciation; 24-25 for the Poor Pronunciation) reveal an under-estimation of the speaker's real age, though the median values (26-27 for both tapes) are spot on (Table 8.7). It is noticeable that the better of the two version yields the lower mode, but no conclusion can be drawn since the Mann-Whitney *U* test and the Median test reveal that there is no significant difference in the evaluations of the two tapes ($p > .05$). The fact that the Good Pronunciation (with 11 phonemic errors) has no 'ageing' effect here might possibly suggest that for native Finnish listeners any mispronunciation, however slight or strong, is enough to counter (or inhibit) the ageing effect witnessed with an excellent pronunciation.

It is again noticeable that the range and distribution of estimated ages is extremely wide, the spread covering a good twenty years.

For Speaker 3 the English-speaking informants view the Good Pronunciation as being produced by an older speaker than what is actually the case, and the Poor Pronunciation by someone younger (see Table 8.8 and Figure 8.7).

Table 8.8 Speaker 3: 26.5-year-old man

Tape	**Mode**	**Median**	**Mean**
1: 11 mistakes – 'Good'	28-29	28-29	30.1
2: 47 mistakes – 'Poor'	24-25	24-25	26.5

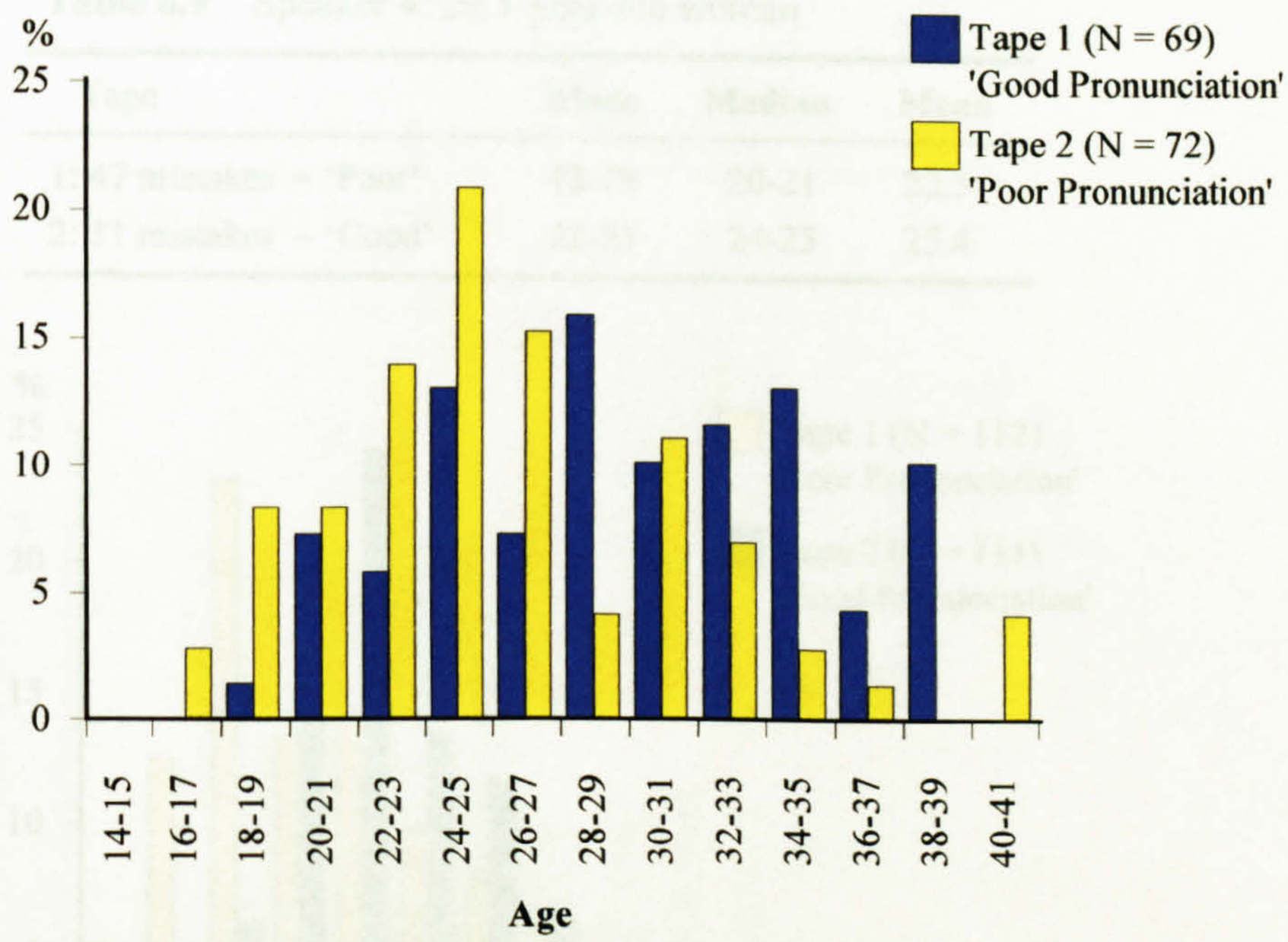


Figure 8.7 English-speaking informants listening to Speaker 3

The ageing effect of better pronunciation here—male speaker, native English listeners—is clear: there is a significant difference in the evaluations of the two tapes (Mann-Whitney *U* test and Median test, $p < .001$). Since no such effect was found with Speaker 2 (a man), though there was for Speaker 1 (a woman), one is tempted perhaps to surmise that the Poor Pronunciation here (47 phonemic mistakes) is so bad (in the ears of native English listeners) that the speaker is necessarily viewed as being younger (i.e. 'juvenated') relative to the Good Pronunciation (11 mistakes). Similarly the Good Pronunciation is so good that the speaker is necessarily viewed as being older: 'No older—and necessarily more experienced—foreigner could possibly speak so badly,' v. 'No young foreigner could possibly speak so well.'

The range and distribution of estimated ages is again extremely wide.

The Mann-Whitney *U* test and the Median test reveal that the Finnish-speaking and English-speaking listeners estimate the age of the Good Pronunciation in different ways ($p < .01$), and that of the Poor Pronunciation in the same way ($p > .05$). Contrast Figure 8.6 and Figure 8.7. In other words, when the phonemic

accuracy is high, the native English listeners think the speaker is older than what the native Finnish listeners think, but when the pronunciation is bad (i.e. lots of phonemic inaccuracies), both native English listeners and native Finnish listeners estimate the age in the same way.

Speaker 4

Finnish-speaking informants listening to Speaker 4, a woman aged 20.1 years, would again seem to downgrade age with Poor Pronunciation (47 mispronounced phonemes) and upgrade age with Good Pronunciation (11 mispronounced phonemes). See Table 8.9 and Figure 8.8.

Table 8.9 Speaker 4: 20.1-year-old woman

Tape	**Mode**	**Median**	**Mean**
1: 47 mistakes – 'Poor'	18-19	20-21	22.3
2: 11 mistakes – 'Good'	22-23	24-25	25.4

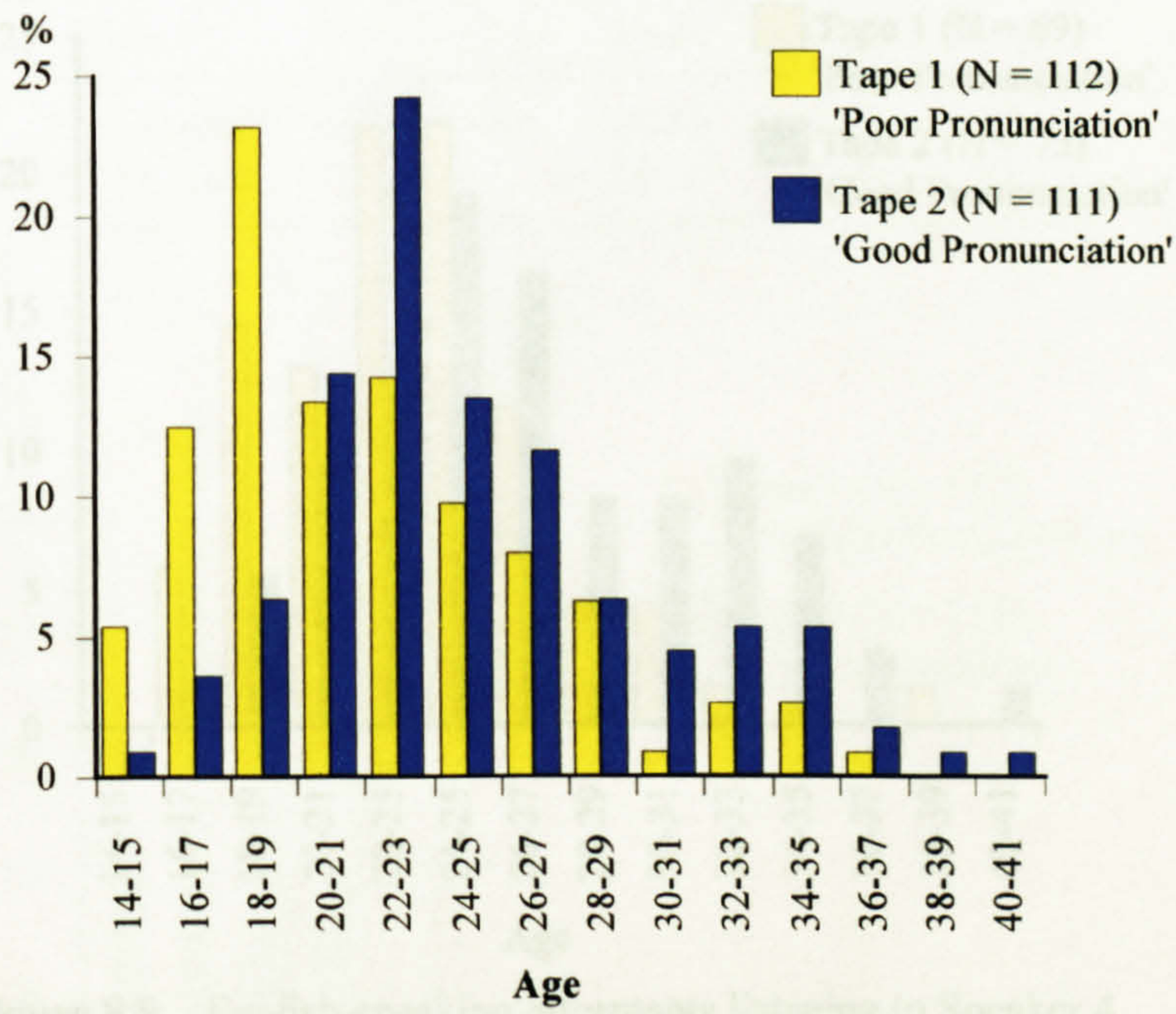


Figure 8.8 Finnish-speaking informants listening to Speaker 4

The mode for both versions fails to capture the real age of the speaker, that for the Poor Pronunciation being 18-19 (slight under-estimation), that for the Good Pronunciation being 22-23 (over-estimation). See Table 8.9. The median for the Poor Pronunciation is appropriate (20-21), that for the Good Pronunciation not (24-25). As with Speaker 1 (also a woman), the Finnish-speaking listeners would seem to interpret better pronunciation from a woman as indicating that the speaker is older than what she actually is. Both the Mann-Whitney *U* test and the Median test show that there is a significant difference in the evaluations of the two tapes ($p < .001$ and $p < .01$ respectively).

The startling finding here is once again the wide range of ages ascribed to the speaker. With the Poor Pronunciation 5.3% of informants thought the speaker was 14-15 years of age, while 6.3% thought her to be 32 or over (32-37). The Good Pronunciation has an even wider range from 14-15 through to 40-41.

The two groups of English-speaking informants estimate the age of this fourth speaker as being slightly older than what she actually is (Table 8.10 and Figure 8.9).

Table 8.10 Speaker 4: 20.1-year-old woman

Tape	**Mode**[a]	**Median**	**Mean**
1: 47 mistakes – 'Poor'	22-23 24-25	22-23	23.7
2: 11 mistakes – 'Good'	24-25	26-27	27.2

[a] The data for Tape 1 is bi-modal.

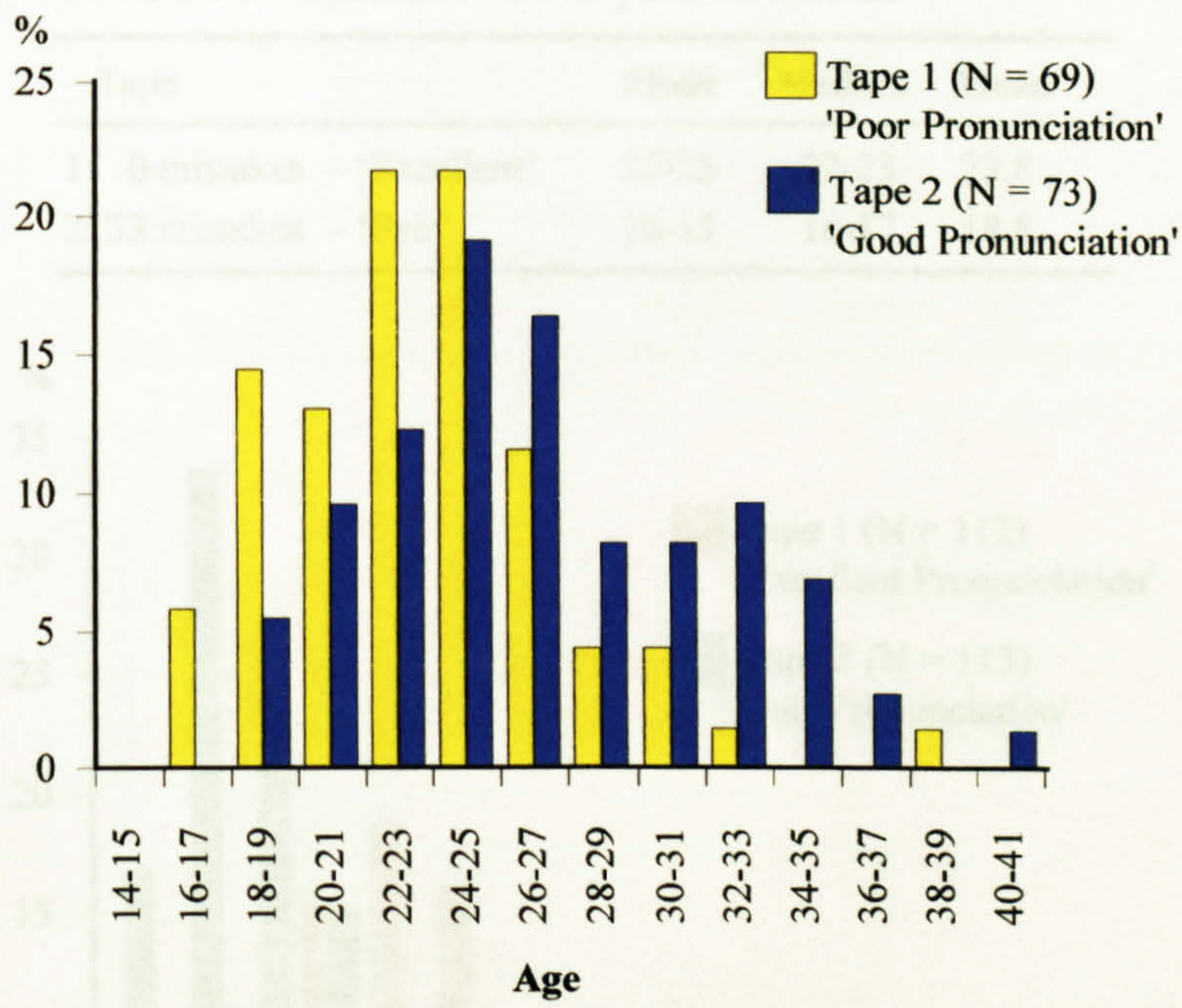


Figure 8.9 English-speaking informants listening to Speaker 4

However, what is interesting is that once again (similar to the results with Speaker 1) the version with fewer phonemic mistakes bestows an older aura than the version with more phonemic mistakes. Both the Mann-Whitney *U* test and the Median test confirm that there is a significant difference in the evaluations of the two tapes ($p < .001$). Note that the version with fewer mistakes is not 'perfect' (there are 11 mistakes) but it is 'near enough' to have an ageing effect: better pronunciation presumably suggests experience, maturity and general linguistic knowledge, though only a qualitative study would clarify this point. Greater phonemic accuracy certainly suggests 'years,' i.e. an older speaker. By contrast, worse pronunciation (here four times worse in terms of phonemic mistakes per

same text) apparently prompts the impression of inexperience, immaturity, lack of linguistic knowledge . . . and fewer 'years,' i.e. a younger speaker.

The range and distribution of estimated ages is wide but would seem to be narrower for the Poor Pronunciation than for the Good Pronunciation.

The difference between the Finnish-speaking listeners and the English-speaking listeners in the estimation of age of both versions of this fourth speaker is mildly significant, the Mann-Whitney U test yielding $p < .05$ for both versions, the Median test showing no difference for the Poor Pronunciation ($p > .05$) and a moderate difference for the Good Pronunciation ($p < .05$). Contrast Figure 8.8 and Figure 8.9.

Speaker 5

Finnish-speaking informants listening to Speaker 5, a woman aged 20.75 years, yet again downgraded age for Fair Pronunciation and upgraded age for Excellent Pronunciation (Table 8.11 and Figure 8.10).

Table 8.11 Speaker 5: 20.75-year-old woman

Tape	**Mode**	**Median**	**Mean**
1: 0 mistakes – 'Excellent'	22-23	22-23	23.8
2: 33 mistakes – 'Fair'	16-17	16-17	18.8

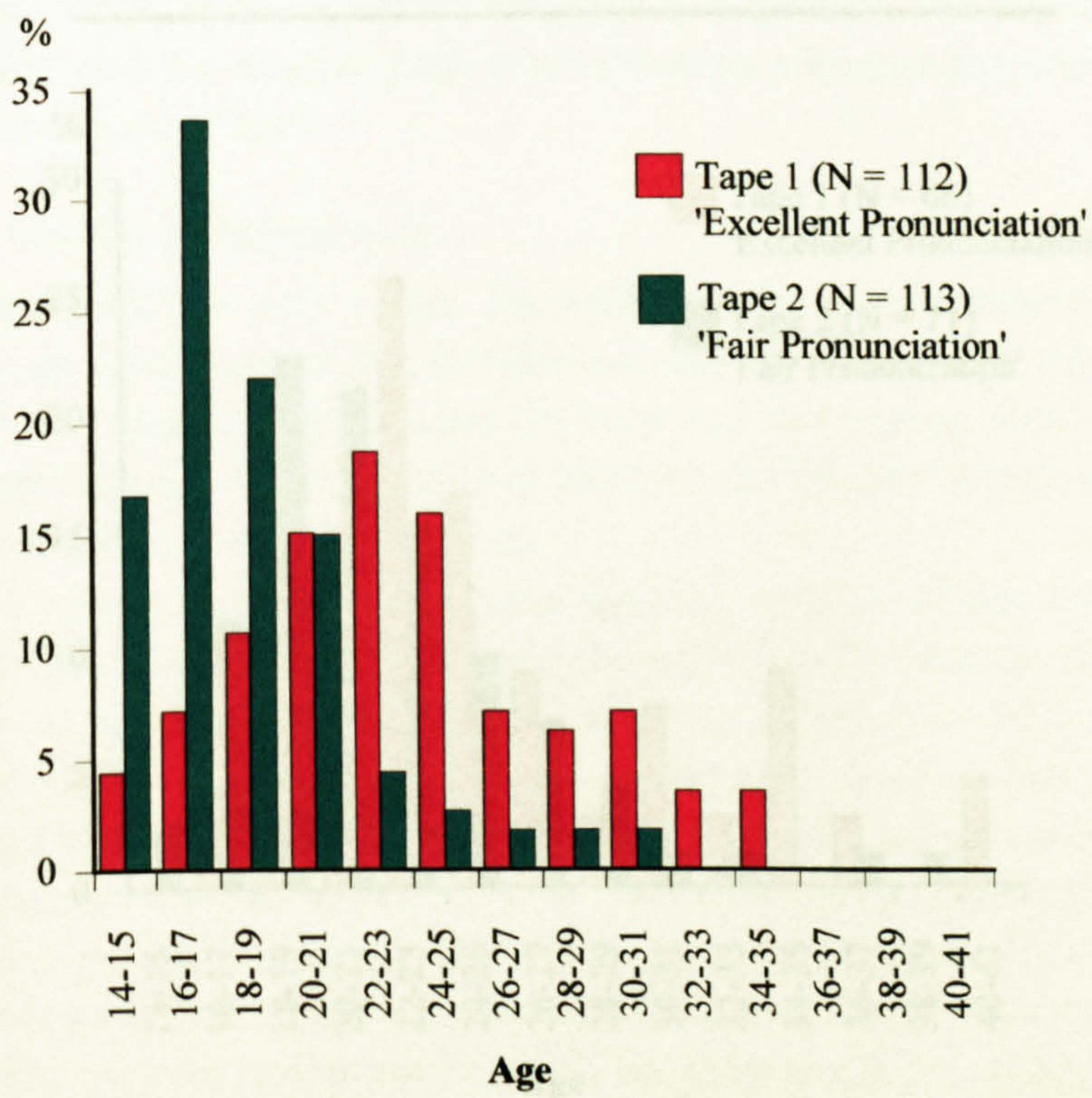


Figure 8.10 Finnish-speaking informants listening to Speaker 5

Both the Mann-Whitney U test and the Median test show that there is a significant difference in the evaluations of the two tapes ($p < .001$).

As with speaker 4 (also a woman), the mode for both versions fails to pinpoint the real age of Speaker 5: the mode for the Excellent Pronunciation over-estimates the age, while the mode for the Fair Pronunciation under-estimates the age—and under-estimates it quite badly. For native Finnish listeners here, bad pronunciation clearly prevents years from accumulating in the impression of the age of this young speaker.

The range and distribution of proposed ages is particularly narrow for the Fair Pronunciation, and hence recalls the situation found with Speaker 2 (Figure 8.4, P. 178). The suggestion presented then that gender of speaker may be a factor influencing the range and distribution hence seems to be refuted by the evidence offered by this woman speaker, though the range for Speaker 5's Excellent Pronunciation is of course wide, spanning 20 years.

The English-speaking informants over-estimate the age of Speaker 5 with her Excellent Pronunciation, but are more or less accurate in their estimation of her age with her Fair Pronunciation (see Table 8.12 and Figure 8.11. Cf. Table 8.11 and Figure 8.10).

Table 8.12 Speaker 5: 20.75-year-old woman

Tape	**Mode**	**Median**	**Mean**
1: 0 mistakes – 'Excellent'	22-23	24-25	26.5
2: 33 mistakes – 'Fair'	18-19	20-21	22.2

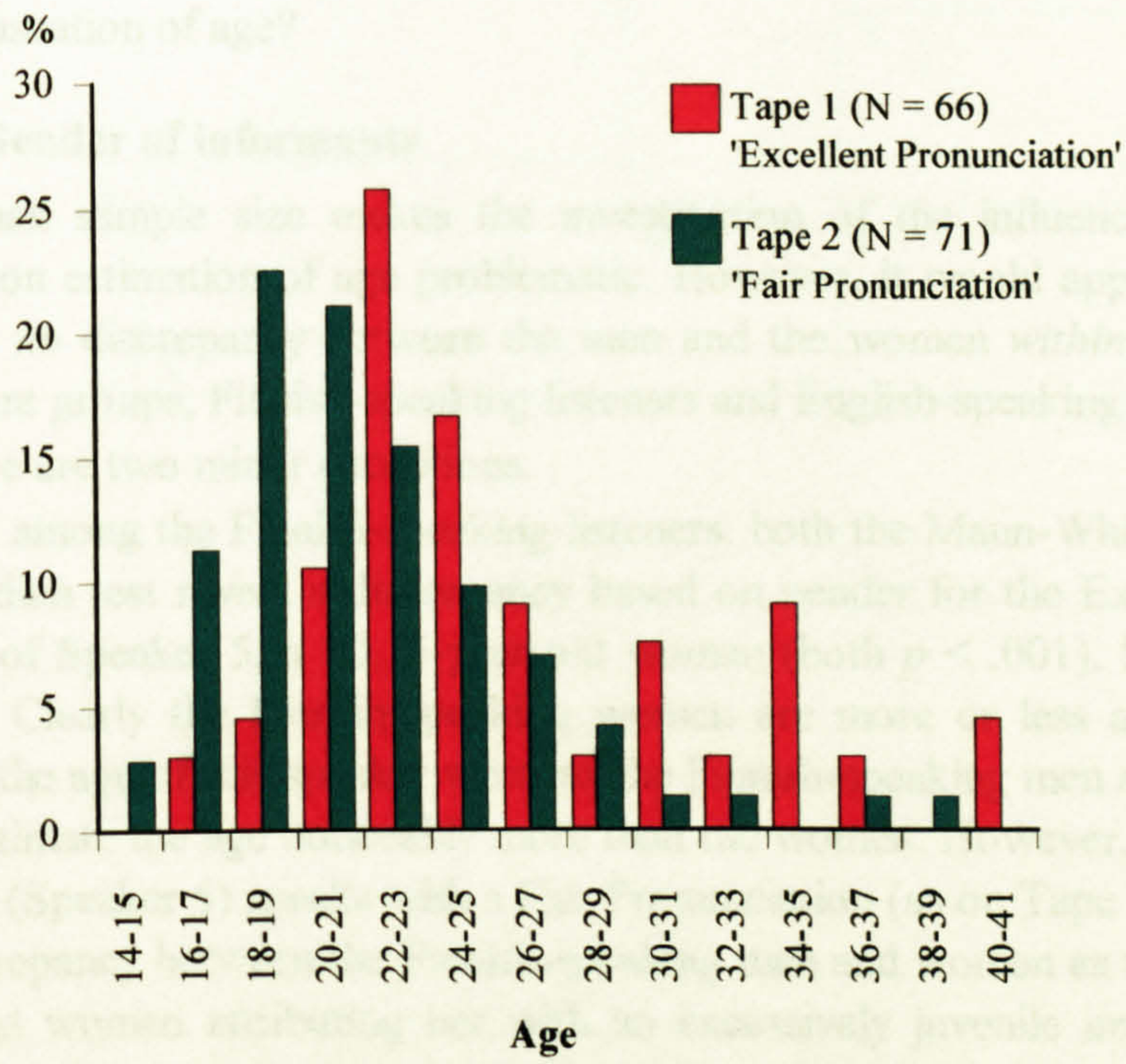


Figure 8.11 English-speaking informants listening to Speaker 5

Both the Mann-Whitney *U* test and the Median test reveal that there is a significant difference in the evaluations of the two tapes ($p < .001$).

The group of informants who heard the Fair Pronunciation, estimate the age reasonably accurately, the mode revealing slight under-estimation, the median being spot-on. However, as with Speaker 4 (also a woman), the English-speaking informants who heard the Excellent Pronunciation are prompted to over-estimate the age: better pronunciation can again be interpreted as bestowing more years to the impression of age.

The Finnish-speaking listeners and the English-speaking listeners clearly differ in their evaluations of the two versions of this speaker. The Mann-Whitney *U* test shows that the level of significance for the Fair Pronunciation is $p < .001$, that for the Excellent Pronunciation $p < .01$; the Median test confirms the former, but not the latter ($p > .05$). The Finnish-speaking listeners are noticeably less accurate in their estimation of the age of the Fair Pronunciation than the English-speaking listeners, and seriously 'under-age' the speaker.

The startling discrepancies between the evaluations within each language group of the different pronunciation versions of the same speaker, and between the evaluations made by Finnish-speaking listeners and by English-speaking listeners, are perhaps appreciated more from a graphic presentation of all the evaluations together. This is achieved most simply by recourse to the means of the responses (Figure 8.12, P. 187).[2] Whether the two groups of listeners would be more accurate in their estimations of age if the speakers spoke in their mother-tongue, Finnish, rather than in a foreign language is a question that regrettably cannot be answered with the present data. Does speaking a foreign language always lead to the obfuscation of age?

8.1.1. Gender of informants

The small sample size makes the investigation of the influence of informant gender on estimation of age problematic. However, it would appear that overall there is *no* discrepancy between the men and the women *within* the two major informant groups, Finnish-speaking listeners and English-speaking listeners.

There are two minor exceptions.

First among the Finnish-speaking listeners, both the Mann-Whitney *U* test and the Median test reveal a discrepancy based on gender for the Excellent Pronunciation of Speaker 5, a 20.75-year-old woman (both $p < .001$). See Table 8.13, P. 188. Clearly the Finnish-speaking women are more or less accurate in estimating the age of this speaker whereas the Finnish-speaking men are not: the men over-estimate the age noticeably more than the women. However, when the same woman (Speaker 5) speaks with a Fair Pronunciation (as on Tape 2), then there is no discrepancy between the Finnish-speaking men and women as to her age, both men and women attributing her with an excessively juvenile image, viz. 15-16 years of age (women's mode is 16-17 and median 18-19; men's mode and median

[2] The figure is modelled on that of Arthur et al. (1974:260) which diagrammatically presents the mean ranking of the dialect range of four pairs of voice guises as perceived on 10-point scales.

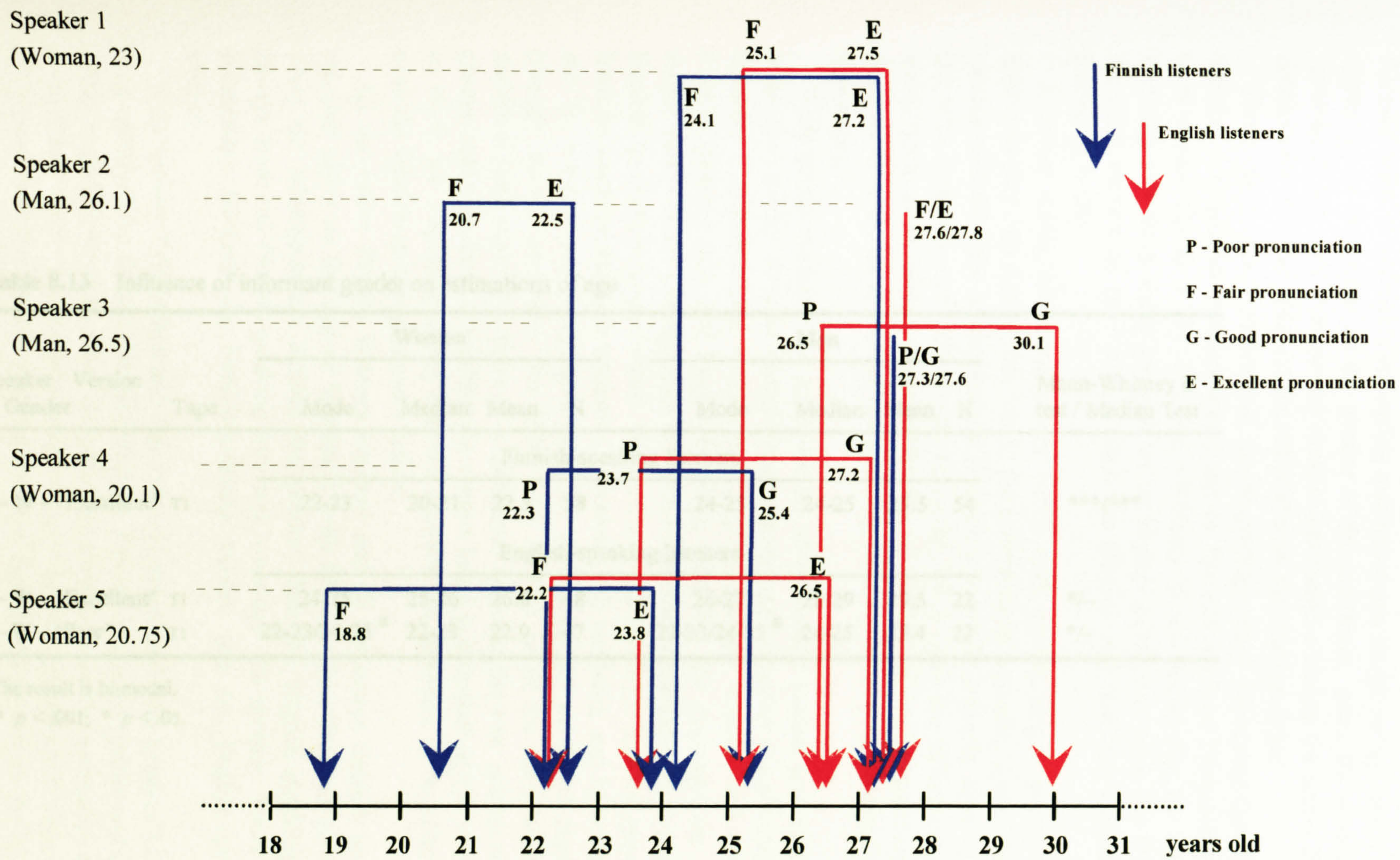


Figure 8.12 Means of estimated ages of speakers

Table 8.13 Influence of informant gender on estimations of age

Speaker Version Gender	Tape	Women Mode	Women Median	Women Mean	Women N	Men Mode	Men Median	Men Mean	Men N	Mann-Whitney *U* test / Median Test
		Finnish-speaking listeners								
5 – W – 'Excellent'	T1	22-23	20-21	22.3	58	24-25	24-25	25.5	54	***/***
		English-speaking listeners								
1 – W – 'Excellent'	T1	24-25	25-26	26.6	48	26-27	28-29	29.5	22	*/–
4 – W – 'Poor'	T1	22-23/24-25 [a]	22-23	22.9	47	22-23/24-25 [a]	24-25	25.4	22	*/–

[a] The result is bi-modal.

*** $p < .001$; * $p < .05$.

are 16-17; see also Table 8.11). This speaker, therefore, with her Excellent Pronunciation, is unusual in prompting a dichotomous image for age which would appear to be dependent on the gender of the informant, exact for women, but 'over-aged' for men: there are no similar discrepancies among the native Finnish listeners for the other speakers for either version that they produce.

The second exception involves the English-speaking informants listening to Speaker 1 (Excellent Pronunciation) and to Speaker 4 (Poor Pronunciation)—both women (Table 8.13). The Mann-Whitney *U* test shows that the English-speaking men think these speakers to be slightly older than what the English-speaking women do ($p < .05$), though the Median test does not confirm this finding ($p > .05$).

In other words, informant gender within each mother-tongue group would appear to play only a negligible role in the estimation of a speaker's age: men and women with the same mother-tongue (Finnish or English) generally seem to share the same impression of speaker age. However, when there are differences, these would seem to be limited to the estimation of the age of *women* speakers, and particularly when these women are speaking with a *phonemically accurate* pronunciation.

While gender contrasts are negligible for *within* mother-tongue groupings, this is not the case for *between* mother-tongue groupings: a number of contrasts exist between Finnish-speaking women and English-speaking women, and between Finnish-speaking men and English-speaking men, but the contrasts are not a simple mirror of the discrepancies between Finnish-speaking listeners overall and English-speaking listeners overall. The men and women divide into ten groups to listen to ten pronunciation versions (5 speakers x 2 versions), and both the men and the women, separately, contrasted for mother-tongue on six of these ten (though not the same six). See Table 8.14, P. 190. In all the cases where the statistical tests show that there is a difference between the two mother-tongue groups, and regardless of whether the version heard is Excellent or Good or Fair or Poor, the English-speaking listeners estimate the speakers to be *older* than what the Finnish-speaking listeners believe.

All the native Finnish listeners together and all the native English listeners together agree on the age of Speaker 1 with her Excellent Pronunciation, but when the informants are divided for gender, then the English-speaking *men* diverge from the Finnish-speaking *men* and 'over-age' the speaker. There is no discrepancy between the Finnish-speaking women and the English-speaking women.

With Speaker 2 (Excellent and Fair Pronunciations) and with Speaker 3 (Good Pronunciation) the finding that all the native Finnish listeners together and all the native English listeners together disagree with each other on the estimation of age is confirmed even when the two mother-tongue listener groups are broken down for gender: both the Finnish-speaking men and women disagree with the English-speaking men and women respectively.

Table 8.14 Significant contrasts in estimations of age (influence of informant's gender)

Speaker Gender	Version	Tape	Finnish-speaking listeners Mode	Median	*N*	English-speaking listeners Mode	Median	*N*	Mann-Whitney *U* test / Median test
					Men				
1 – W –	'Excellent'	T1	22-23/24-25 [a]	24-25	54	26-27	28-29	22	*/ –
2 – M –	'Excellent'	T2	20-21	20-21	54	22-23	22-23	23	***/**
	'Fair'	T1	18-19	18-19	54	24-25/28-29/30-31 [a]	28-29	22	***/***
3 – M –	'Good'	T1	22-23	26-27	54	24-25/32-33 [a]	29-30 [b]	22	*/–
4 – W –	'Poor'	T1	18-19/20-21 [a]	20-21	54	22-23/24-25 [a]	24-25	22	**/*
5 – W –	'Fair'	T2	16-17	16-17	53	18-19/20-21 [a]	20-21	23	**/*
					Women				
2 – M –	'Excellent'	T2	22-23	22-23	60	26-27	27-28 [b]	50	***/***
	'Fair'	T1	16-17/18-19 [a]	20-21	58	24-25	26-27	48	***/***
3 – M –	'Good'	T1	26-27	26-27	59	28-29	28-29	47	*/ –
4 – W –	'Good'	T2	22-23	22-23	59	24-25	26-27	50	**/*
5 – W –	'Excellent'	T1	22-23	20-21	58	22-23	24-25	46	***/**
	'Fair'	T2	16-17	18-19	60	18-19	20-21	48	***/***

[a] The result is multi-modal.

[b] The median falls between two values.

*** $p < .001$; ** $p < .01$; * $p < .05$.

The mildly significant difference in the evaluations of the age of Speaker 4 (both the Poor and Good Pronunciations) made by the native Finnish listeners all together and by the native English listeners all together is not entirely confirmed by the mother-tongue groups when divided for gender. In the case of the Poor Pronunciation the Finnish-speaking *men* disagree with the English-speaking *men* (but the women agree), and in the case of the Good Pronunciation Finnish-speaking *women* disagree with the English-speaking *women* (but the men agree).

As for Speaker 5 the disagreement between the two mother-tongue groups found for both the Fair and the Excellent Pronunciations persists for the Fair Pronunciation when the two groups are divided for gender: the Finnish-speaking men and women and the English-speaking men and women, respectively, are significantly different in their evaluations of age. However, for the Excellent Pronunciation only the Finnish-speaking *women* and the English-speaking *women* disagree whereas the men do not.

8.1.2. Age of informants

Analyses were made to ascertain whether the age of informants had a bearing on the estimation of the age of the speakers. However, it must be remembered that the exact age of informants was not known since informant age was calculated from 'year-of-test' minus 'year-of-birth' (see section 6.2.3.). The calculated age is potentially only accurate to within ± 11 months.

The age distributions of the informants can be seen in Table 8.15.

Table 8.15 Ages of informants

	N	Mean age	Range	*sd*
Finnish-speaking listeners	227	22.99	17–51	4.28
– women	119	23.54	17–51	4.98
– men	108	22.39	19–37	3.27
English-speaking listeners	143	29.34	17–69	13.13
– women	98	28.27	17–60	11.87
– men	45	31.62	18–69	15.42

The mean age of the English-speaking listeners was slightly older than that of the Finnish-speaking listeners as a result of attempting to boost numbers by including a few students from adult evening classes. Thus 14 of the 143 English-speaking informants were not 'normal' undergraduate or postgraduate students, though they were all graduates attending university-organised evening courses: their mean age was 55.4 years, range 39–69, *sd* 10.02.

When an arbitrary division between younger and older listeners is set at the 22/23 border line, the distribution of informants is as follows in Table 8.16, P. 192. When using this artificial border line for age, it was found that there are no differences between the younger and older *Finnish*-speaking listeners: Mann-Whitney *U* test and Median test, $p > .05$.

Table 8.16 Distribution of informants by age

	<= 22 years of age *N*	>= 23 years of age *N*
Finnish-speaking listeners		
Tape 1	71	42
Tape 2	70	44
English-speaking listeners		
Tape 1	36	34
Tape 2	30	43

The case is almost identical with the *English*-speaking listeners, i.e. there is to all intents and purposes no difference between the younger and older English-speaking informants. However, there is one small exception involving the Fair Pronunciation (Tape 2) of the first speaker, Speaker 1 (a 23-year-old woman): the results show a difference at the $p < .05$ level (Mann-Whitney *U* test), though this result is not confirmed by the Median test ($p > .05$). See Table 8.17.

Table 8.17 Estimation of Speaker 1's age ('Fair Pronunciation')

	Mode	Median	Mean
Younger English-speaking listeners	24-25	23-24 [a]	23.7
Older English-speaking listeners	24-25	24-25	26.2

[a] The median falls between two values

Since this is the first evaluation in the set and is a jump into the unknown for the informants, one should not perhaps make too much of it. However, the younger native English listeners basically think Speaker 1 younger, the older listeners think her older. The contrast is perhaps more striking on reading the cumulative percentage of estimations given up to and including the 24-25 value: the value for younger listeners is 80%, that for older listeners only 55.8% (Figure 8.13, P. 193).

Analyses were also carried out on *between* language group younger and older informants and it was found that the contrasts between the younger Finnish-speaking listeners and the younger English-speaking listeners on the one hand, and between the older Finnish-speaking listeners and the older English-speaking listeners on the other hand, led to several significant differences for all Speakers except Speaker 1 (Table 8.18, P. 194).

As far as the younger informants are concerned, the between language groups differ for five of the ten voice guises (though the significance level for one of them is borderline: Speaker 5, Excellent Pronunciation). In all five cases the young English-speaking listeners think the speaker is older than what the young Finnish-speaking listeners think.

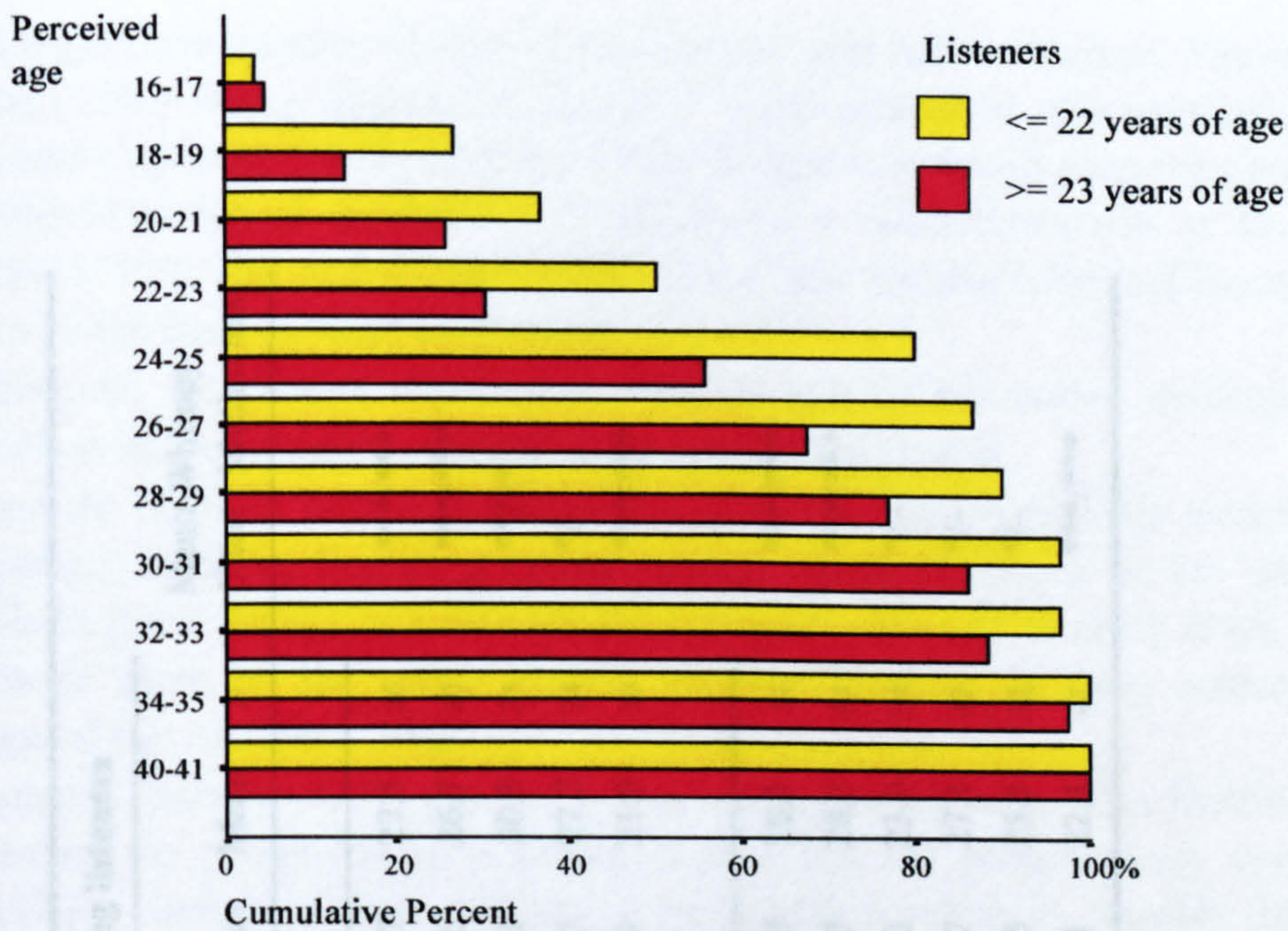


Figure 8.13 Distributions (cum%) of estimations of Speaker 1's age

As for the older informants, the between language groups differ for six of the ten voice guises, but three of these six are also borderline: Speaker 4, Poor Pronunciation and Good Pronunciation; Speaker 5, Excellent Pronunciation. Once again the old English-speaking listeners consider the speaker to be older than the old Finnish-speaking listeners do.

One may conclude that for Speaker 1 the uniformity of evaluations of age made by the Finnish-speaking listeners and the English-speaking listeners overall is repeated even when the informant groups are broken down for age: there is no hidden age factor in either language group. For Speaker 3 (a 26.5-year-old man) the dichotomous approach to the two pronunciation guises (uniformity of the Finnish-speaking listeners' and the English-speaking listeners' evaluations of the Poor Pronunciation, but non-uniformity of those of the Good Pronunciation) is here seen to rest on a differentiation of the *younger* Finnish-speaking and English-speaking listeners alone. For Speaker 4 (a 20.1-year-old woman) the mildly significant difference in evaluations made by the Finnish-speaking and English-speaking listeners overall is here shown to result from a differentiation of the *older* Finnish-speaking and English-speaking listeners. Finally, for Speakers 2 and 5, the difference in evaluations of age seen between the two language groups overall goes right across the age spectrum: the difference in evaluations is not restricted to either younger or older listeners, but is the same for both.

Table 8.18 Significant contrasts in estimations of age (influence of informant's age)

Speaker Gender	Version	Tape	Finnish-speaking listeners Mode	Median	Mean	*N*	English-speaking listeners Mode	Median	Mean	*N*	Mann-Whitney *U* test / Median test
						Younger					
2 – M –	'Fair'	T1	16-17/18-19 [a]	20-21	20.8	70	24-25	26-27	27.5	36	***/***
	'Excellent'	T2	20-21	20-21	22.2	70	22-23	24-25	26.7	30	***/***
3 – M –	'Good'	T1	24-25	26-27	27.6	71	32-33	30-31	30.8	35	**/**
5 – W –	'Excellent'	T1	24-25	22-23	24.1	71	22-23	24-25	27.2	34	*/–
	'Fair'	T2	16-17	16-17	18.5	70	18-19/20-21 [a]	20-21	21.9	30	***/***
						Older					
2 – M –	'Fair'	T1	18-19	19-20 [b]	20.5	42	22-23/26-27 [a]	26-27	28.1	34	***/***
	'Excellent'	T2	22-23	22-23	23.0	44	26-27	26-27	28.3	43	***/***
4 – W –	'Poor'	T1	18-19	22-23	22.2	41	22-23	22-23	23.9	34	*/–
	'Good'	T2	22-23	22-23	24.7	43	24-25	26-27	27.4	43	*/–
5 – W –	'Excellent'	T1	22-23	22-23	23.3	41	22-23	24-25	25.8	32	*/–
	'Fair'	T2	16-17	18-19	19.4	43	18-19	20-21	22.4	41	***/***

[a] The result is bi-modal.

[b] The median falls between two values.

*** $p < .001$; ** $p < .01$; * $p < .05$.

8.1.3. Summary

In an experiment conducted with native speakers and native listeners, Ptacek and Sander (1966) found 'impressive accuracy' in the estimation of a speaker's age. Ryan and Capadano (1978) confirmed this, though they found a centring tendency, overestimation at the bottom of the range v. underestimation at the top. Helfrich (1979) also claimed that people have little trouble assessing the age of others on the basis of voice cues alone.

However, the present experiment, with its mix of non-native speakers and native/non-native listeners, yielded totally unexpected results.

First the range of estimated ages given by native Finnish listeners listening to compatriots speaking English is extremely large: many Finns just do not seem to be able to guess the age of a Finn speaking English with any accuracy at all. Even the range given by the native English listeners listening to these unidentified foreigners was extremely wide.

Second, Helfrich (1979) notes that tempo and pausing have been identified as facilitating the perception of a speaker's age, but the present study involved recordings which kept these features as reasonably constant as possible (though not identical) in the two guises of each speaker. The present results thus would seem at least to limit the influence of these factors though not entirely exclude them.

Third, Helfrich's (1979) claim that voice quality (phonatory and articulatory settings, i.e. timbre) is decisive in age estimation needs revising—at least for foreign-accented speech—since the present experiment held voice quality constant (as much as possible) in the two guises, and yet despite this, estimations varied widely. In a real-life situation one can imagine a Finn with poor English skills being nervous and insecure when speaking English, and hence speaking in a more strained voice with possibly a higher pitch than a Finn with good English skills, confident and at ease, but this element of emotional strain and insecurity leading to a change in voice quality was not present here. In the present study it was phonemic accuracy that increased perceived age, and phonemic inaccuracy that decreased it.

Fourth, there is basically no gender differentiation *within* the language groups: Finnish-speaking men estimate age in the same way as Finnish-speaking women, as do English-speaking men and women. However, there are certain discrepancies between the *men* of the two language groups and between the *women* of the two language groups.

Fifth, there is basically no age differentiation within the language groups; the younger Finnish-speaking listeners estimate age in the same way as the older Finnish-speaking listeners, as do the younger and older English-speaking listeners. However, there were a few contrasts between the younger listeners of both language groups, and between the older listeners of each language group.

Sixth, both English-speaking listeners and Finnish-speaking listeners attribute higher age to speakers with high phonemic accuracy (except for one version of Speakers 3 and 2 respectively: Figure 8.12, P. 187) than to those with low pho-

nemic accuracy, though why this is so is impossible to explain with the present non-qualitative data. However, one might speculate that both native English listeners—whatever their experience of listening to foreigners speaking English—and native Finnish listeners assume that only experience—and hence age—leads to phonemic accuracy in a foreign language. One might also surmise that when speech is recognised as being deficient in linguistic and communicative competence, both native and non-native listeners label the speaker as 'immature,' or—to borrow Piaget's terminology for the burgeoning language of children (Piaget 1971)—'egocentric' rather than 'socialised.' Hence the age of the speaker with poor pronunciation skills is under-estimated. It is of course highly likely that native listeners listening to any foreign-accented speaker under-estimate not only biological age but also intellectual maturity and cognitive skills, a presupposition that has become the butt of many people's jokes:

> ... and there was a droll disposition, not only on the part of Mr. Podsnap, but of everybody else, to treat [the foreign gentleman] as if he were a child who was hard of hearing. (Charles Dickens. *Our Mutual Friend*)

Seventh, it is remarkable that many Finnish-speaking listeners and quite a number of English-speaking listeners crossed the box '14-15' or '16-17' for Speaker 4 (a 20.1-year-old woman) and Speaker 5 (a 20.75 year-old woman): Figures 8–11. This is perhaps something worth warning Finnish learners about in the classroom (particularly a school classroom): poor pronunciation skills may lead to 'serious' under-estimation of age. Of course how serious this under-estimation is depends on your viewpoint. However, twenty-year-olds who find themselves considered to be three years younger than they actually are will not laugh the matter off in the same way as forty-year-olds who are thought to be 37. The pedagogic significance of phonemic inaccuracy in a foreigner's speech leading to 'under-ageing' should not be dismissed lightly: teenage and young adult students will not thank us for failing to take this fact into account in the pronunciation classroom.

Viewing this matter in a social light, from the perspective of the evaluator, it is possible that when native listeners of English under-estimate the age of foreigners speaking English, they also under-estimate their maturity, experience and intelligence and as a result, acting and behaving according to this false premise, they then treat the foreigners with inappropriate condescension. The foreign speaker of English may then in turn consider the native speaker of English to be snobbish, patronising and pretentious. Such speculative conclusions could easily be confirmed in further studies.

The ramifications of under-estimating the age of speakers on the basis of vocal cues and foreign accent are enormous.

8.2. Estimated excellence of the pronunciation of speakers

The second question that listeners had to answer concerned the quality of the pronunciation: was it basically 'good' or 'bad' (Figure 8.14, P. 197).

How good do you think this foreigner's pronunciation of English is?

excellent pronunciation ☐ ☐ ☐ ☐ ☐ ☐ ☐ awful pronunciation

Figure 8.14

N.B. The following labels for the answer boxes will be used in the tables and bar-charts, going from 'excellent pronunciation' (+) to 'awful pronunciation' (-):

ex+ very+ fair+ middle fair- very- ex-

The intention with this question was to establish whether the listeners in fact recognise that there is a difference in the phonemic accuracy of the versions of the speakers they hear. If they do not, then it is hardly likely that they will form differential impressions of the fourteen speaker traits to be evaluated in Question 3, and no credence or value could be attached to any such impressions that might be found. Since the informants were assigned either Tape 1 or Tape 2 at random, the procedure was to discover whether the two informant groups are drawn from the same population, i.e. whether they have the same responses. If the level of 'excellence' of pronunciation is stochastically greater in one group or the other, then the null hypothesis must be rejected and the conclusion is that the two groups of informants come from populations with different distributions, i.e. that the stimulus material is not just different—as it most categorically is—but is recognisably different.

The results for the four separate groups of informants (Finnish-speaking listeners Tape 1, Finnish-speaking listeners Tape 2, English-speaking listeners Tape 1 and English-speaking listeners Tape 2) were non-uniform, Chi Square analysis revealing a level of significance of $p < .001$. However, when the median is 'middle,' visual checking of the associated bar-chart might prompt one to be careful here since a bi-modal result might indicate a symmetrical split between positive and negative values. Such is the case with the English-speaking listeners evaluating the Fair Pronunciation (Tape 2) of Speaker 1 (see Table 8.20 and Figure 8.16, P. 199), and indeed the Kolmogorov-Smirnov one-sample test (which depends on the *maximum* deviation from the expected cumulative distribution) not surprisingly shows this bi-modal distribution to be insufficiently deviant to reject the null hypothesis ($p > .05$). When the Kolmogorov-Smirnov one-sample test was used to check all the other results, all except three were found to confirm the Chi-Square analyses at the $p < .001$ significance level. The three exceptions were significant at the $p < .01$ level and all involved English-speaking listeners: Speaker 4 Tape 1 Poor Pronunciation (Table 8.26, Figure 8.22, P. 206); Speaker 4 Tape 2 Good Pronunciation (Table 8.26, Figure 8.22, P. 206); Speaker 5 Tape 2 Fair Pronunciation (Table 8.28, Figure 8.24, P. 208).

The non-uniform distribution across the 7-point scale, therefore, needs to be interpreted with caution, since it is perhaps more important to know not so much whether the seven points on the scale are favoured unequally, but whether the

positive (or negative) pole of the scale is favoured more than the other. (This same element of caution will be necessary in the evaluation of the fourteen traits in Question 3.)

Speaker 1

The Finnish-speaking listeners clearly evaluated the pronunciation of Speaker 1 in two different ways, positively and negatively, according to which tape they heard, Excellent Pronunciation or Fair Pronunciation (Table 8.19 and Figure 8.15).

Table 8.19 Speaker 1: 23-year-old woman

Tape	Mode	Median
1: 0 mistakes – 'Excellent'	very+	fair+
2: 31 mistakes – 'Fair'	very-	fair-

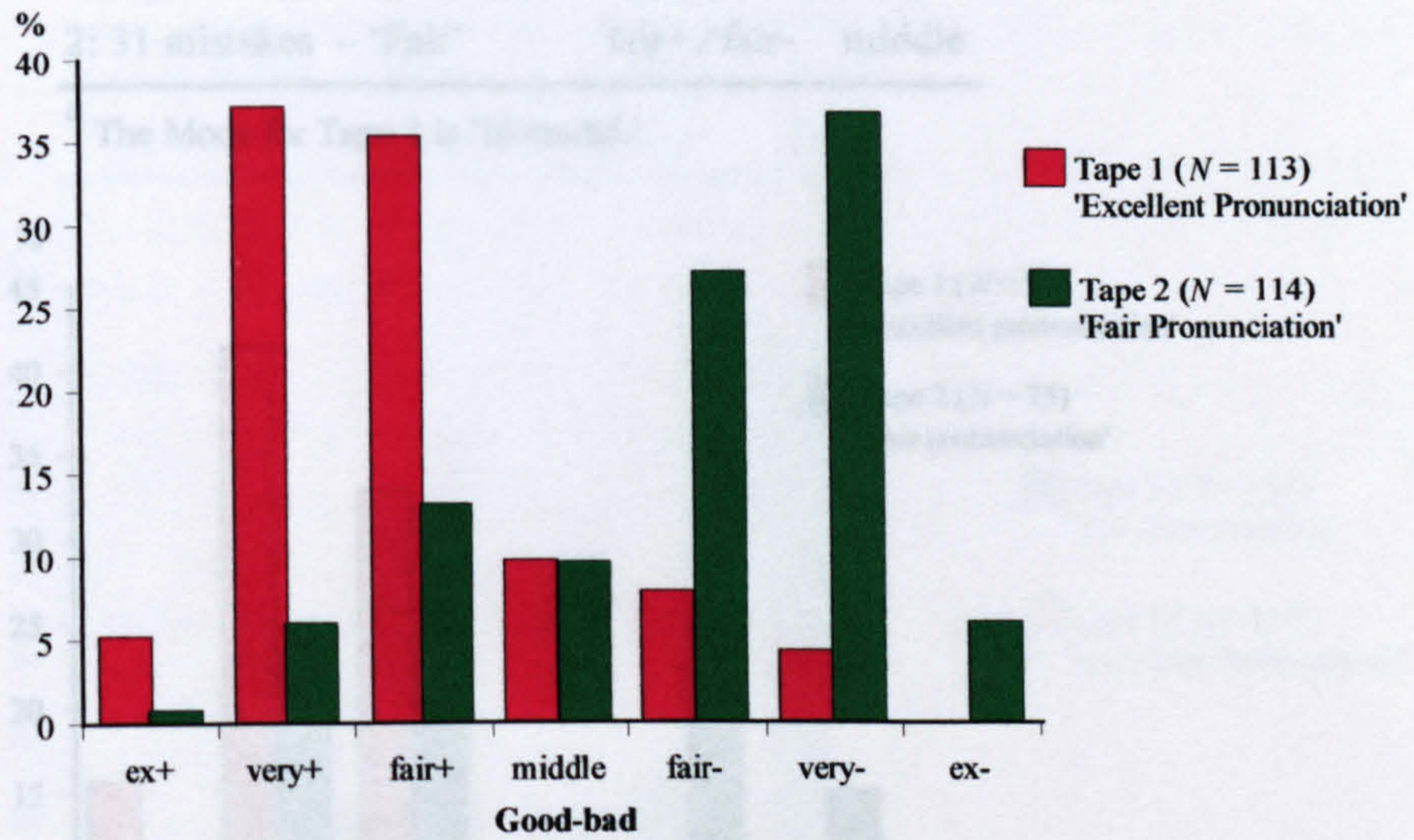


Figure 8.15 Finnish-speaking listeners evaluating quality of Speaker 1

It should be remembered that with this first speaker the informants have no anchor point for comparison purposes. Despite this fact, the contrast in the data is stark, and both the Median Test and the Mann-Whitney *U* test confirm that the difference is statistically significant ($p < .001$).

Interestingly the Excellent Pronunciation is not considered to be 'ex+,' even though it is phonemically without blemish. Either Finnish-speaking informants are uncomfortable about using the extreme limit of the marking scale (the so-called error of severity: see Section 2.1.2.), or they recognise something in the pronunciation which marks the speaker as a non-native speaker, and hence they deem the pronunciation to be less than perfect. (Unfortunately the experiment included no native speaker with 'perfect' pronunciation to see whether informants would rate that as 'excellent.')

The Fair Pronunciation is immediately recognised as being weak and accordingly given a negative evaluation. There is no hesitation here, no mercy: the accent is bad. It is impossible to say whether informants instantly guessed the

speaker to be a Finn, but it is reasonable to assume that they did. Hence one may conclude that Finns—presumably as a result of the barrage of garbled English that they hear every day on television and radio and elsewhere—are well up to the task of evaluating Finnish-accented English and are more than willing to denigrate a speaker when denigration is deserved.

The English-speaking listeners are not so clearly divided on the positive–negative axis in their evaluations of the two versions of Speaker 1, though there is a significant difference between the evaluations of the two tapes: one version is perceived as being 'better' in quality than the other: both Median Test and Mann-Whitney *U* test, $p < .001$ (Table 8.20 and Figure 8.16).

Table 8.20 Speaker 1: 23-year-old woman

Tape	Mode [a]	Median
1: 0 mistakes – 'Excellent'	very+	very+
2: 31 mistakes – 'Fair'	fair+/fair-	middle

[a] The Mode for Tape 2 is 'bi-modal.'

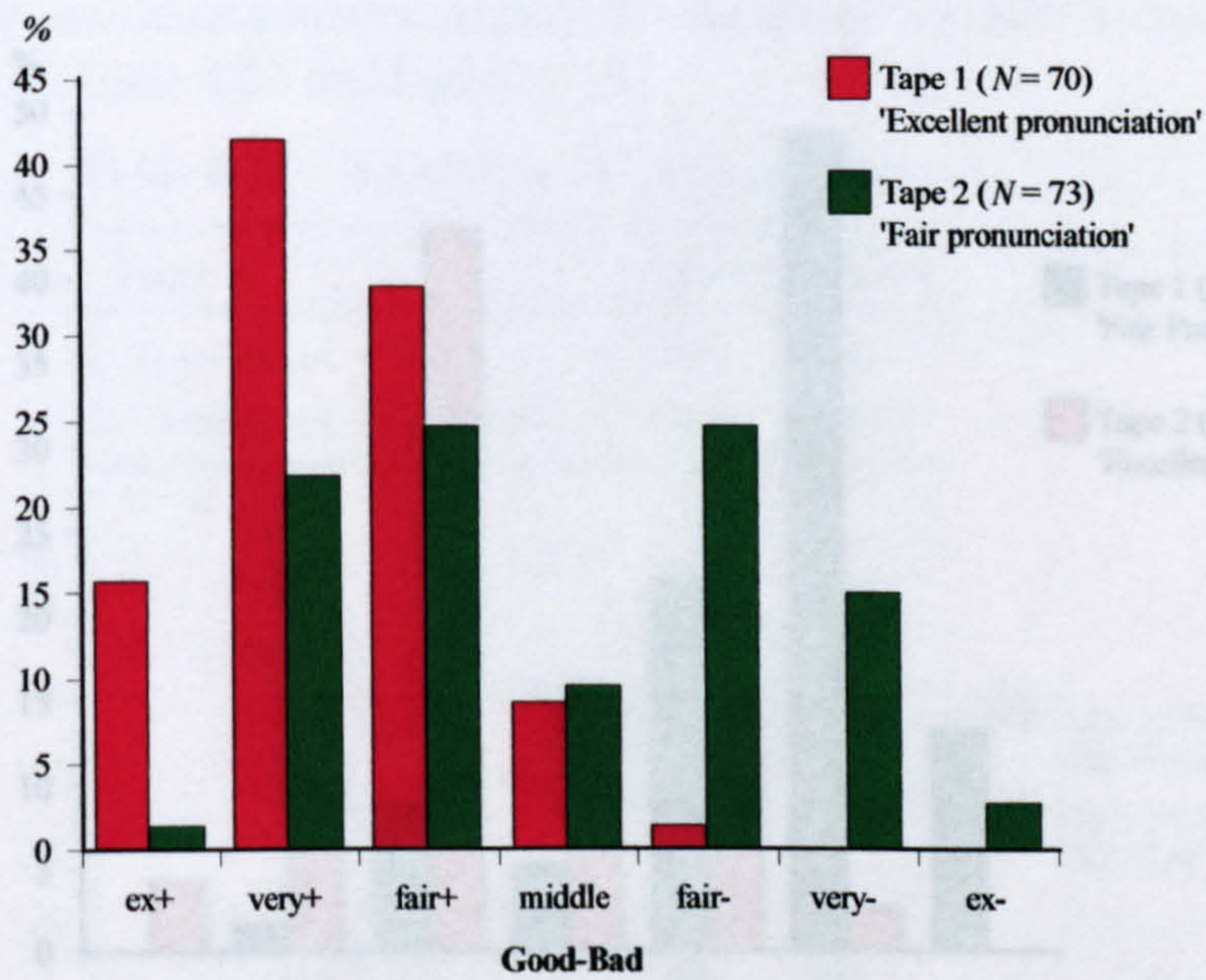


Figure 8.16 English-speaking listeners evaluating quality of Speaker 1

Interestingly the majority of English-speaking listeners (just like the Finnish-speaking listeners) do not deem the Excellent Pronunciation to be 'ex+': once again, either the end point of the scale is shunned, or else the trace of a foreign accent (despite the phonemic accuracy) is sufficient to prevent the attribution of the label 'excellent.'

The Fair Pronunciation does not earn a clear negative evaluation. Although presumably recognised as containing phonemic inaccuracies, it appears to pose difficulties for the informants: what is the bench mark to be used for this first evaluation? The results are split bi-modally and almost symmetrically between a positive and a negative evaluation: positive 48.0%; 'middle' 9.6%; negative

42.5%. With this initial shot in the dark, some native English listeners are more tolerant, some less tolerant, but with no objective external standard available, it is noteworthy that only 9.6% chose the middle category: the box was not specifically labelled as 'neutral,' 'central,' 'middle' or whatever on the questionnaire, though the instructions on the first page described it as being 'neutral' ('neutraali' in the Finnish version). Relatively few informants, in other words, opted for a bland non-committal 'neither-one-nor-the-other' category.

Speaker 2

The Finnish-speaking listeners easily recognised Speaker 2's pronunciation as being 'better' or 'worse' depending on the tape heard, Fair Pronunciation or Excellent Pronunciation (Table 8.21 and Figure 8.17).

Table 8.21 Speaker 2: 26.1-year-old man

Tape	Mode	Median
1: 31 mistakes – 'Fair'	very-	very-
2: 0 mistakes – 'Excellent'	fair+	fair+

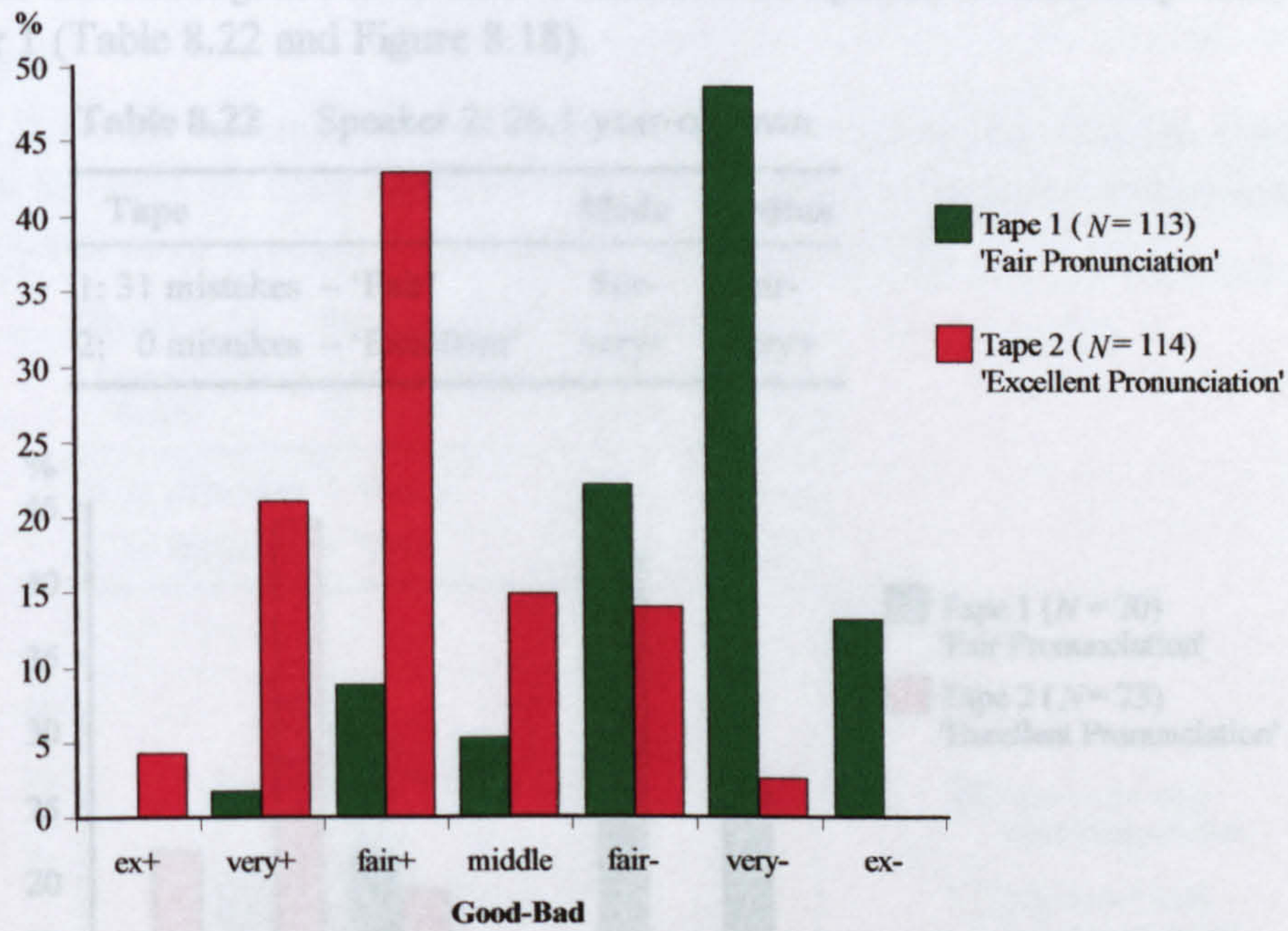


Figure 8.17 Finnish-speaking listeners evaluating quality of Speaker 2

The difference between the two impressions is statistically significant: both Median Test and Mann-Whitney *U* test, $p < .001$.

The overall impression gained from the Excellent Pronunciation by the Finnish-speaking listeners is only 'fair+,' which is nonetheless a higher (more positive) grade than the evaluation they gave to the Fair Pronunciation of Speaker 1. However, one might be tempted to claim that the Excellent Pronunciation of this male speaker is being downgraded more (for whatever trace of a foreign accent he might have, despite his phonemic accuracy) than the Excellent Pronunciation of Speaker 1 (a woman): the same level of phonemic accuracy in a man would

appear to be insufficient of itself to earn as high a rating on a scale of quality as the same level in a woman. However, such a claim would be highly contentious since several intervening variables (such as prosodic and/or suprasegmental features, and order of presentation of the versions) may well be exerting an influence here. More noteworthy is the fact that the number of informants who gave the Excellent Pronunciation a negative or 'middle' rating is considerable: 2.6% 'very-,' 14% 'fair-,' and 14.9% 'middle,' total 31.5%.

The Fair Pronunciation is clearly rated as 'very-': 48.7 % of informants chose 'very-,' and 84.1% chose either 'ex-' or 'very-' or 'fair-.' The contrast with the first speaker is stark for the corresponding figure there for the total of 'negative votes' is only 70.1%: the suspicion again arises that the man is being downgraded more for Fair Pronunciation than the woman. It would seem that a Finnish woman speaking English is upgraded and a Finnish man speaking English is downgraded, but the evidence from just these first two speakers is naturally not enough to claim this gender differentiation with any confidence. The order of presentation of speakers may be playing a decisive role here.

The two groups of English-speaking listeners appear to have been much more confident and homogeneous in their evaluation of Speaker 2 than they were for Speaker 1 (Table 8.22 and Figure 8.18).

Table 8.22 Speaker 2: 26.1-year-old man

Tape	Mode	Median
1: 31 mistakes – 'Fair'	fair-	fair-
2: 0 mistakes – 'Excellent'	very+	very+

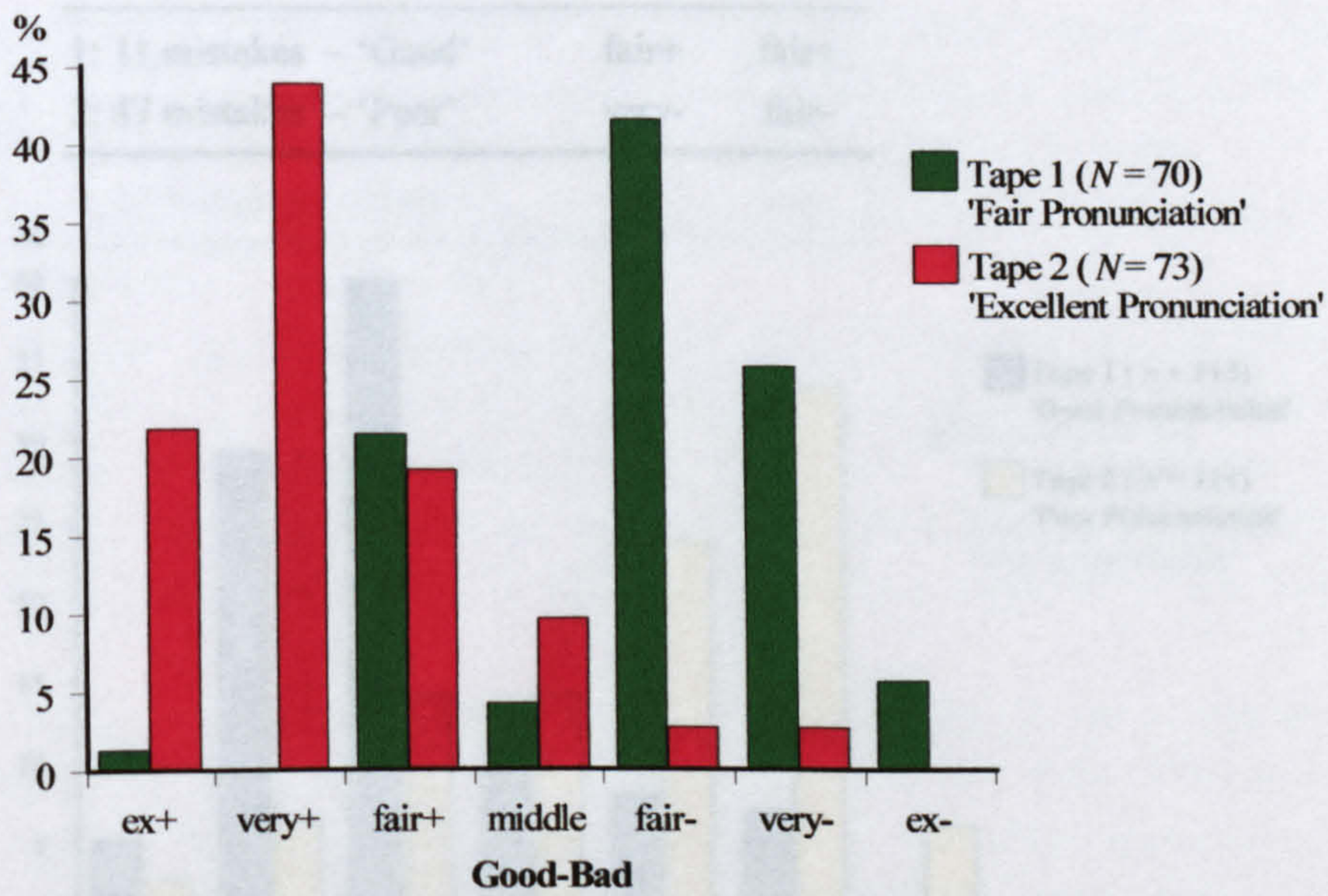


Figure 8.18 English-speaking listeners evaluating quality of Speaker 2

There is again a significant difference in the evaluation of the two tapes: both Median Test and Mann-Whitney *U* test, $p < .001$.

The Excellent Pronunciation is again perceived as being only 'very+,' presumably whatever traces of a foreign accent there may be in the speech being sufficient once more to prevent the native listeners of English from using the end box 'excellent.'

The Fair Pronunciation clearly prompts a negative evaluation ('fair-' 41.4%, 'very-' 25.7%, 'ex-' 5.7%: total 72.8%), though some informants are admittedly drawn to a positive evaluation ('fair+' 21.4%, or 'ex+' 1.4%). These positive responses are perhaps surprising in that the evaluation of this version of the speaker has the Excellent Pronunciation of Speaker 1 as a bench mark. However, the fact is that informants can decide for themselves what level of phonemic inaccuracy deserves the label 'weak,' 'bad,' 'awful' and so on: no objective criterion can be used here; the scale is not anchored externally. Nevertheless, the evaluation of the Fair Pronunciation here betrays none of the hesitation or prevarication that characterises the reactions of the English-speaking listeners who heard the Fair Pronunciation of Speaker 1 first: native listeners would appear to be uncertain how to react to foreigners when first hearing them, to be hesitant, tolerant. But once a bench mark has been set—as is the case for this second speaker—then their critical appraisal is more confident.

Speaker 3

Judging from the values for mode and median, it would appear that the Finnish-speaking listeners are once again clear in their positive or negative evaluations of the two versions of Speaker 3 (Table 8.23 and Figure 8.19).

Table 8.23 Speaker 3: 26.5-year-old man

Tape	Mode	Median
1: 11 mistakes – 'Good'	fair+	fair+
2: 47 mistakes – 'Poor'	very-	fair-

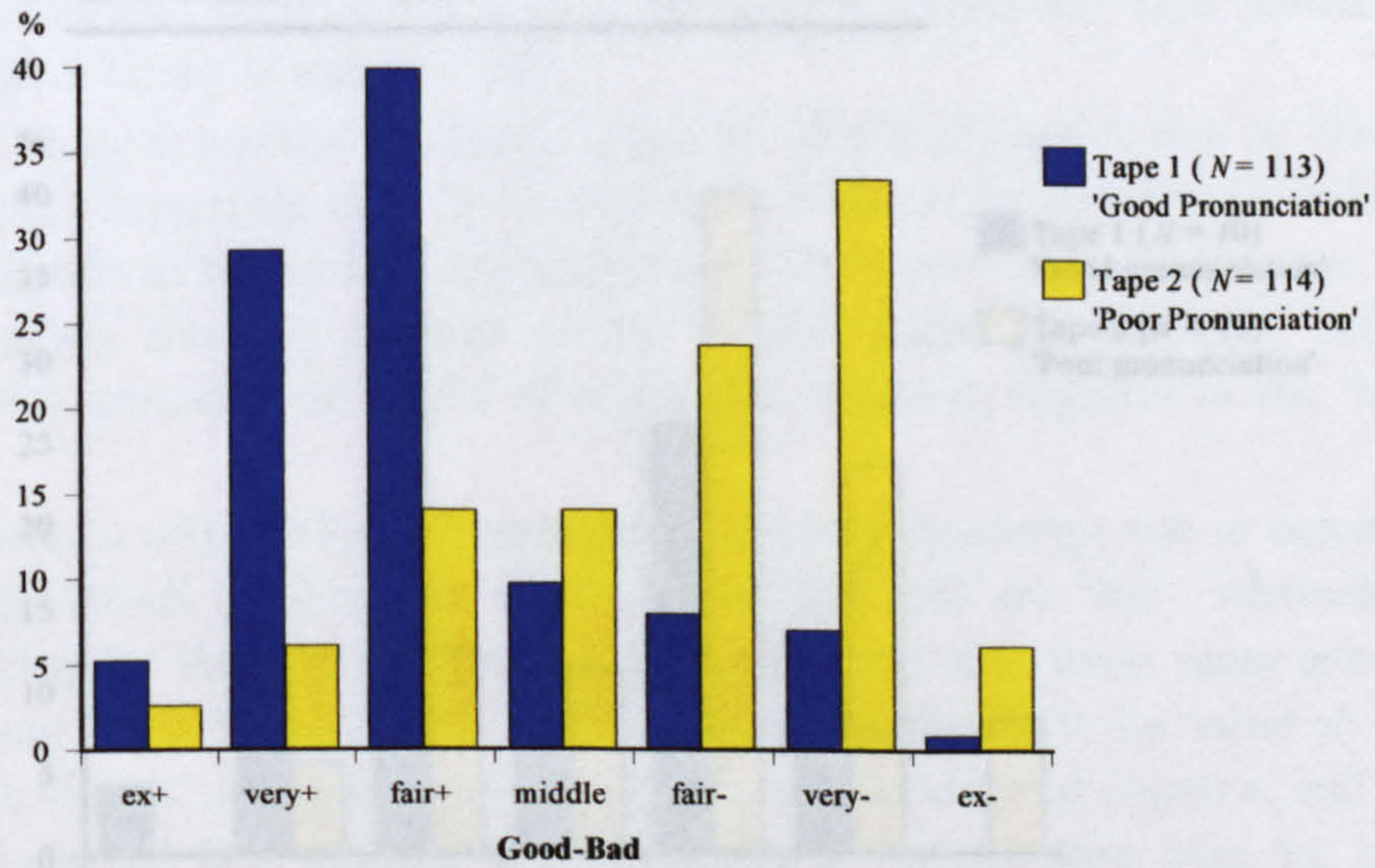


Figure 8.19 Finnish-speaking listeners evaluating quality of Speaker 3

The difference between the two evaluations here is statistically significant: both Median Test and Mann-Whitney *U* test, $p < .001$.

The Good Pronunciation provokes a positive response, with relatively few scores appearing on the negative end of the scale, or on the middle point. Reassuringly the group of informants who heard this tape clearly swing back to the positive end of the scale here from the negative evaluation they gave the Fair Pronunciation of Speaker 2, but they do not seem to swing to quite as far back as the positive evaluation they gave the Excellent Pronunciation of Speaker 1.

The Poor Pronunciation provokes a negative evaluation—as one might have expected—though a total of 22.7% of informants judged the accent positively and a further 14% in the middle, total 36.7%. This is surprising: the listeners in this informant group heard first a Fair Pronunciation, then second an Excellent Pronunciation, and now a Poor Pronunciation which is worse than either of the previous two. Nevertheless, the evaluation does *not* swing violently to the negative end of the scale: indeed the rating is moderately *more* positive overall than that for the Fair Pronunciation of the first speaker (for the Fair Pronunciation only 20.2% of informants judged the accent positively and a further 9.6 % in the middle, total 29.8%). One is tempted to suggest that familiarity breeds tolerance, though once again any interpretation here must be made cautiously because of the possible influence of intervening variables and of order of presentation.

The English-speaking listeners would also seem to be clearly divided between positive and negative in their reactions to the two versions of Speaker 3 (Table 8.24 and Figure 8.20). There is again a significant difference in the evaluation of the two tapes: both Median Test and Mann-Whitney *U* test, $p < .001$.

Table 8.24 Speaker 3: 26.5-year-old man

Tape	Mode	Median
1: 11 mistakes – 'Good'	fair+	fair+
2: 47 mistakes – 'Poor'	fair-	fair-

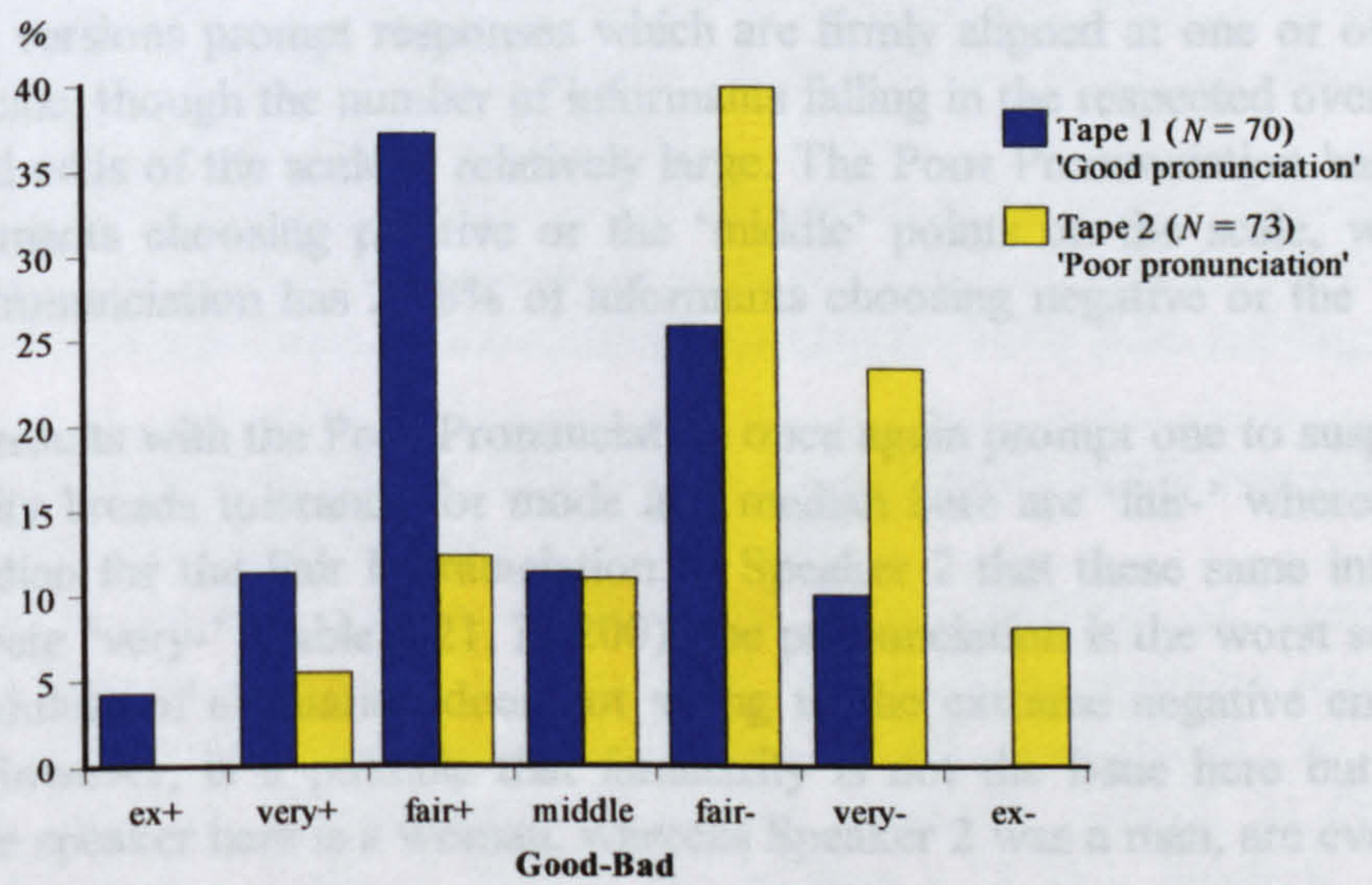


Figure 8.20 English-speaking listeners evaluating quality of Speaker 3

However, the bar chart for the evaluation of the Good Pronunciation reveals another situation where a considerable proportion of the informants judge the pronunciation to be negative and not positive: 35.7% claim the accent to be negative (10% 'very-' + 25.7% 'fair-'), and a further 11.4% 'middle,' total 47.1%. On the other hand, it is again reassuring that the English-speaking listeners' overall evaluation of this speaker, whose standard of pronunciation is between that of the first and second speakers on this tape, has swung back from that awarded the second speaker towards—but not as far as—that awarded the first speaker.

The informants' responses to the Poor Pronunciation are more clearly focused at one end of the scale, the negative end: only 17.8% of informants chose either 'very+' or 'fair+,' and 11% 'middle,' total 28.8%. The swing from the positive evaluation of Speaker 2 (Excellent Pronunciation) to this negative evaluation is obvious: the hesitation and uncertainty which seems to have coloured the evaluation of the first Speaker (Fair Pronunciation) has been dispelled. One could argue that the Fair Pronunciation of the first speaker was neither too bad nor too good to prompt a uni-modal result, that a foreigner's mispronunciation must be noticeably bad before native English listeners move en masse to a negative, less tolerant appraisal (the 22.4% of informants who gave a negative evaluation of the Fair Pronunciation of Speaker 2 would seem to give qualified support to such a view). However, the clear swing from positive to negative in the evaluations of Speaker 2 and Speaker 3 here can also be interpreted as support for the claim that the native English listeners listening to Tape 2 were puzzled by the first speaker, having no external standard available to perform the task they had been asked to do.

Speaker 4

The two groups of Finnish-speaking listeners are also distinct from each other in their negative and positive reactions to the Poor Pronunciation and Good Pronunciation respectively of Speaker 4 (Table 8.25 and Figure 8.21, P. 205).

The difference between the two evaluations is significant: both Median Test and Mann-Whitney U test, $p < .001$.

Both versions prompt responses which are firmly aligned at one or other end of the scale, though the number of informants falling in the respected overall non-favoured ends of the scale is relatively large. The Poor Pronunciation has 37.1% of informants choosing positive or the 'middle' points on the scale, while the Good Pronunciation has 29.8% of informants choosing negative or the 'middle' points.

The results with the Poor Pronunciation once again prompt one to suspect that familiarity breeds tolerance for mode and median here are 'fair-' whereas mode and median for the Fair Pronunciation of Speaker 2 that these same informants heard were 'very-' (Table 8.21, P. 200): the pronunciation is the worst so far but the pendulum of evaluation does not swing to the extreme negative end of the scale. However, is it possible that familiarity is not the issue here but gender: since the speaker here is a woman, whereas Speaker 2 was a man, are evaluations

being upgraded for the feature 'female speaker'? The data allow no such conclusion to be drawn, though the suspicion lingers that it is possible.

Table 8.25 Speaker 4: 20.1-year-old woman

Tape	Mode	Median
1: 47 mistakes – 'Poor'	fair-	fair-
2: 11 mistakes – 'Good'	fair+	fair+

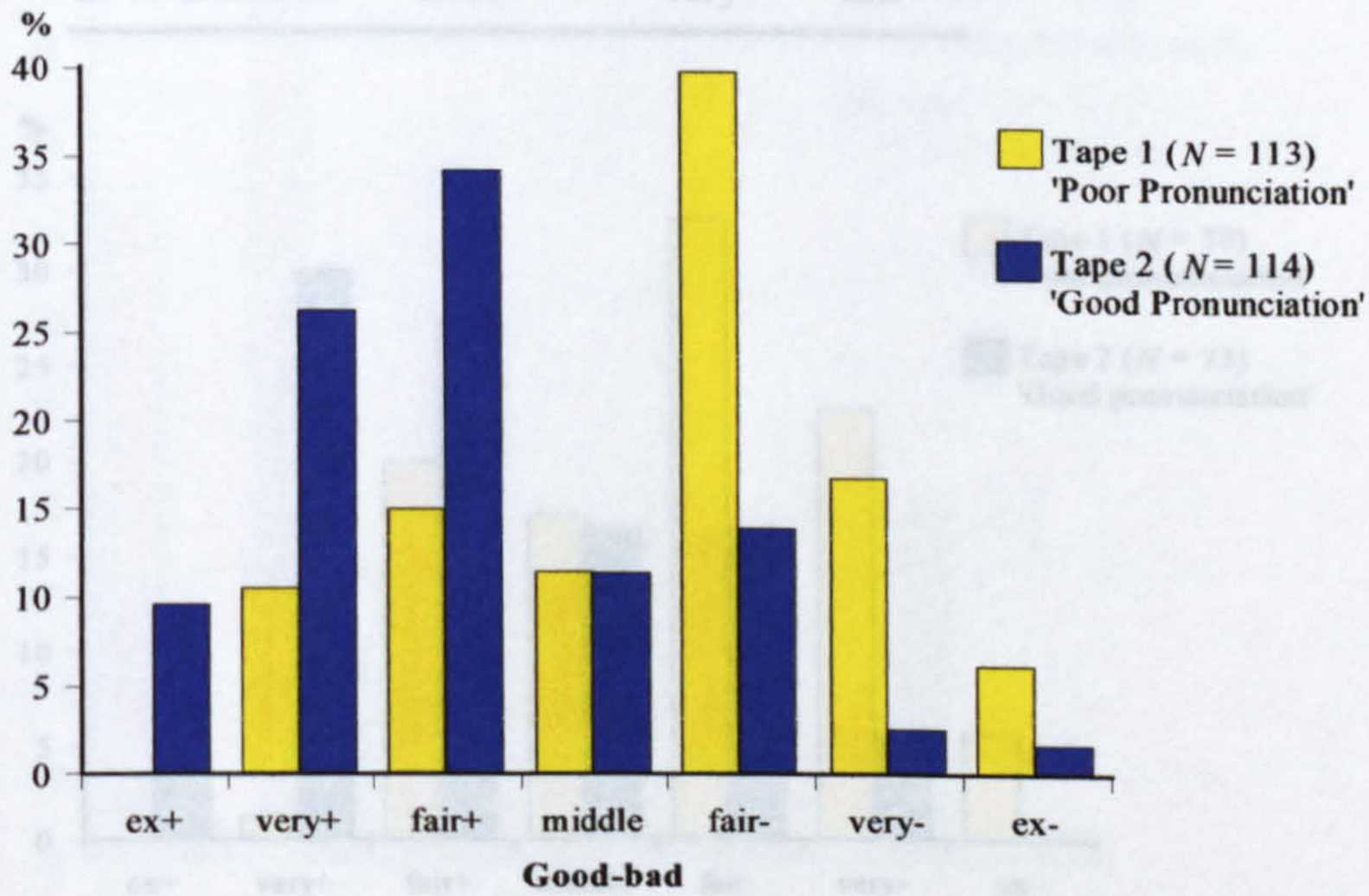


Figure 8.21 Finnish-speaking listeners evaluating quality of Speaker 4

The data for the Good Pronunciation here lends some support to this interpretation for 70.1% of informants gave this version a positive evaluation, the corresponding figure for the Excellent Pronunciation of Speaker 2 being almost identical: 68.5%. In other words, the appraisal of the Good Pronunciation of Speaker 4 (a woman) would seem to confirm the impression that the Excellent Pronunciation of Speaker 2 (a man) was *down*graded, and that hence, yes, the *Poor* Pronunciation of Speaker 4 is *up*graded for the feature 'female speaker.' There is of course nothing in the data which proves this gender bias decisively, but the signs for suspecting this are already quite strong.

The English-speaking listeners would also seem to be clearly divided in their reactions to the versions of Speaker 4, negatively and positively for the Poor Pronunciation and the Good Pronunciation respectively (Table 8.26 and Figure 8.22, P. 206).

There is again a significant difference in the evaluation of the two tapes: both Median Test and Mann-Whitney U test, $p < .001$.

However, the bar chart again highlights certain problems of interpretation. Both the Poor Pronunciation and the Good Pronunciation prompt a considerable number of 'middle' responses: 17.1% and 16.4% respectively. In addition, although the Poor Pronunciation is viewed negatively overall, 21.4% of the ratings are positive (total of 'middle' and positive = 38.5%). Similarly, although the Good Pronunciation is viewed positively overall, 21.9% of the responses are

negative (total of 'middle' and negative = 38.3%). The impression is that both versions here are prompting responses which, although not level across the five central categories, are moving in that direction.

Table 8.26 Speaker 4: 20.1-year-old woman

Tape	Mode	Median
1: 47 mistakes – 'Poor'	fair-	fair-
2: 11 mistakes – 'Good'	very+	fair+

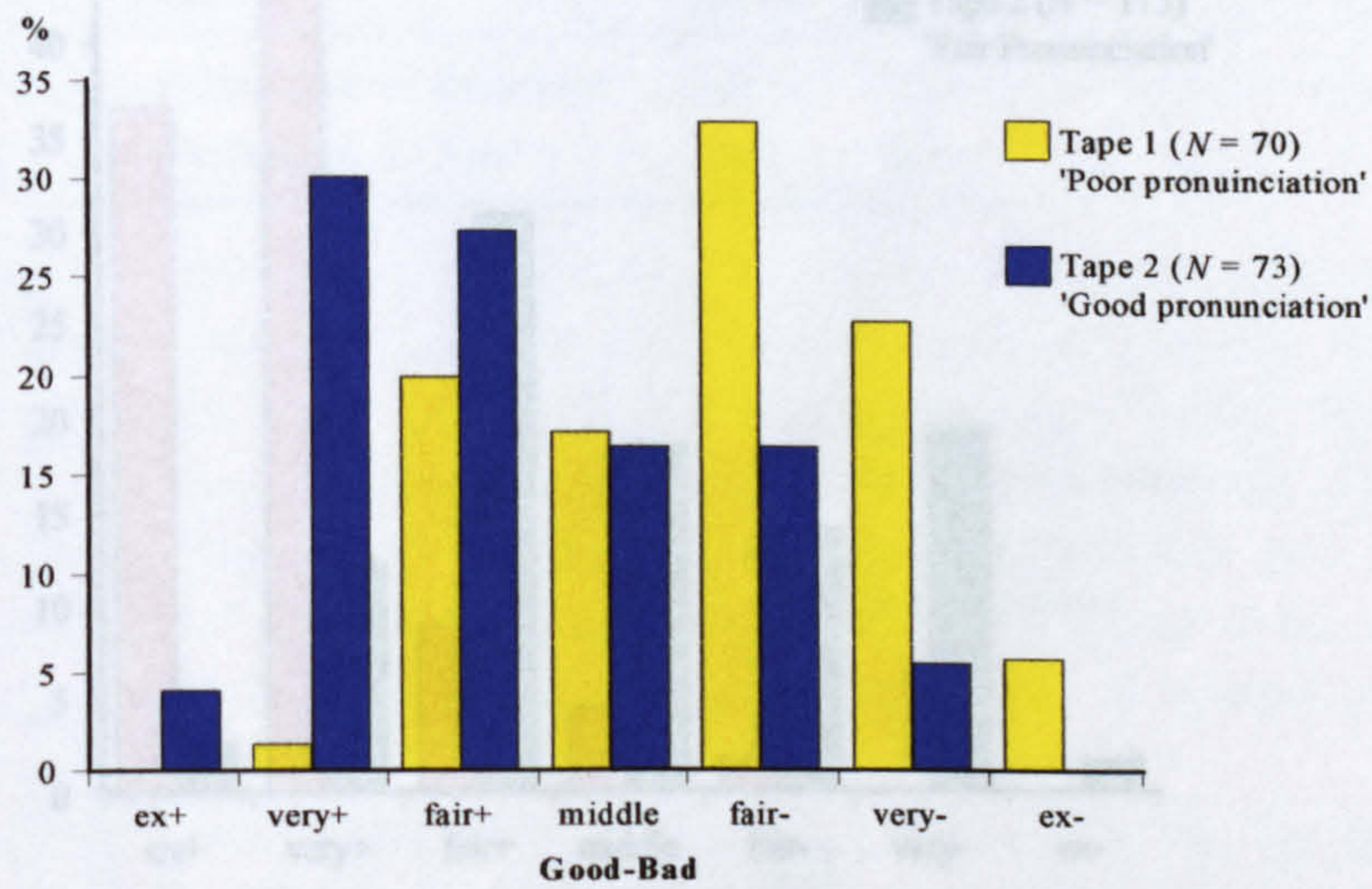


Figure 8.22 English-speaking listeners evaluating quality of Speaker 4

The Poor Pronunciation is the worst pronunciation that the native English listeners have heard so far, but despite this the mode and median do not swing to 'very-' let alone 'ex-': they are both 'fair-,' just as the Fair Pronunciation of Speaker 2 is. The overall negative vote ('fair-' + 'very-' + 'ex-') is 61.5%, which is indeed *less* than the corresponding figure for Speaker 2: 72.8%. Either increasing task and accent familiarity are leading to a numbing of critical acumen, or else intervening variables are distorting the evaluations, the prime suspect again being—I would contend—gender of voice: this woman's Poor Pronunciation is bad, worse than the Fair Pronunciation of Speaker 2, a man, and yet overall the rating is more positive than that of the Fair Pronunciation of Speaker 2.

The Good Pronunciation is reassuringly rated as being better than the Poor Pronunciation of Speaker 3 but not as good as the Excellent Pronunciation of Speaker 2: the swing of the critical pendulum is consistent with level of mispronunciation.

Speaker 5

The Finnish-speaking listeners clearly rate Speaker 5 positively for the Excellent Pronunciation but are less decisive and less homogeneous in their rating of the Fair Pronunciation (middle-tending-to-positive) (Table 8.27 and Figure 8.23, P. 207).

Table 8.27 Speaker 5: 20.75-year-old woman

Tape	Mode	Median
1: 0 mistakes – 'Excellent'	very+	very+
2: 33 mistakes – 'Fair'	fair+	middle

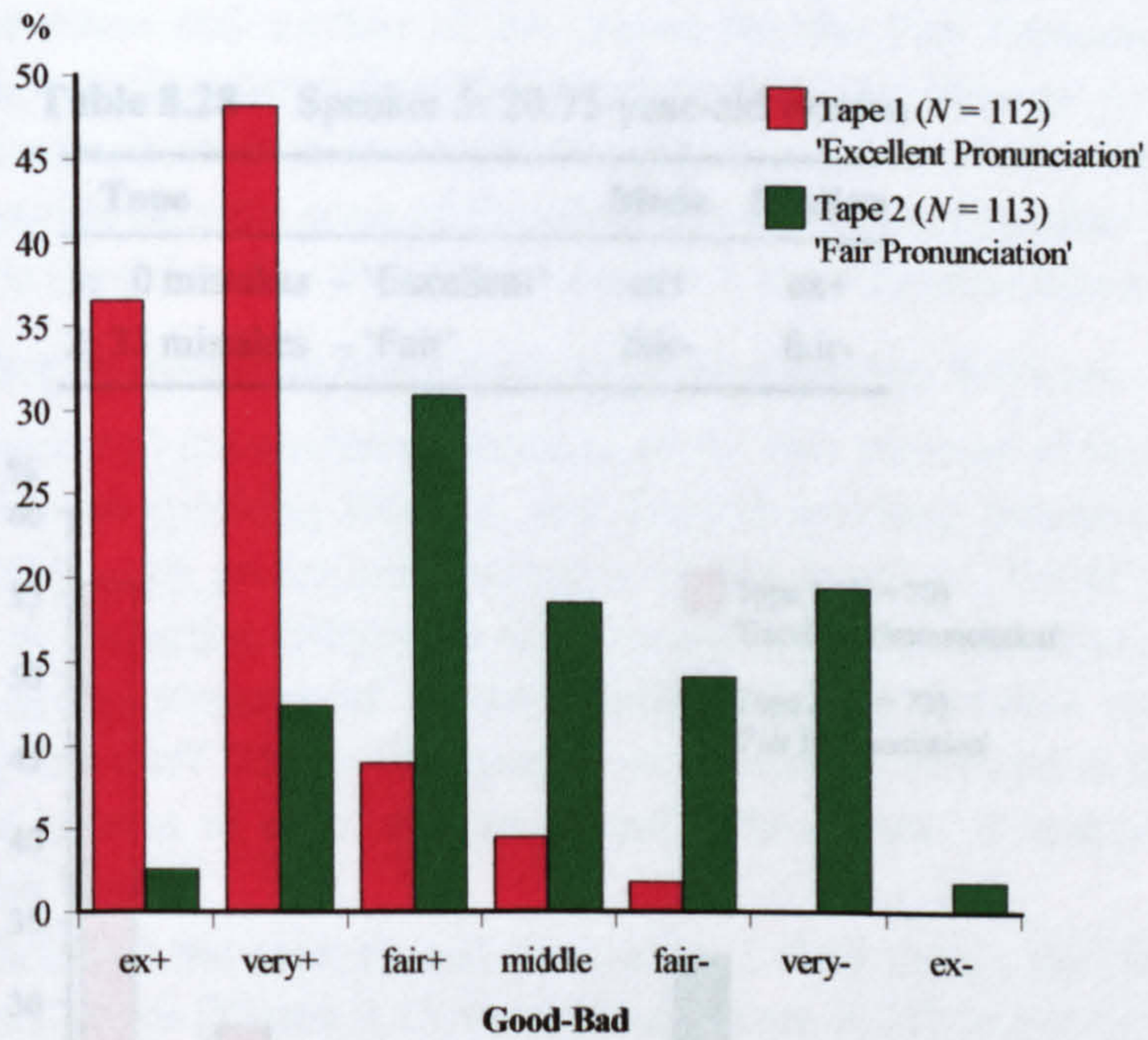


Figure 8.23 Finnish-speaking listeners evaluating quality of Speaker 5

The contrast between the two sets of data are statistically significant: both Median Test and Mann-Whitney *U* test: $p < .001$.

That such a large number (84.1%) of the native Finnish listeners who heard the Excellent Pronunciation rate it as either 'ex+' or 'very+' is truly remarkable, as also of course the fact that so few (6.2%) rated it as 'middle' or 'fair-': the congruity is striking. It is also manifest that this last speaker with her Excellent Pronunciation is rated more positively overall than the first speaker with her Excellent Pronunciation: various intervening variables may be the cause for this, but the fact that the first speaker had no reference point for undertaking the evaluation must be recognised as playing a part here.

By contrast the informants who heard the Fair Pronunciation would seem to have had problems in evaluating the speaker, even though they now had four other speakers to use as points of comparison; the mode is 'fair+,' but the median is 'middle' and that label ('middle') would seem to capture the impression given by the bar chart more accurately. The responses of the informants are not distributed uniformly across the 7-point scale, but there is a tendency to uniformity across the five central points, suggesting uncertainty and ambiguity in how to rate this speaker. Reassuringly the rating is more negative than that of the Good Pronunciation of Speaker 4 and the Excellent Pronunciation of Speaker 2, and more

positive than the Poor Pronunciation of Speaker 3. However, it would appear to be more positive than the Fair Pronunciation of the first speaker, possibly again a result of the fact that the first speaker had no reference point for the informants to start off with.

The two random groups of English-speaking informants are sharply differentiated in their evaluations of the two versions of Speaker 5 (Table 8.28 and Figure 8.24).

Table 8.28 Speaker 5: 20.75-year-old woman

Tape	Mode	Median
1: 0 mistakes – 'Excellent'	ex+	ex+
2: 33 mistakes – 'Fair'	fair-	fair-

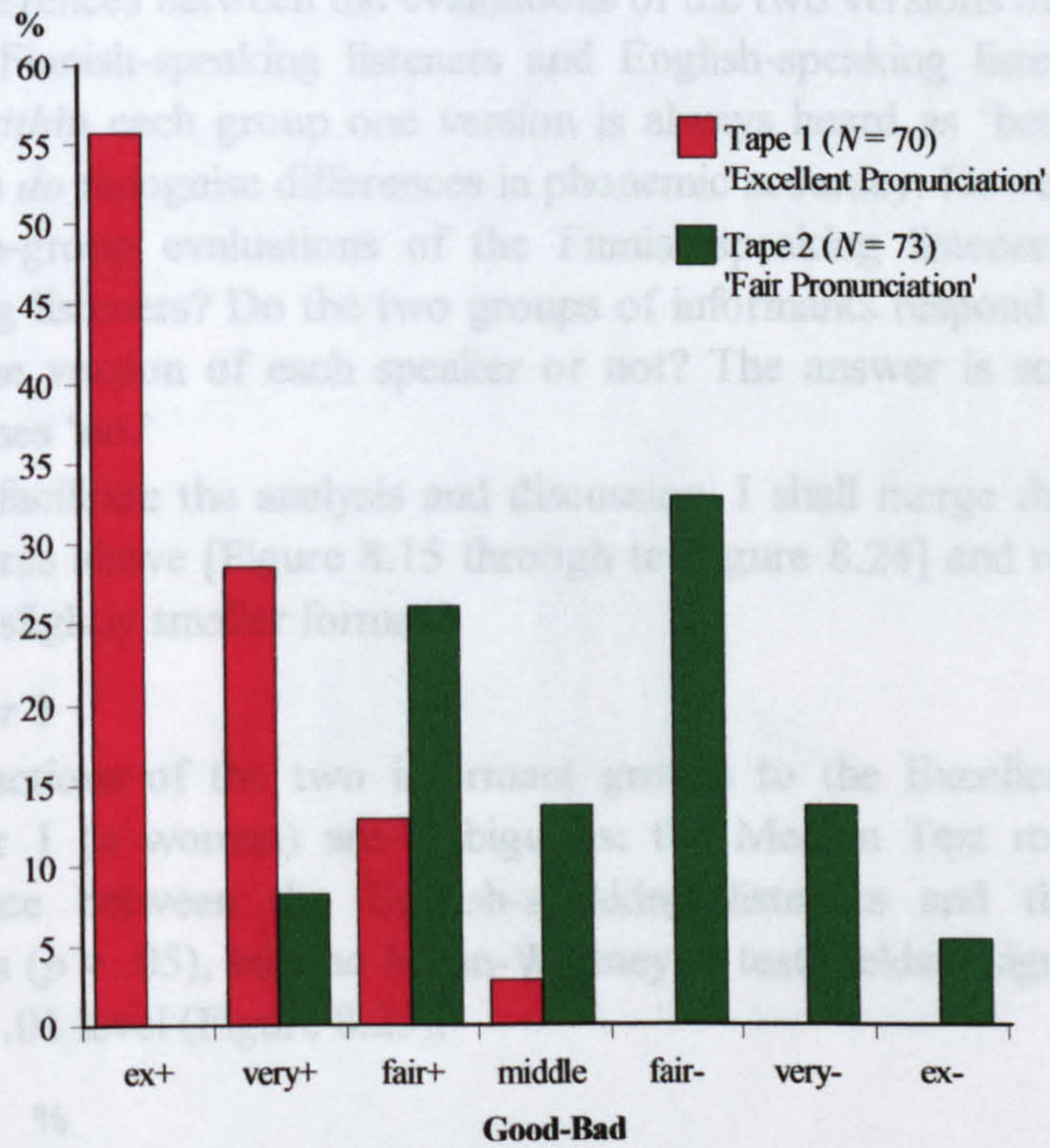


Figure 8.24 English-speaking listeners evaluating quality of Speaker 5

There is again a significant difference in the evaluation of the two tapes: both Median Test and Mann-Whitney U test, $p < .001$.

The Excellent Pronunciation of Speaker 5 prompts an extraordinary 55.7% of informants to tick the 'ex+' box on the scale. A further 41.5% chose positive categories, the positive total being 97.2%. The 'middle' category earned 2.9% of the votes, and the negative categories earned no votes at all. This rating is more positive than that for the Excellent Pronunciation of the first speaker: either this woman's voice has some prosodic or suprasegmental features (e.g. rhythm, tempo, pitch contours, voice quality) that provoke a more positive reaction, or else experience with the previous four speakers allows the informants to place this

speaker at her deserved high level. The corollary is necessarily that the first speaker was under-rated or downgraded through lack of experience and the lack of a firm reference point.

The reactions of the second group of English-speaking informants listening to the Fair Pronunciation of Speaker 5 are more spread out across the scale, recalling to some extent the bi-modal result obtained with Speaker 1's Fair Pronunciation. The mode and median of the scores for the Fair Pronunciation here are 'fair-,' though they were 'fair+/fair-' and 'middle' for Speaker 1: in other words the evaluation here is more negative overall, and indeed sits comfortably in its relative position on the scale of Excellent Pronunciation (Speaker 2) > Good Pronunciation (Speaker 4) > Fair Pronunciation > Poor Pronunciation (Speaker 3).

8.2.1. Finnish-speaking listeners v. English-speaking listeners

The differences between the evaluations of the two versions of each speaker made by the Finnish-speaking listeners and English-speaking listeners separately are clear: *within* each group one version is always heard as 'better' than the other; listeners *do* recognise differences in phonemic accuracy. However, what about the *between*-group evaluations of the Finnish-speaking listeners and the English-speaking listeners? Do the two groups of informants respond in the same way to the same version of each speaker or not? The answer is sometimes 'yes,' and sometimes 'no.'

(To facilitate the analysis and discussion, I shall merge the pertinent parts of the figures above [Figure 8.15 through to Figure 8.24] and re-present them here using a slightly smaller format.)

Speaker 1

The reactions of the two informant groups to the Excellent Pronunciation of Speaker 1 (a woman) are ambiguous: the Median Test reveals no significant difference between the English-speaking listeners and the Finnish-speaking listeners ($p > .05$), but the Mann-Whitney U test yields a significant difference at the $p < .01$ level (Figure 8.25).

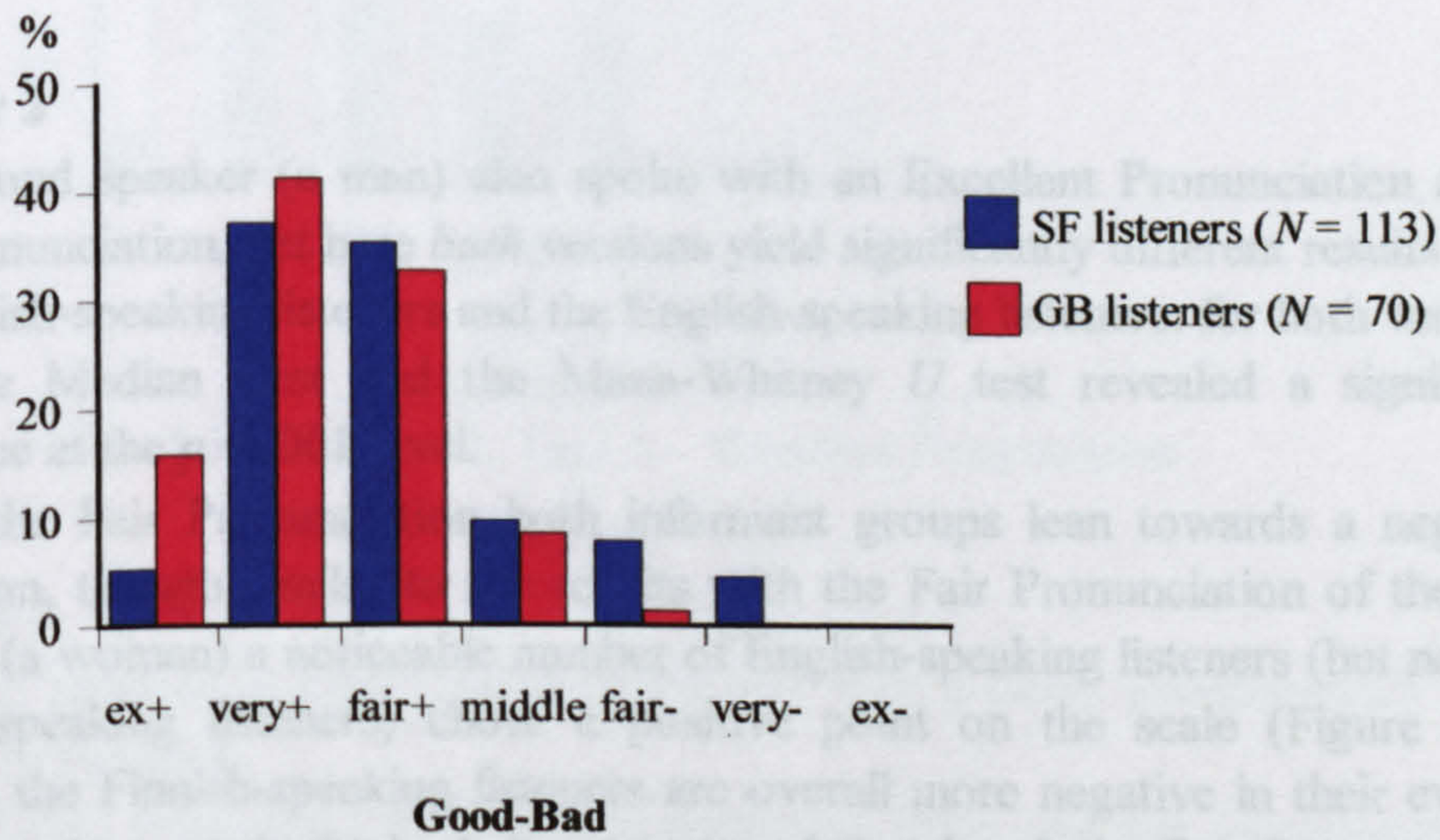


Figure 8.25 Speaker 1, Tape 1 – 'Excellent Pronunciation'

Casual visual interpretation of Figure 8.25 perhaps suggests the result of the Median Test is more acceptable (i.e. no significance), since there is certainly a degree of 'level-pegging' here: each informant group's expected scores above and below the combined median are no more or less than what one would expect by chance. Nevertheless, the English-speaking listeners are marginally more positive overall in their evaluations than the Finnish-speaking listeners, and the dispersion of the two group's scores across the 6 ranks used is indeed significantly different (the Mann-Whitney *U* test). In other words the native English listeners are marginally more tolerant than the native Finnish listeners.

The results for Speaker 1's Fair Pronunciation are clear-cut: both the Median Test and the Mann-Whitney *U* test show a significant difference between the evaluations of the native Finnish listeners and the native English listeners: $p < .001$. This is obviously a result of the bi-modal result with the English-speaking listeners: these listeners seem unsure what to do (this is a jump in the dark), whereas the Finnish-speaking listeners clearly lean towards a negative evaluation (Figure 8.26). Here too the English-speaking listeners are prepared to be more tolerant than the Finnish-speaking listeners in their first contact with this foreign-accented speech.

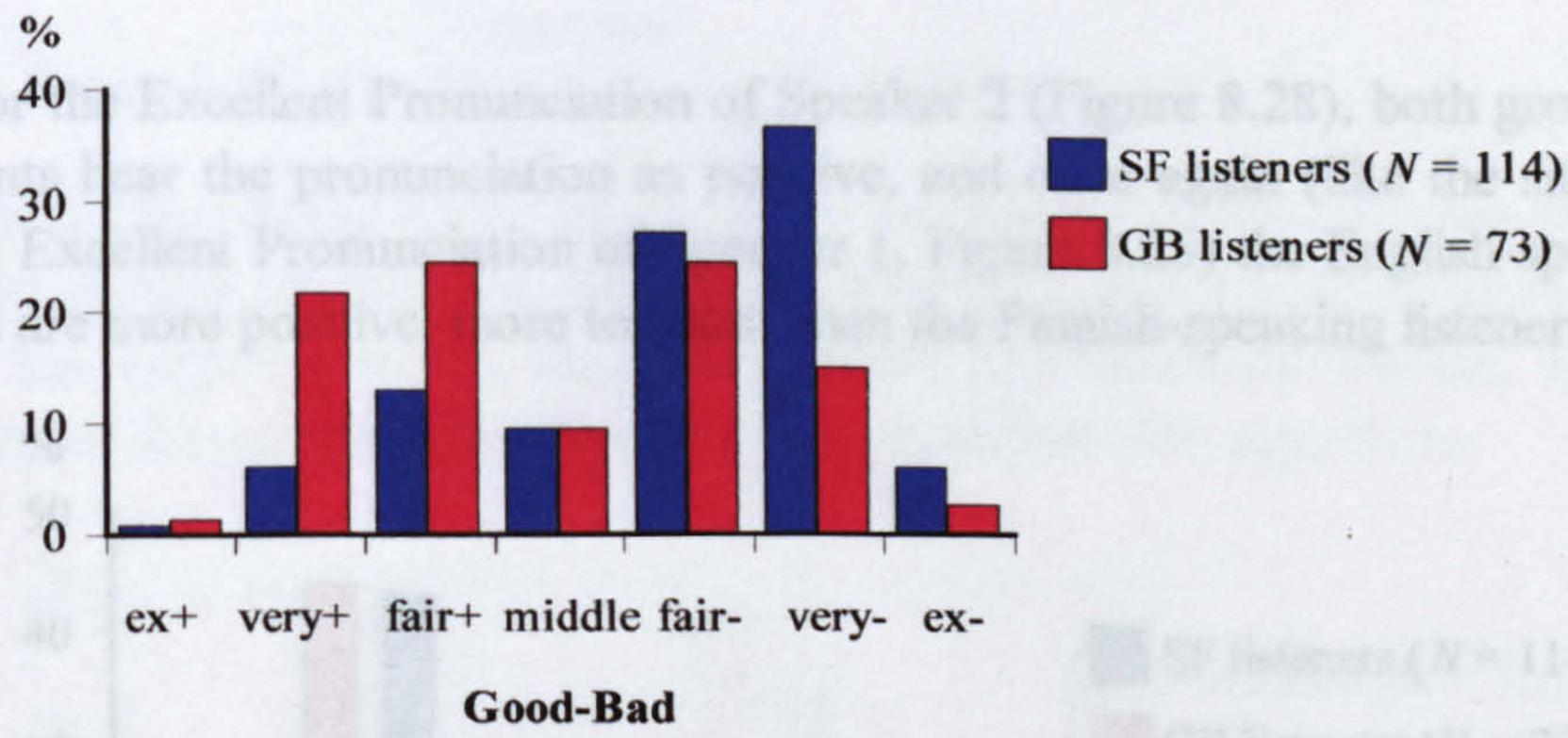


Figure 8.26 Speaker 1, Tape 2 – 'Fair Pronunciation'

Speaker 2

The second speaker (a man) also spoke with an Excellent Pronunciation and a Fair Pronunciation, yet here *both* versions yield significantly different results from the Finnish-speaking listeners and the English-speaking listeners: for both versions both the Median Test and the Mann-Whitney *U* test revealed a significant difference at the $p < .001$ level.

For the Fair Pronunciation both informant groups lean towards a negative evaluation, though similar to the results with the Fair Pronunciation of the first speaker (a woman) a noticeable number of English-speaking listeners (but not the Finnish-speaking listeners) chose a positive point on the scale (Figure 8.27, P. 211): the Finnish-speaking listeners are overall more negative in their evaluations. In other words for both Speaker 1 and Speaker 2 the Fair Pronunciation prompts a more positive rating from the native English listeners than from the

native Finnish listeners, though the situation is more obvious here with Speaker 2 (a man) than with Speaker 1 (a woman). This may be a result of the gender of the speaker, but it is impossible to say this for certain; order of presentation may be exerting an influence as well.

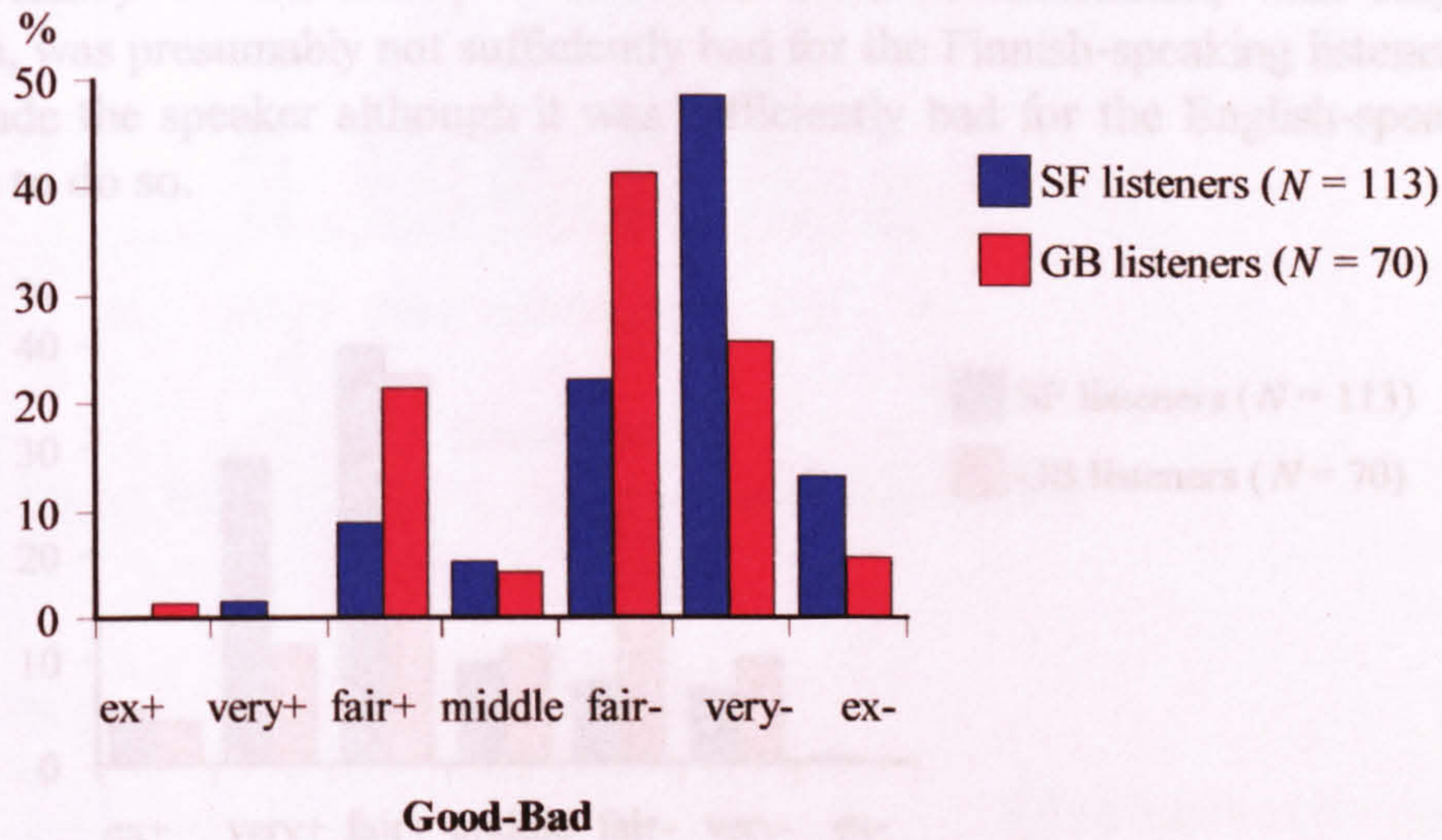


Figure 8.27 Speaker 2, Tape 1 – 'Fair Pronunciation'

As for the Excellent Pronunciation of Speaker 2 (Figure 8.28), both groups of informants hear the pronunciation as positive, and once again (like the situation with the Excellent Pronunciation of Speaker 1, Figure 8.25) the English-speaking listeners are more positive, more tolerant, than the Finnish-speaking listeners.

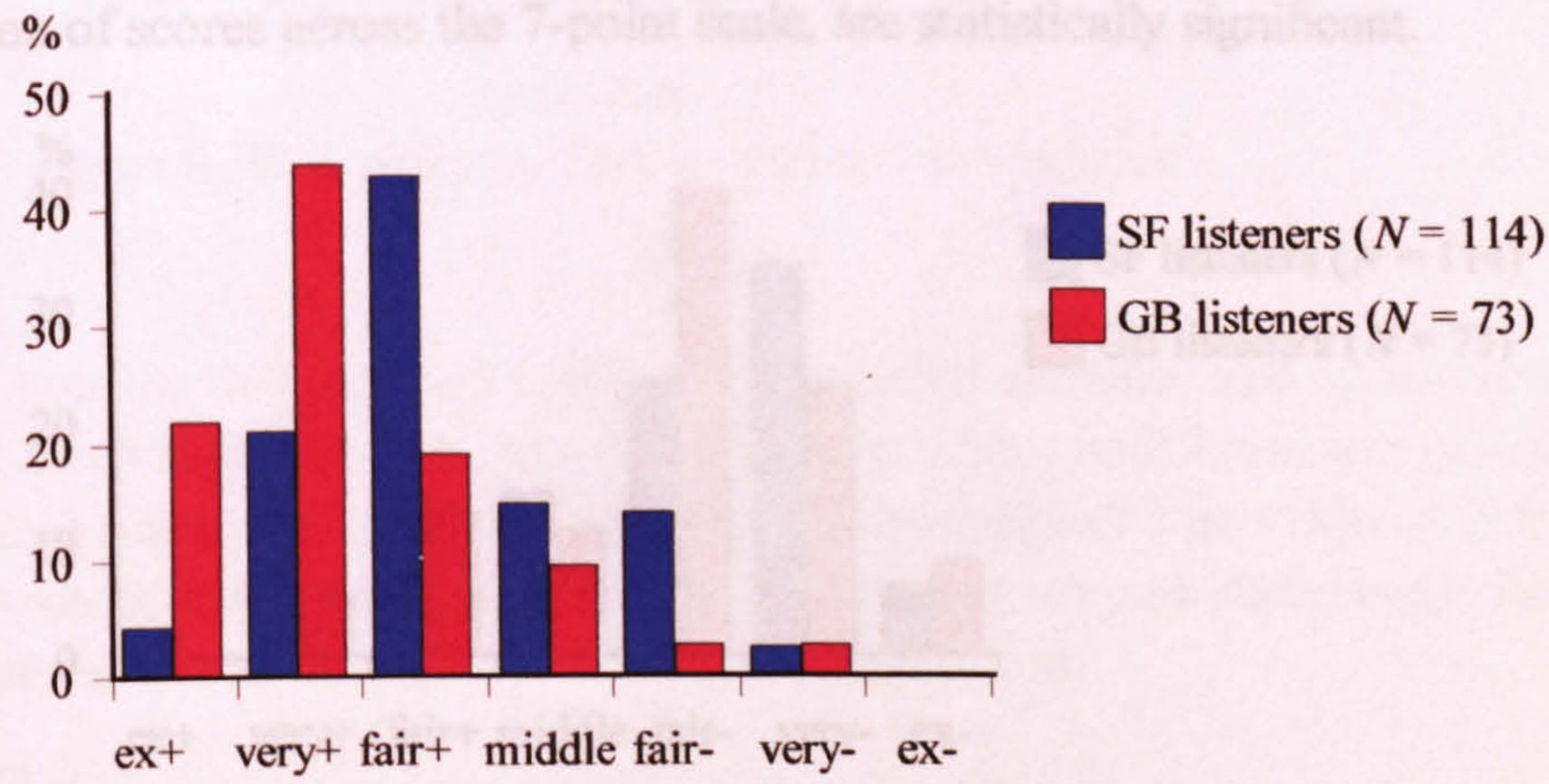


Figure 8.28 Speaker 2, Tape 2 – 'Excellent Pronunciation'

Speaker 3

The third speaker (a man) produced the text with a Good Pronunciation and a Poor Pronunciation.

The Good Pronunciation prompted responses from both the Finnish-speaking listeners and the English-speaking listeners with mode and median as 'fair+'

(Figure 8.19, P. 202 and Figure 8.20, P. 203). However, visual examination of the bar chart in Figure 8.29, below, clearly shows that the Finnish-speaking listeners were overall more positive in their evaluation of the Good Pronunciation than the English-speaking listeners, a point confirmed by both the Median Test and the Mann-Whitney *U* test: both $p < .01$. The Good Pronunciation, with only 11 mistakes, was presumably not sufficiently bad for the Finnish-speaking listeners to downgrade the speaker although it was sufficiently bad for the English-speaking listeners to do so.

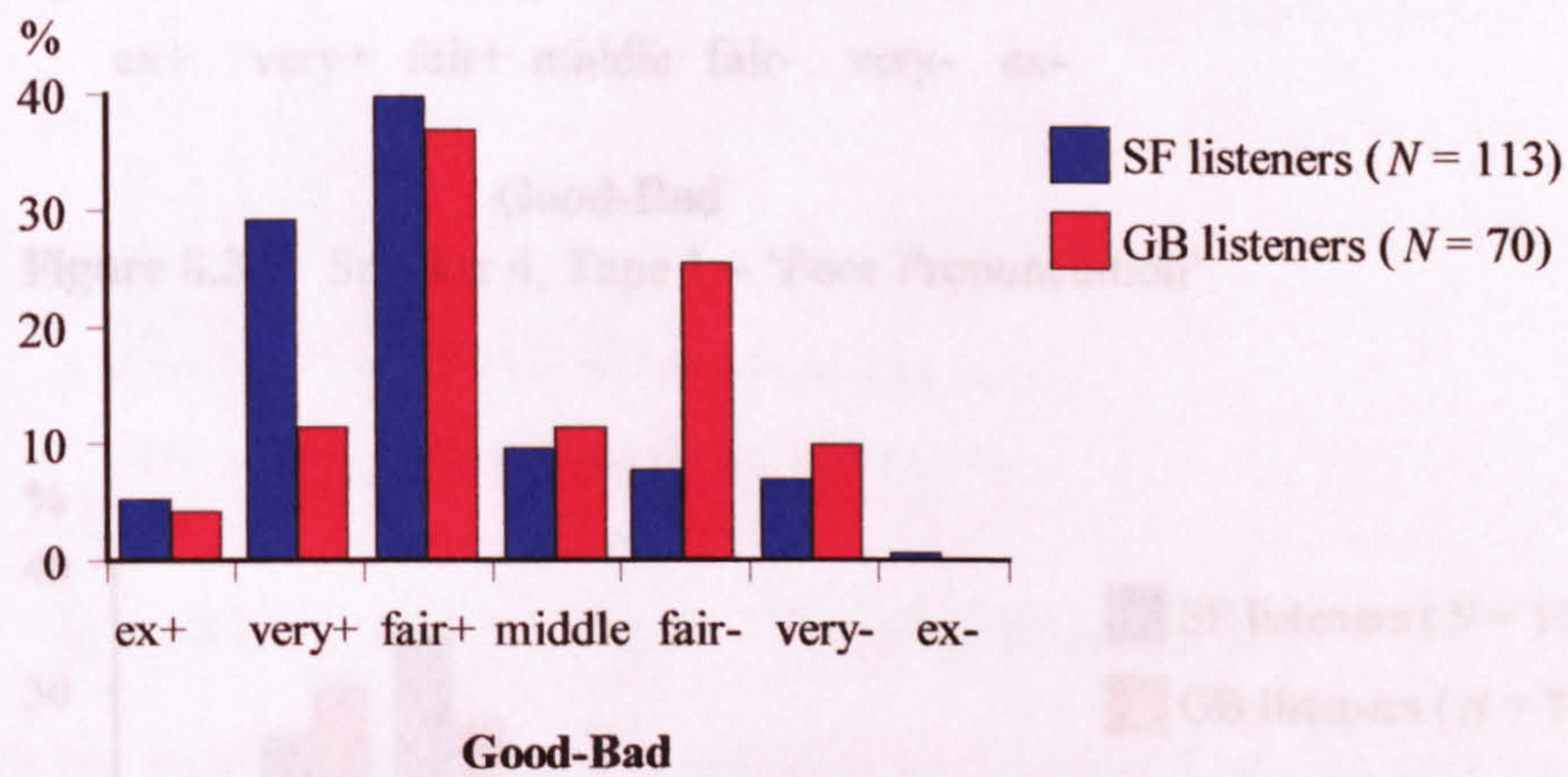


Figure 8.29 Speaker 3, Tape 1 – 'Good Pronunciation'

The evaluation of the Poor Pronunciation was approximately the same for both informant groups: Median Test and Mann-Whitney *U* test $p > .05$ (Figure 8.30). Neither the expected scores above and below the combined median, nor the dispersion of scores across the 7-point scale, are statistically significant.

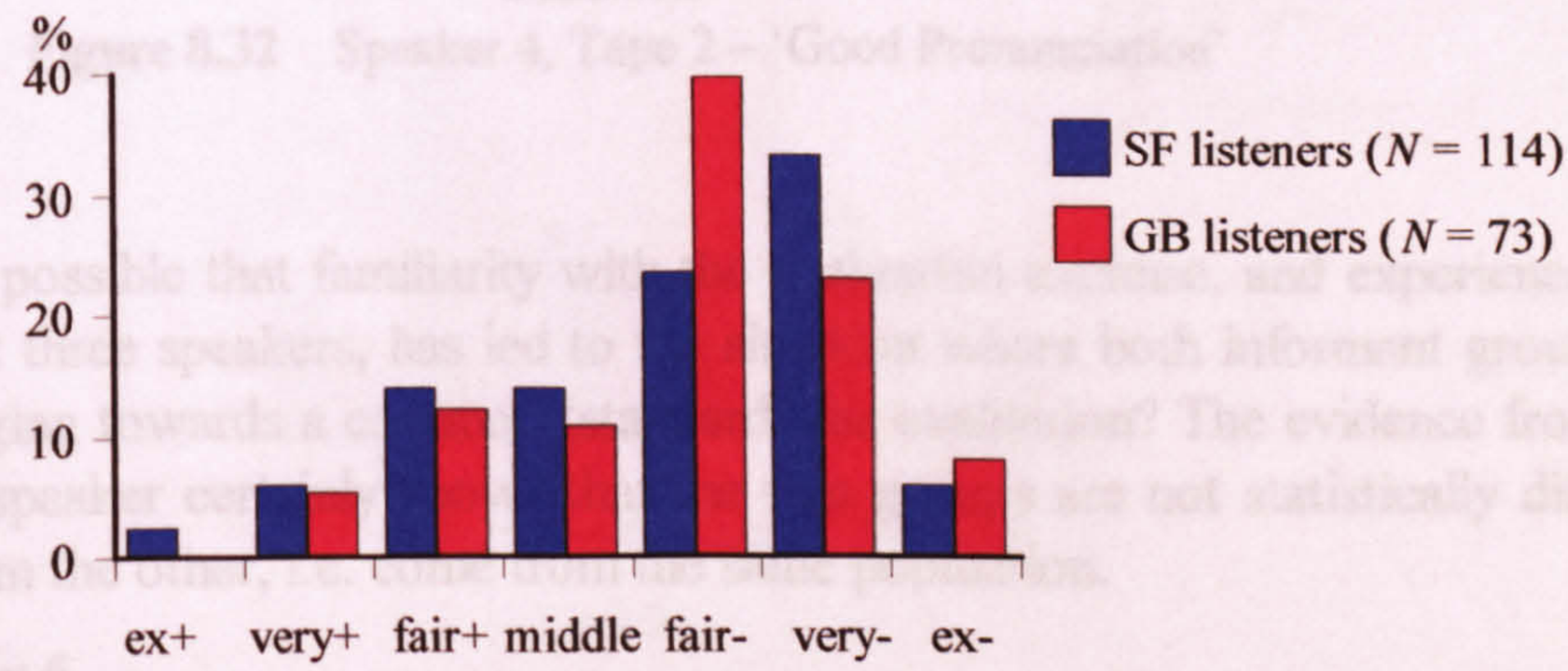


Figure 8.30 Speaker 3, Tape 2 – 'Poor Pronunciation'

Speaker 4

The fourth speaker (a woman) also produced a Good Pronunciation and a Poor Pronunciation, but this time the two groups of informants, Finnish-speaking listeners and English-speaking listeners, were basically identical in their evaluations of both versions: Median Test and Mann-Whitney *U* test $p > .05$ for both versions (Figure 8.31 and Figure 8.32, P. 213).

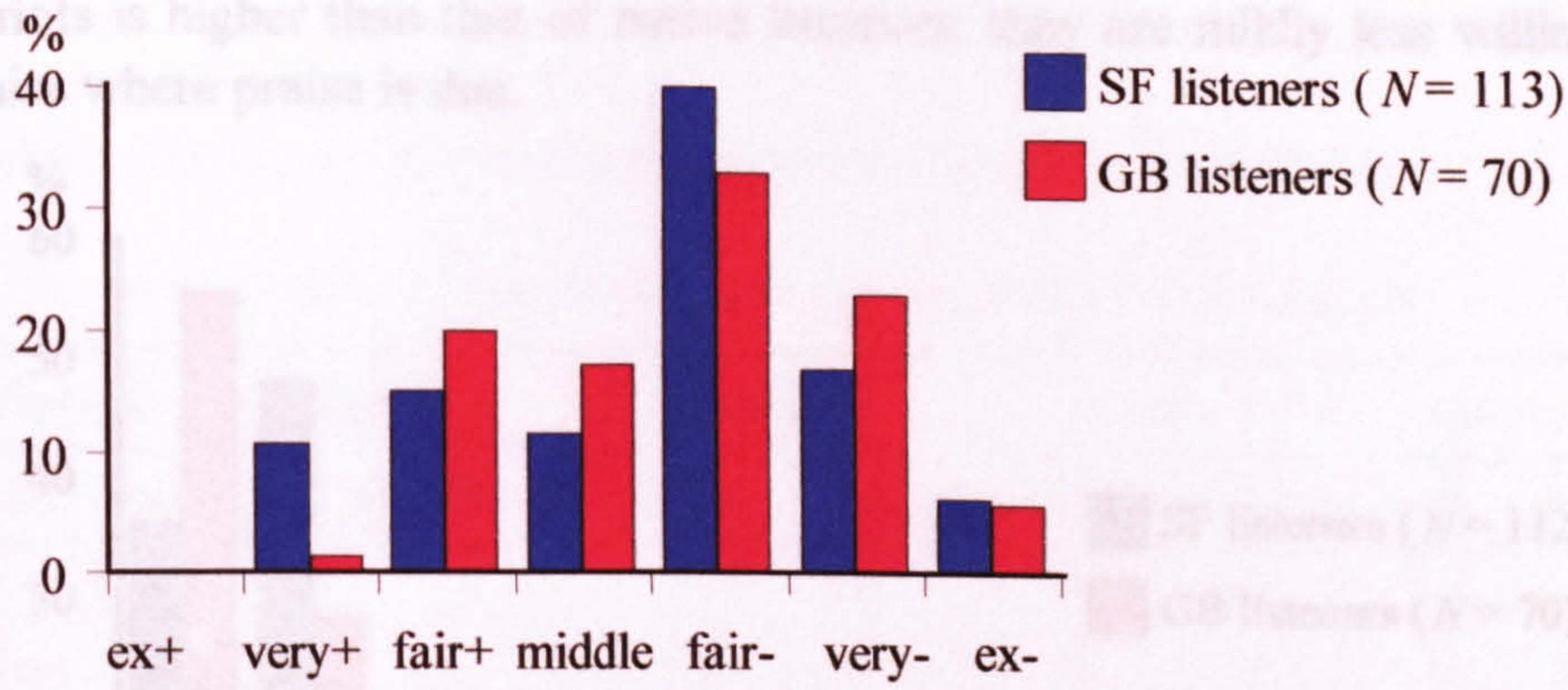


Figure 8.31 Speaker 4, Tape 1 – 'Poor Pronunciation'

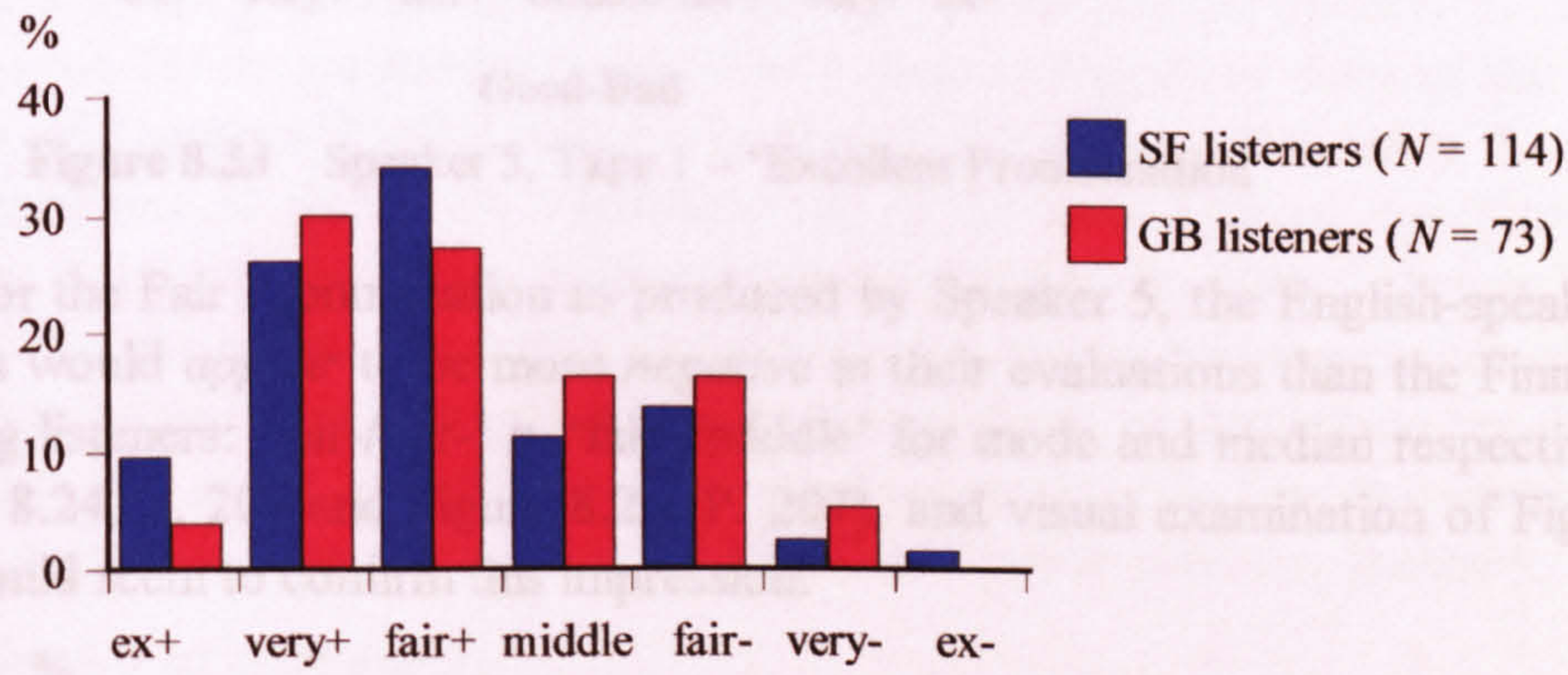


Figure 8.32 Speaker 4, Tape 2 – 'Good Pronunciation'

Is it possible that familiarity with the evaluation exercise, and experience with the first three speakers, has led to the situation where both informant groups are converging towards a common 'standard' for evaluation? The evidence from this fourth speaker certainly shows that the two groups are not statistically different one from the other, i.e. come from the same population.

Speaker 5

Speaker 5 offers the interesting situation where the between-group differences are different for each version heard.

For the Excellent Pronunciation the English-speaking listeners are more positive towards the speaker than the Finnish-speaking listeners (Figure 8.33, P. 214). Both the Median Test and the Mann-Whitney U test indicate the difference to be significant at the modest $p < .05$ level. Since the native Finnish listeners show themselves to be less tolerant than the native English listeners, it would seem that the standard of English pronunciation that Finns demand of their

compatriots is higher than that of native listeners: they are mildly less willing to give praise where praise is due.

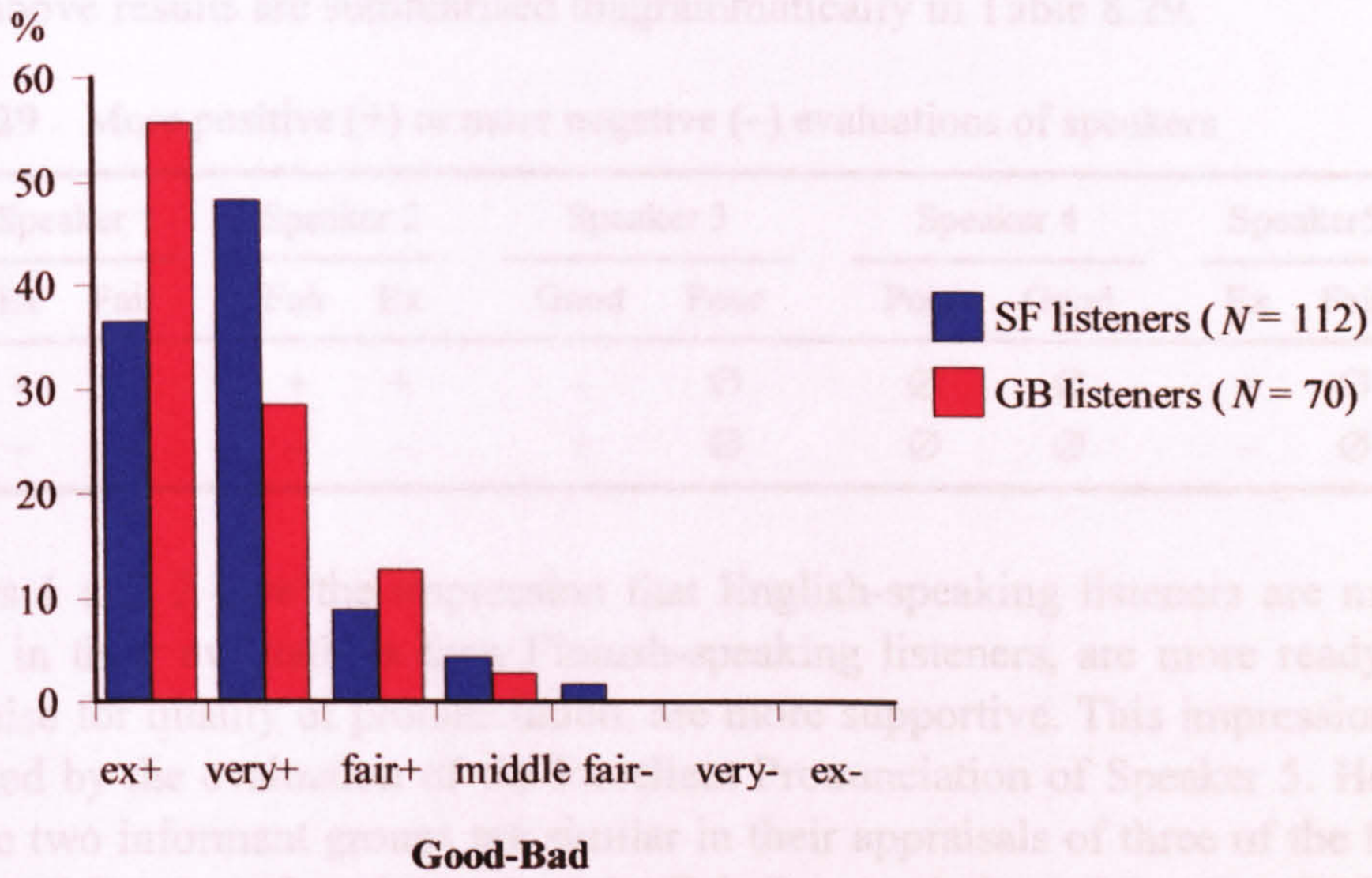


Figure 8.33 Speaker 5, Tape 1 – 'Excellent Pronunciation'

As for the Fair Pronunciation as produced by Speaker 5, the English-speaking listeners would *appear* to be more *negative* in their evaluations than the Finnish-speaking listeners: 'fair-/fair-' v. 'fair+/middle' for mode and median respectively (Figure 8.24, P. 208 and Figure 8.23, P. 207), and visual examination of Figure 8.34 would seem to confirm this impression.

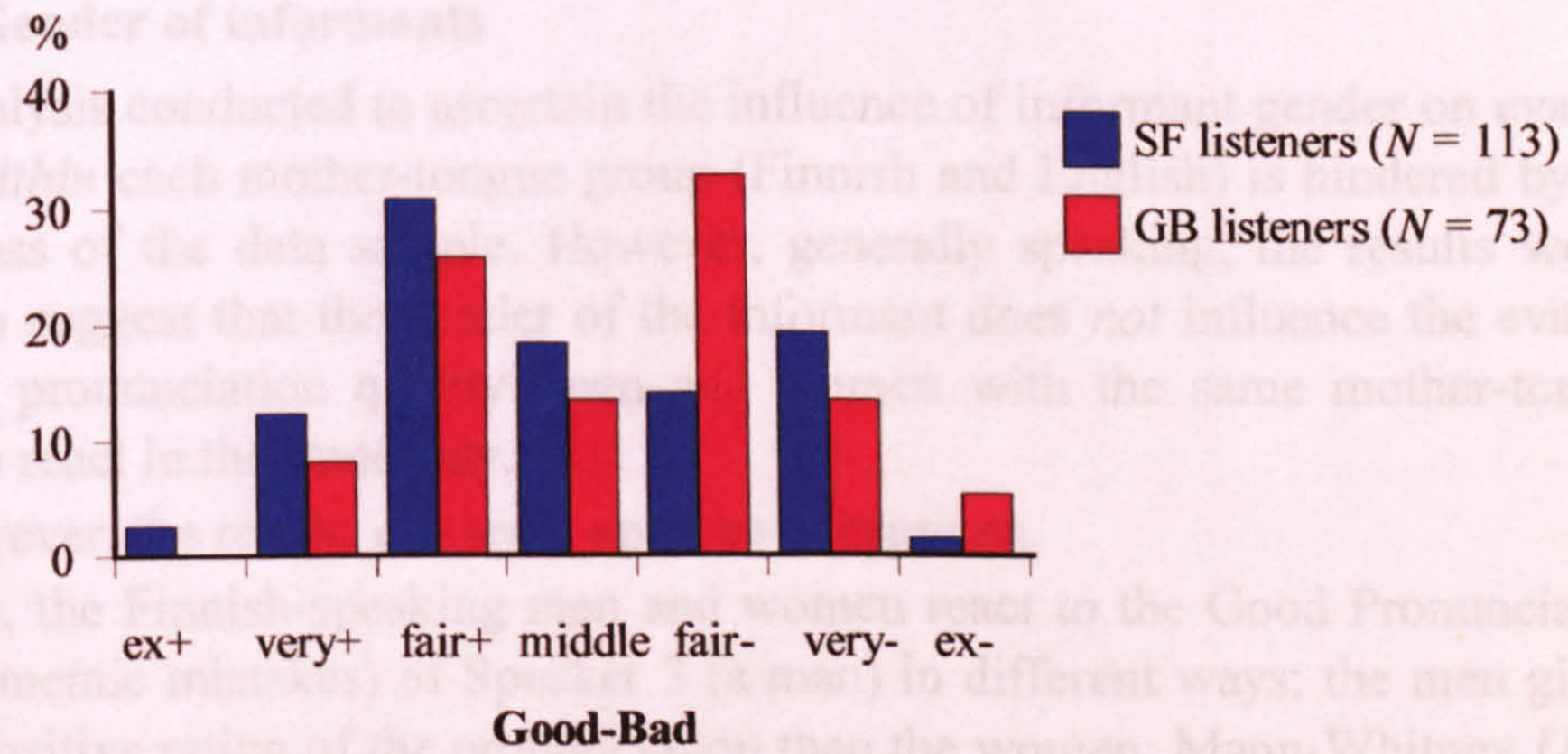


Figure 8.34 Speaker 5, Tape 2 – 'Fair Pronunciation'

However, *neither* the Median Test *nor* the Mann-Whitney *U* test confirm this impression: neither test indicates a significant difference between the evaluations of the two groups ($p > .05$).[3] On the basis of the modes and medians, the native English listeners would *seem* to be less tolerant than the native Finnish listeners,

[3] The Chi Square test for a cross-tabulation on mother-tongue groups gives a value of $p < .05$, but is unreliable as 4 of the 14 cells have expected counts less than 5, and one cell with zero falls below the expected count of 1.18.

downgrading mispronunciation where downgrading is 'deserved,' but *statistical tests do not confirm this impression.*

The above results are summarised diagrammatically in Table 8.29.

Table 8.29 More positive (+) or more negative (–) evaluations of speakers

	Speaker 1		Speaker 2		Speaker 3		Speaker 4		Speaker5	
	Ex	Fair	Fair	Ex	Good	Poor	Poor	Good	Ex	Fair
GB	+	+	+	+	–	Ø	Ø	Ø	+	Ø
SF	–	–	–	–	+	Ø	Ø	Ø	–	Ø

Speakers 1 and 2 give the impression that English-speaking listeners are more tolerant in their evaluations than Finnish-speaking listeners, are more ready to give praise for quality of pronunciation, are more supportive. This impression is confirmed by the evaluation of the Excellent Pronunciation of Speaker 5. However, the two informant groups are similar in their appraisals of three of the four versions of Speakers 3 and 4 and of the Fair Pronunciation of Speaker 5. Only the evaluations of the Good Pronunciation of Speaker 3 show the Finnish-speaking listeners being more positive than the English-speaking listeners. It is possible that the order of presentation of the voice guises, and the gender of the voices, are hidden variables here, but one could perhaps interpret these results as suggesting that as far as first impressions are concerned (the first two speakers) English-speaking listeners are more positive than Finnish-speaking listeners.

8.2.2. Gender of informants

The analysis conducted to ascertain the influence of informant gender on evaluations *within* each mother-tongue group (Finnish and English) is hindered by the smallness of the data sample. However, generally speaking, the results would seem to suggest that the gender of the informant does *not* influence the evaluation of pronunciation quality: men and women with the same mother-tongue seem to react in the same way.

However, the results do throw up three exceptions.

First, the Finnish-speaking men and women react to the Good Pronunciation (11 phonemic mistakes) of Speaker 3 (a man) in different ways: the men give a more positive rating of the pronunciation than the women: Mann-Whitney *U* test $p < .01$, and Median test $p < .05$. The men also tend to give a more positive rating than the women with the Poor Pronunciation (47 phonemic mistakes) of this same speaker, though the two statistical tests used give contradictory results: Median test $p < .05$, but Mann-Whitney *U* test $p > .05$. See Table 8.30, P. 216.

Second, the Finnish-speaking men and women do not seem to react to the Poor Pronunciation of Speaker 4 (a woman) in the same way, the men giving a slightly more positive rating than the women: Mann-Whitney *U* test $p < .05$, but not confirmed by the Median test, $p > .05$. English-speaking men and women also react in different ways to this same version of the same speaker (the Poor

Table 8.30 Within-groups gender contrasts in evaluations of quality of pronunciation

Speaker Version Gender	Tape	Women Mode	Women Median	Women *N*	Men Mode	Men Median	Men *N*	Mann-Whitney *U* test / Median test
				Finnish-speaking listeners				
3 – M – 'Good'	T1	fair+	fair+	59	fair+/very+ [a]	fair+	54	**/*
3 – M – 'Poor'	T2	very-	fair-	60	fair+	fair-	54	-/*
4 – W – 'Poor'	T1	fair-	fair-	59	fair-	fair-	54	*/-
				English-speaking listeners				
2 – M – 'Excellent'	T2	very+	very+	50	very+	fair+	23	*/-
4 – W – 'Poor'	T1	fair-	fair-	48	fair+	middle	22	***/**

[a] The result is bi-modal.

*** $p < .001$; ** $p < .01$; * $p < .05$.

Pronunciation of Speaker 4), the men also giving a more positive rating than the women: Mann-Whitney *U* test $p < .001$, Median test $p < .01$.

Third, English-speaking men and women would appear to differ as to the quality of Speaker 2's Excellent Pronunciation, this time the women tending to award a more positive rating than the men: Mann-Whitney *U* test $p < .05$, but unconfirmed by the Median test, $p > .05$.

Gender, therefore, seems to play little role *within* each mother-tongue group.

The situation is only slightly different when the influence of gender is analysed in conjunction with *between* mother-tongue groupings: Finnish-speaking women v. English-speaking women, and Finnish-speaking men v. English-speaking men (Table 8.31, P. 218). Here there are seven significantly contrasting results out of a possible total of twenty.[4] In six of these cases the native English listeners are more positive (more tolerant of mispronunciation) in their evaluations than the native Finnish listeners. The one exception (the last row of Table 8.31) involves Finnish-speaking men and English-speaking men listening to the Good Pronunciation of Speaker 3 (a man).

In brief, there are two sets of listeners here, men and women, and each set has ten groups for speaker-evaluation purposes (5 speakers x 2 versions each). Of the ten evaluations made by the groups of *women, five* reveal a difference between the Finnish-speaking listeners and the English-speaking listeners making up the group, the latter always being more positive than the former. Of the ten evaluations made by the groups of *men*, only *two* show a difference between the Finnish-speaking listeners and the English-speaking listeners, and of these, one (Speaker 3, Good Pronunciation) has the Finnish-speaking men being more positive than the English-speaking men, and one (Speaker 1, Fair Pronunciation) has the English-speaking men being more positive than the Finnish-speaking men.[5]

In other words the difference between the two mother-tongue groups of listeners, where it exists, would seem to be focused more in the women than in the men. English-speaking women are more tolerant, generous and supportive of unidentified foreign-accented speech than Finnish-speaking women are of compatriots speaking English. In other words, if a difference is being sought between Finnish listeners and English listeners, then it would be better to look for it among the women than among the men. A broad generalisation of the situation is presented diagrammatically in Figure 8.35, P. 219 ('SF' and 'GB' are again abbreviations for Finnish-speaking listeners and English-speaking listeners respectively).

---

[4] 5 speakers x 2 versions each x 2 groups for listener gender = 20.

[5] For six of the seven results here the significant values for Cramér's *V* (indicating the strength of association for the mother tongue of informants) range from .32 to .53, but not much can be made of this as the basic limitations of the Chi Square procedure are violated: some cells in the cross-tabulations equal zero, and some cells have values below 5.

Table 8.31 Between-groups gender contrasts in evaluations of quality of pronunciation

Speaker Gender	Version Tape	Finnish-speaking listeners Mode	Median	*N*	English-speaking listeners Mode	Median	*N*	Mann-Whitney *U* test / Median test
				Women listeners				
1 – W – 'Excellent'	T1	fair+	fair+	59	very+	very+	48	**/*
1 – W – 'Fair'	T2	very-	fair-	60	fair+	middle/fair+ [a]	50	***/**
2 – M – 'Excellent'	T2	fair+	fair+	60	very+	very+	50	***/***
2 – M – 'Fair'	T1	very-	very-	59	fair-	fair-	48	***/***
5 – W – 'Excellent'	T1	very+	very+	58	ex+	ex+	48	**/**
				Men listeners				
1 – W – 'Fair'	T2	very-/fair- [b]	fair-	54	very+	middle	23	*/*
3 – M – 'Good'	T1	fair+/very+ [b]	fair+	54	fair+	fair+	22	**/ –

[a] The median falls between the two values.

[b] The result is bi-modal.

*** $p < .001$; ** $p < .01$; * $p < .05$.

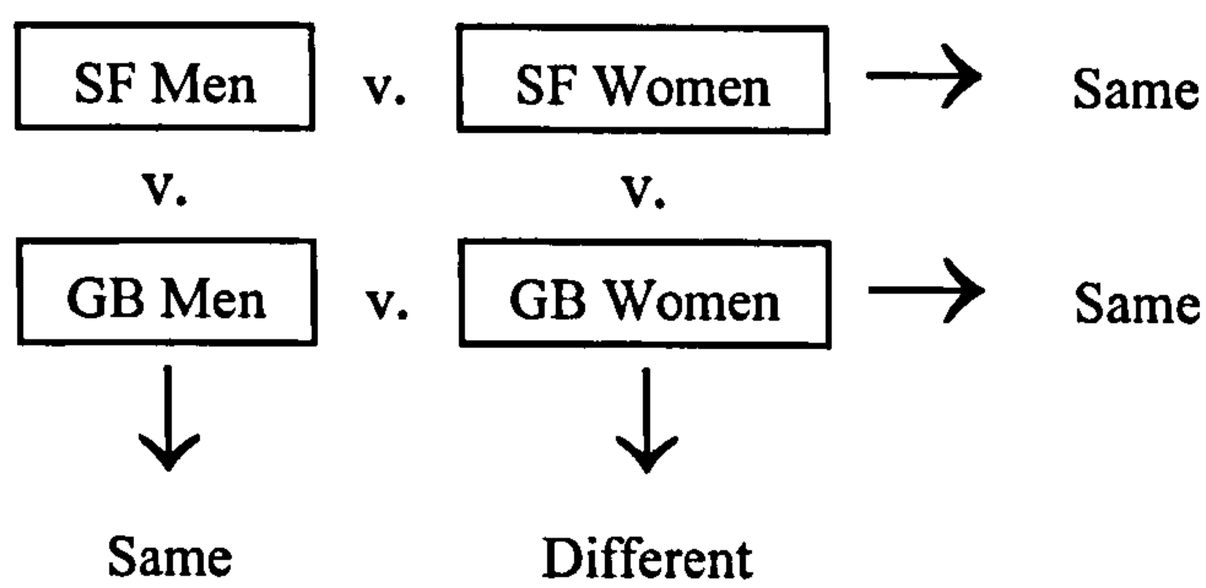


Figure 8.35 Probable contrasts in evaluations of quality of pronunciation made by informant groups selected on the basis of mother tongue and gender

8.2.3. Summary

The results reveal several important points.

First, it is clear that within the two groups of Finnish-speaking and English-speaking listeners, people generally recognise the standards of pronunciation of each individual speaker on the two tapes as being different, even when heard in isolation: the 'poorer pronunciations' are recognised as being weaker than those of the 'better pronunciations' even though they are not heard side by side. The Median Test and the Mann-Whitney *U*-test for two independent groups (Finnish-speaking listeners Tape 1 v. Finnish-speaking listeners Tape 2; English-speaking listeners Tape 1 v. English-speaking listeners Tape 2) reveal that the results of the evaluations of a speaker's version on Tape 1 are significantly different from those of the same speaker's second version on Tape 2. All the results are significant at the $p < .001$ level. This result in itself justifies the analysis of the data collected from the next question, Question 3: the real differences in the degree of accentedness of the speakers are noticeable and may therefore have a differential effect on perceptions of the fourteen traits of the speakers.

Second, it seems that everyone has an established internal personal standard by which to judge a speaker's pronunciation. Naturally one expects native-speakers of English to have a standard by which they can evaluate a foreigner's pronunciation of English: how closely does the pronunciation approximate some native accent of English? However, it appears that Finnish-speaking listeners also have some internal personal standard to refer to when listening to compatriots speaking English: the two groups of Finnish-speaking informants (those who heard one version of a speaker, and those who heard the other) evaluated all five speakers—even the first speaker, which has no preceding 'model' to serve as a point of reference—in different ways. A Finn's pronunciation of English is recognisably 'awful,' 'better,' 'much better,' and so on, and the range of evaluation would seem to suggest that accent is perceived atomistically, step by step, feature by feature (see section 4.5.1.3 above).

However, the result that Finns hear the versions the speakers produce as being variously 'better' and 'worse' should be interpreted with caution. The two groups

of listeners hear different tapes and they accurately identify the versions on their tapes as being different, but how listeners have understood the idea of 'better' and 'worse' here is impossible to elucidate. Have listeners evaluated the speakers in terms of 'more/less like a native speaker'? Or have they evaluated the speakers in terms of 'more/less like a Finnish speaker'? Or what? The version which is heard may be recognised as being less noticeably Finnish-accented, but paradoxically that does not necessarily mean that it is known to be more 'accurate,' more like a native-speaker's accent.

Third, the Finnish-speaking listeners usually differ from the English-speaking listeners in their evaluations of the same version of the same speaker (Finnish-speaking listeners Tape 1 v. English-speaking listeners Tape 1; Finnish-speaking listeners Tape 2 v. English-speaking listeners Tape 2). The Finns would seem to be less tolerant and more critical of their compatriots' standard of pronunciation (in-group/peer group disloyalty) than the native English listeners (Cf. the results of Fayer & Krasinksi 1987). They also seem to be more willing to make an immediate evaluation (perhaps because they recognise or assume the speakers to be fellow Finns), whereas the English-speaking listeners seem to be loath to commit themselves—at least with the first speaker—or to view the speakers negatively before a standard has become established (as is the case for the fourth and fifth speakers).

Finally there are two instances where groups of informants seem to have problems in evaluating a speaker, i.e. where the result is bi-modal and/or the median is 'middle.' The first involves the English-speaking listeners listening to the Fair Pronunciation of Speaker 1. The informants know the woman speaking on the tape is a foreigner, a non-native speaker, but the ambivalence and dispersion of responses is presumably a result of the lack of a bench mark and external standard for this first speaker: how bad must the mispronunciation be for a native-listener of English to call it bad? The second case involves the Finnish-speaking listeners evaluating the Fair Pronunciation of Speaker 5: the informants presumably recognise the woman they hear as being a Finn, and they now have four speakers with which to compare this last one, and yet the mode/median ends up as 'fair+'/'middle.' Why this speaker is not viewed more negatively is puzzling.

8.3. The fourteen traits

The third question consisted of fourteen 7-point scales with bi-polar adjectives marking the poles (Figure 8.36, P. 221).

> (N.B. The whole questionnaire with the full set of fourteen traits, both the English versions and the Finnish versions, is presented in Appendix D. The English–Finnish translation equivalents are presented side by side in Table 6.8, Chapter 6. In the following presentation and discussion of the results only the English names for the traits will be used even though the Finnish-speaking informants naturally filled in questionnaires which had the trait names in Finnish.)

Characteristics

Basing your judgement merely on the voice of the speaker, where would you place the speaker on the following 7-point scales?

educated	☐	☐	☐	☐	☐	☐	☐	uneducated
stupid	☐	☐	☐	☐	☐	☐	☐	intelligent
etc.								etc.

Figure 8.36

The aim of Question 3 was to discover whether the responses of the informants listening to Tape 1 and of the informants listening to Tape 2, when evaluating the fourteen traits, were the same or different. If different, then the null hypothesis that everyone forms the same impression of the speakers regardless of the level of mispronunciation they hear would have to be rejected, i.e. the different levels of mispronunciation of the five speakers on the two tapes do prompt different images.

Chi Square analysis was used to determine whether the observed distribution of responses was uniform across the 7-point nominal/ordinal scales for each of the fourteen traits involved, i.e. whether the associated probability of occurrence of the range of responses on each scale was greater than that of chance. The minimum level of significance used was the modest one of $\alpha < .05$.

The Mann-Whitney U test and the Median test for two independent groups were used to ascertain the level of significance of the contrast between the evaluations of Tape 1 and Tape 2, between the responses of the English-speaking listeners and the Finnish-speaking listeners, and between the men and the women within each language group, the minimum level again being set at the modest one of $\alpha < .05$.

Speaker 1

23-year-old woman
Tape 1: 0 mistakes 'Excellent Pronunciation'
Tape 2: 31 mistakes 'Fair Pronunciation'

All the evaluations of the 14 traits made by the two groups of Finnish-speaking listeners (the Excellent Pronunciation and the Fair Pronunciation) were non-uniform, and highly significantly so, at the $p < .001$ level (Chi Square test). Of the evaluations made by the English-speaking listeners, all except two evaluations were similarly non-uniform, and highly significantly so, at the $p < .001$ level. All evaluations which are not significant are marked in the Tables below as 'NA' (Not Applicable).

The general impression of Speaker 1 that is common to both the Finnish-speaking listener groups—the same impression being gained regardless of the version heard—is that she is: *educated, hardworking, intelligent, self-confident, friendly, honest, kind, reliable, serious* and *sympathetic* (Table 8.32).

Table 8.32 Evaluations of Finnish-speaking informants listening to Speaker 1

	Tape 1 [a]: 'Excellent Pronunciation'		Tape 2 [b]: 'Fair Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	fair+	fair+	fair-	fair-	***/***
Educated	very+	very+	fair+	fair+	***/***
Hardworking	very+	fair+	fair+	middle	**/**
Intelligent	very+	fair+	fair+	middle	***/***
Rich	fair+	middle	fair-	middle	***/***
Self-confident	very+	fair+	fair+	middle	**/**
Successful	fair+	fair+	fair-	middle	***/***
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	fair+	
Honest	very+	fair+	fair+	fair+	
Kind	fair+	fair+	fair+	fair+	
Reliable	very+	fair+	fair+	fair+	
Serious	very+	very+	very+	fair+	*/–
Sympathetic	fair+	fair+	fair+	middle	–/*
Trustworthy	fair+	fair+	fair-	middle	***/***

[a] $N = 113$ for all traits. [b] $N = 114$ for all traits except *ambitious* where $N = 113$.
*** $p < .001$; ** $p < .01$; * $p < .05$.

However, for six of these ten traits the evaluation leans in the same direction but varies in degree: the speaker is heard as being more *educated, hardworking, intelligent, self-confident, serious* and *sympathetic* when the pronunciation is Excellent than when it is Fair. (However, the Median test and the Mann-Whitney *U* test disconfirm each other for the last two, *serious* and *sympathetic.*)

For the remaining four traits still not mentioned, the Finnish-speaking listeners see Speaker 1 in a 'positive' or 'middle-but-tending-to-positive' light[6] when hearing the Excellent version (*ambitious, rich, successful, trustworthy*), but see her in a 'negative' or 'middle-but-tending-to-negative' light when hearing the Fair version (*unambitious, poor, unsuccessful, untrustworthy*).

It is striking that all the traits listed as being concerned with status/competence, but only three listed as relating to solidarity/benevolence (and two of these

[6] Where the mode and median are divided between a 'middle' and a 'fair' (+ or -) evaluation, I shall call this a 'middle-but-tending-to-positive/negative' result.

are borderline since the two statistical tests contradict each other), are evaluated significantly differently depending on which version the informants hear: broadly speaking the traits of status/competence vary between the two versions, whereas those for solidarity/benevolence are stable, thus appearing to be unaffected by the level of phonemic accuracy.

The two groups of English-speaking listeners (Tape 1, Tape 2) gain the same general impression of Speaker 1 regardless of which version they hear (Table 8.33). The English-speaking listeners consistently hear the speaker as being: *educated, hardworking, intelligent, rich* and *successful* (status/competence) and also *friendly, honest, kind, reliable, serious, sympathetic* and *trustworthy* (solidarity/benevolence).

Table 8.33 Evaluations of English-speaking informants listening to Speaker 1

	Tape 1 [a]: 'Excellent Pronunciation'		**Tape 2 [b]: 'Fair Pronunciation'**		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	NA	NA	fair-	middle	
Educated	very+	very+	very+	very+	–/*
Hardworking	very+	fair+	fair+	fair+	
Intelligent	very+	very+	very+	very+	–/*
Rich	fair+	fair+	fair+	middle	**/**
Self-confident	fair+	fair+	NA	NA	
Successful	very+	fair+	fair+	fair+	–/*
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	fair+	
Honest	very+	fair+	fair+	fair+	
Kind	fair+	fair+	very+	fair+	
Reliable	very+	fair+	fair+	fair+	
Serious	very+	very+	very+	very+	*/–
Sympathetic	fair+	fair+	very+	fair+	
Trustworthy	very+	fair+	fair+	fair+	

[a] $N = 70$ for all traits. [b] $N = 73$ for all traits.
** $p < .01$; * $p < .05$.

For five of these traits the evaluation of the speaker varies in degree depending on the version heard, but the significance level is borderline for four of them since only one or the other statistical test confirms it at the $p < .05$ level. According to the Mann-Whitney *U* test, the speaker is thought to be more *rich* and more *serious* when the Excellent Pronunciation is evaluated than when the Fair Pronunciation is evaluated; according to the Median test, the speaker is thought to be more *educated, intelligent* and *successful.*

The Excellent Pronunciation voice guise gives the impression that the speaker is *self-confident*, though listeners of the Fair Pronunciation fail to produce a result which is statistically significant. In similar fashion the Fair Pronunciation voice guise suggests the speaker is *unambitious* ('middle-but-tending-to-negative', the only negative evaluation of the speaker regardless of tape heard), though this time it is the listeners of the Excellent version who fail to yield a set of responses which is statistically significant. Presumably we can interpret these two last traits as being inappropriate for their respective versions.

When relying on the results of the Mann-Whitney *U* test alone, the difference in degree of phonemic accuracy of Speaker 1 would seem to have little effect on the evaluation the native English listeners make of this speaker: the image of the speaker remains almost exactly constant. The striking feature of the data is indeed the near congruity of responses across the two versions, or, put another way, the tolerance for the level of phonemic inaccuracy heard with the Fair Pronunciation. Of the seven status/competence traits, five are 'positive' for both versions, four equally so, one (*rich*) differentially so. Of the seven solidarity/benevolence traits, all are 'positive' for both versions, six equally so, only one (*serious*) differentially so. One possible reason for this uniformity (as already postulated for the bi-modal evaluation of the degree of pronunciation excellence for this speaker) may be the lack of an external standard for this first speaker.

However, when the perhaps less powerful Median test is used, the picture is slightly different, the congruity less marked. No traits of solidarity/benevolence are appraised differentially, though four traits of status/competence are: *educated, intelligent, rich* and *successful.* With the Median test there is perhaps a danger of committing a Type I error, i.e. of rejecting H_0 when in fact it is true.

Differences between the Finnish-speaking listeners and the English-speaking listeners in their perception of the speaker are not marked and involve mainly (but not only) the Fair Pronunciation version (Table 8.34, P. 225).

Both groups of listeners (Finnish-speaking listeners and English-speaking listeners) are in broad agreement with each other when the pronunciation is Excellent, but they differ from each other when the pronunciation is Fair. Where differences exist, they all—with the exception of *self-confident*—reveal the Finnish-speaking listeners as being significantly more negative in their evaluations of the speaker than the English-speaking listeners: the latter are more tolerant of the phonemic inaccuracies. Since the data for *self-confident* from the English-speaking listeners is not significantly non-uniform, the discrepancy between the native English listeners and the native Finnish listeners that is revealed here by the Median test is perhaps best ignored.

In brief, the impression that the Finnish-speaking listeners have of this female speaker depends on the version they hear: they upgrade all seven status/competence traits (plus *trustworthy*, and marginally *serious* and *sympathetic*) with the Excellent Pronunciation and downgrade them with the Fair Pronunciation (Table 8.32). The English-speaking listeners, on the other hand, have broadly the same impression of the speaker—apart from the *rich* trait and four borderline traits—regardless of the person's phonemic accuracy (Table 8.33). The English-speaking

Table 8.34 Significant contrasts on various traits of Speaker 1

Trait	Finnish-speaking listeners		English-speaking listeners		Mann-Whitney *U* test / Median test	GB listeners +/-[a]
	Mode	Median	Mode	Median		
	Tape 1: 'Excellent Pronunciation'					
	(*N* = 113)		(*N* = 70)			
Status/Competence						
Educated	very+	very+	very+	very+	*/**	+
Intelligent	very+	fair+	very+	very+	***/**	+
Rich	fair+	middle	fair+	fair+	***/***	+
	Tape 2: 'Fair Pronunciation'					
	(*N* = 114[b])		(*N* = 73)			
Status/Competence						
Ambitious	fair-	fair-	fair-	middle	**/**	+
Educated	fair+	fair+	very+	very+	***/***	+
Hardworking	fair+	middle	fair+	fair+	*/–	+
Intelligent	fair+	middle	very+	very+	***/***	+
Rich	fair-	middle	fair+	middle	***/***	+
Self-confident	fair+	middle	NA	NA	–/*	NA
Successful	fair-	middle	fair+	fair+	***/***	+
Solidarity/Benevolence						
Sympathetic	fair+	middle	very+	fair+	***/***	+
Trustworthy	fair-	middle	fair+	fair+	***/**	+

[a] The English-speaking listeners' perceptions of the speaker are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. [b] For *ambitious* $N = 113$.
*** $p < .001$; ** $p < .01$; * $p < .05$.

listeners would appear to be hesitant in judging this first speaker either negatively or positively, apparently using some general sort of bland standard which results in the evaluations of the two versions being broadly identical. Only with the evaluation of the second speaker—if I may anticipate matters—will the English-speaking listeners be prepared to commit themselves one way or the other, for then they can exploit the first speaker as a point of comparison. The Finnish-speaking listeners are not so reluctant to rate this first speaker, perhaps using *their own* (mis-)pronunciation of English as the bench mark.

As for differences in evaluation between the Finnish-speaking listeners and the English-speaking listeners (Table 8.34), these differences focus mainly—but not only—on the evaluation of the status/competence traits of the Fair Pronunciation version: the native Finnish listeners are more negative, more critical, of the phonemic inaccuracy of their compatriot than the more tolerant native English

listeners. Both language group listeners perceive all the solidarity/benevolence characteristics of the speaker in exactly the same way when she makes no phonemic mistakes.

Speaker 2

26.1-year-old man		
Tape 1:	31 mistakes	'Fair Pronunciation'
Tape 2:	0 mistakes	'Excellent Pronunciation'

All the evaluations of the 14 traits made by the two groups of Finnish-speaking listeners (Fair Pronunciation and Excellent Pronunciation) were non-uniform, and highly significant at the $p < .001$ level (Chi Square test). Of the evaluations of the 14 traits made by English-speaking informants all except three were similarly non-uniform at the $p < .001$ level: two results were only significant at the $p < .01$ level (Excellent Pronunciation: *ambitious* and *trustworthy*), but they will be analysed further together with the others since they pass the minimum $\alpha < .05$ threshold; one result (Fair Pronunciation: *successful*, $p > .05$) failed to reach the set minimum level of significance and will be marked 'NA' in the following tables.

The general impression of Speaker 2 common to all the Finnish-speaking listeners—regardless of the version heard—is that he is: *educated, friendly, honest, kind, reliable* and *serious* (Table 8.35, P. 227).

However, while three of these six traits (*friendly, honest* and *kind*) are evaluated similarly for the two versions (neither the Mann-Whitney *U* test nor the Median test revealing any significant difference between the two informant groups: $p > .05$), the other three traits (*educated, reliable* and *serious*) lean more in the 'positive' direction when the Excellent Pronunciation is evaluated than when the Fair Pronunciation is evaluated. (Note, however, that the Median test does not confirm the differential evaluation for *serious*.)

One trait (*rich–poor*) prompted a 'middle' (but statistically significant) response for the Fair Pronunciation, but is evaluated as mildly *rich* (fair+) for the Excellent Pronunciation.

Seven further traits are seen in either a 'negative' or a 'positive' light depending on the version heard: the Fair Pronunciation prompts the native Finnish listeners to see the speaker as *unambitious, lazy, unintelligent, hesitant/unsure, unsuccessful, unsympathetic* and *untrustworthy*; but the Excellent Pronunciation prompts them to see him as the opposite, i.e. as *ambitious, hardworking, intelligent, self-confident, successful, sympathetic* and *trustworthy*. The negative evaluations of the Fair Pronunciation seem to be more concerned—though not exclusively—with work values than with social values.

The fact that no difference is made for *friendly* and *kind* is surprising in that a difference is made for *sympathetic*. At first blush one might expect the three traits to be evaluated similarly, but this is not the case for this male speaker.

Table 8.35 Evaluation of Finnish-speaking informants listening to Speaker 2

	Tape 1 [a]: 'Fair Pronunciation'		Tape 2 [b]: 'Excellent Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	fair-	fair-	very+	fair+	***/***
Educated	fair+	middle	very+	very+	***/***
Hardworking	fair-	fair-	fair+	fair+	***/***
Intelligent	fair-	middle	fair+	fair+	***/***
Rich	middle	middle	fair+	fair+	***/***
Self-confident	fair-	middle	very+	fair+	***/***
Successful	fair-	middle	fair+	fair+	***/**
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	fair+	
Honest	fair+	fair+	fair+	fair+	
Kind	fair+	fair+	fair+	fair+	
Reliable	fair+	middle	fair+	fair+	**/*
Serious	fair+	middle	fair+	fair+	*/–
Sympathetic	fair-	middle	fair+	fair+	***/*
Trustworthy	fair-	fair-	fair+	fair+	***/***

[a] $N = 113$ for all traits. [b] $N = 114$ for all traits.
*** $p < .001$; ** $p < .01$; * $p < .05$.

Unlike the results for Speaker 1 (a woman), the results here for Speaker 2 (a man) do not fall quite so clearly into two groups depending on the version of the speaker being evaluated: all seven status/competence traits are differentially evaluated (as was the case with Speaker 1), but in addition, so are four of the seven solidarity/benevolence traits. The level of phonemic accuracy clearly affects the evaluation of Speaker 2 to a greater extent than it does for Speaker 1. One can only surmise that this is a result of the gender of the speaker, or else a result of the informants' greater self-confidence in performing the task in hand.

The general impression of Speaker 2 gained by the two groups of English-speaking listeners—regardless of which version is heard—is that he is: *educated, intelligent, friendly* and *serious* (Table 8.36, P. 228). For two of these traits (*friendly* and *serious*) there is no difference in the evaluation of the two groups (Mann-Whitney *U* test and Median test, $p > .05$); for the other two traits (*educated* and *intelligent*) the evaluation leaned more in the positive direction when the informants heard the Excellent Pronunciation than when they heard the Fair Pronunciation (Mann-Whitney *U* test and Median test, $p < .001$).

One trait (*kind–cruel*) prompts a 'middle' response from both groups of informants (Chi Square $p < .001$ for both groups) so presumably the level of pho-

Table 8.36 Evaluations of English-speaking informants listening to Speaker 2

	Tape 1 [a]: 'Fair Pronunciation'		Tape 2 [b]: 'Excellent Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	fair-	middle	fair+	fair+	**/*
Educated	fair+	middle	very+	very+	***/***
Hardworking	fair-	middle	fair+	fair+	*/–
Intelligent	fair+	fair+	very+	very+	***/***
Rich	middle	middle	fair+	fair+	***/***
Self-confident	fair-	fair-	very+	very+	***/***
Successful	NA	NA	very+	fair+	***/***
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	middle	
Honest	fair-/+ [c]	middle	middle	fair+	*/–
Kind	middle	middle	middle	middle	
Reliable	fair-	middle	fair+	fair+	*/*
Serious	fair+	fair+	fair+	fair+	
Sympathetic	fair+	middle	fair-	fair-	
Trustworthy	middle	middle	fair+	fair+	–/*

[a] $N = 70$ for all traits. [b] $N = 73$ for all traits. [c] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

nemic accuracy (together with the rhythm, volume, voice quality, etc.) fails to create any impression on this dimension.

Two further traits (*rich–poor* and *trustworthy–untrustworthy*) also prompt a 'middle' response from the informants who heard the Fair Pronunciation, though the informants who heard the Excellent Pronunciation clearly lean moderately towards the positive end of the scale. Both the Mann-Whitney *U* test and the Median test confirm a significant difference here for *rich* ($p < .001$), but only the Median test reveals a significant difference for *trustworthy* ($p < .05$).

Another trait (*honest*) is perhaps best interpreted as being 'middle' for the informants who heard the Fair Pronunciation, since the median is 'middle' and the mode is bi-modal, split between 'fair+' and 'fair-' (the result is in fact non-significant according to the Median test: $p > .05$). For the informants who heard the Excellent Pronunciation, on the other hand, the *honest* trait prompts a 'middle-but-tending-to-positive' response.

One trait (*successful*) fails to elicit a statistically significant result ('NA': Chi Square $p > .05$) with the informants who heard the Fair Pronunciation, though the result is clearly positive with those who heard the Excellent Pronunciation. The contrast between the two is significant: Mann-Whitney *U* test and Median test, $p < .001$.

Thus, in one way or another, five traits (*rich* and *successful*; *honest*, *kind* and *trustworthy*) fail to produce results which swing to one end of the scale or the other when informants hear the Fair Pronunciation, and one trait (*kind–cruel*) is similarly 'middle' with those who heard the Excellent Pronunciation. It is as if the level of phonemic inaccuracy limits the English-speaking listeners' ability to form a clear impression one way or the other of the speaker; the listeners are non-committal.

As for the five remaining traits, the English-speaking listeners see Speaker 1 in a positive or negative light depending on which version they hear: for those who heard the Excellent Pronunciation the speaker is *ambitious*, *hardworking*, *self-confident*, *reliable* and—surprisingly—*unsympathetic*, but for those who heard the Fair Pronunciation he is *unambitious*, *lazy*, *hesitant/unsure*, *unreliable* and—surprisingly—*sympathetic*. Why Excellent Pronunciation (no phonemic mistakes) should prompt an *unsympathetic* image while mild phonemic inaccuracy (Fair Pronunciation) should conversely prompt a *sympathetic* impression is a puzzle one can only speculate about: is it the upgrading of the status/competence traits which leads to a corresponding (and inevitable?) downgrading of this social trait?

It is striking that of the ten traits which are evaluated statistically differently by the two groups of English-speaking listeners seven involve traits of status/competence and only three involve traits of solidarity/benevolence (and two of these are borderline with the two statistical tests contradicting each other). This clustering on the status/competence traits which is found here with this male speaker is remarkably similar to the clustering found with the Finnish-speaking listeners with the female speaker (Table 8.32): the positive and negative leanings of the individual traits may not be the same, but it is interesting that it is the status/competence traits (more so than the solidarity/benevolence traits) that are perceived differentially depending on which version of the respective speaker is heard, Fair Pronunciation or Excellent Pronunciation.

Differences between the Finnish-speaking listeners and the English-speaking listeners in their impressions of Speaker 2 embrace pretty evenly not only both Tape 1 and Tape 2 (seven traits and nine traits respectively) but also the status/competence traits and the solidarity/benevolence traits (Table 8.37, P. 230). There are several slight discrepancies here between the results of the Mann-Whitney *U* test and the Median test, but I shall rely more heavily on the former than on the latter, though by doing so there is a danger of my committing a Type I error with *kind*, *reliable* and *honest*, i.e. a danger of rejecting H_0 when in fact it is true.

When the Fair Pronunciation was evaluated, the speaker was heard as being *unambitious* and *lazy* by all, but significantly more so by the native Finnish listeners than by the native English listeners. The speaker was also thought to be *serious* by all, but significantly more so by the English-speaking listeners than by the Finnish-speaking listeners. He was felt to be *unintelligent* but *reliable* by the Finnish-speaking listeners, but the reverse, *intelligent* and *unreliable*, by the English-speaking listeners (the Median test does not support this interpretation

Table 8.37 Significant contrasts on various traits of Speaker 2

Trait	Finnish-speaking listeners		English-speaking listeners		Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Mode	Median	Mode	Median		
	Tape 1: 'Fair Pronunciation'					
	(*N* = 113)		(*N* = 70)			
Status/Competence						
Ambitious	fair-	fair-	fair-	middle	**/***	+
Hardworking	fair-	fair-	fair-	middle	***/***	+
Intelligent	fair-	middle	fair+	fair+	**/**	+
Solidarity/Benevolence						
Honest	fair+	fair+	fair-/+ [b]	middle	***/*	-
Kind	fair+	fair+	middle	middle	*/–	-
Reliable	fair+	middle	fair-	middle	*/–	-
Serious	fair+	middle	fair+	fair+	***/**	+
	Tape 2: 'Excellent Pronunciation'					
	(*N* = 114)		(*N* = 73)			
Status/Competence						
Educated	very+	very+	very+	very+	***/***	+
Intelligent	fair+	fair+	very+	very+	***/***	+
Rich	fair+	fair+	fair+	fair+	**/***	+
Self-confident	very+	fair+	very+	very+	**/**	+
Solidarity/Benevolence						
Friendly	fair+	fair+	fair+	middle	***/*	-
Honest	fair+	fair+	middle	fair+	*/–	-
Kind	fair+	fair+	middle	middle	*/–	-
Serious	fair+	fair+	fair+	fair+	**/*	+
Sympathetic	fair+	fair+	fair-	fair-	***/*	-

[a] The English-speaking listeners' perceptions of the speaker are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

for *reliable*). Finally he was considered to be *honest* and *kind* by the Finns, but these traits prompted only a 'middle' response from the English-speaking listeners (the Median test sees no contrast between the two for *kind*). Broadly, as seen from the last right-hand column, the native Finnish listeners would seem to upgrade solidarity/benevolence traits and the native English listeners status/competence traits relative to each other.

When the Excellent Pronunciation was evaluated, the speaker was heard by both Finnish-speaking listeners and English-speaking listeners as being *educated, intelligent, rich, self-confident, friendly, honest* and *serious*. However, he

appeared significantly more *educated, intelligent, rich, self-confident* and *serious* for the English-speaking listeners, and significantly more *friendly* and *honest* for the Finns (the Median test fails to confirm this result for *honest*). The Finnish-speaking listeners also thought the speaker to be *kind* and *sympathetic*, whereas the English-speaking listeners were 'middle' on the *kind–cruel* trait (a statistically non-significant difference according to the Median test) and considered the speaker to be *unsympathetic* (a statistically significant difference). In other words, for the Excellent Pronunciation, Finnish-speaking listeners tended to upgrade solidarity/benevolence, but English-speaking listeners tended to upgrade status/competence (see the last right-hand column of Table 8.37).

To summarise, the Finnish-speaking listeners clearly have a more negative image of this male speaker when evaluating the Fair Pronunciation than when evaluating the Excellent Pronunciation (Table 8.35): seven traits tend (to a statistically significant degree) not only less towards the positive pole for the Fair Pronunciation than for the Excellent Pronunciation, but more importantly they find themselves on the negative end of the scale. The English-speaking listeners also have a slightly more negative impression of this man—generally speaking—when evaluating the Fair Pronunciation than when evaluating the Excellent Pronunciation (Table 8.36), though the difference is perhaps not quite so marked as with the Finnish-speaking listeners. It is also striking that the traits which reveal this difference in the evaluations of Speaker 2 are mainly status/competence traits (only three solidarity/ benevolence traits—viz. *honest, reliable* and *trustworthy*—display results which contrast significantly between Tape 1 and Tape 2, but then only at the $p < .05$ level). The pattern uncannily resembles the pattern found with the Finnish-speaking listeners evaluating Speaker 1 (Table 8.32): there the native Finnish listeners—listening to a woman, remember—evaluated the status/competence traits differentially; here the native English listeners—listening to a man, notice—also evaluate the status/competence traits differentially. In both within language-group situations separately (Finnish-speaking informants listening to female Speaker 1, and English-speaking informants listening to male Speaker 2), one can say that broadly speaking, apart from a couple of exceptions, the solidarity/benevolence traits earn the same evaluation from the respective groups of listeners regardless of the level of phonemic accuracy.

As for the between language-group situation, contrasting the reactions of Finnish-speaking listeners with those of English-speaking listeners, there is a range of traits for both versions on which the two sets differ, the English-speaking listeners generally being more positive than the Finnish-speaking listeners for the status/competence traits, and more negative for the solidarity/benevolence traits. The English-speaking listeners would appear to be upgrading the speaker's work image (expertise and proficiency), whereas the Finnish-speaking listeners seem to be showing in-group support for his social image.

Speaker 3

26.5-year-old man		
Tape 1:	11 mistakes	'Good Pronunciation'
Tape 2:	47 mistakes	'Poor Pronunciation'

All the evaluations of the 14 traits made by the two groups of both Finnish-speaking and English-speaking listeners (Good Pronunciation and Poor Pronunciation) were non-uniform, and highly significantly so at the $p < .001$ level (Chi Square test). Unlike the data for Speakers 1 and 2, therefore, no evaluations are marked in the following Tables as 'NA' (Not Applicable). The random dispersion of responses that characterised informants' reactions to a couple of the traits for the first two speakers is no longer to be found.

The Finnish-speaking listeners have the impression that Speaker 3—regardless of the version they hear—is: *hardworking, intelligent, friendly, honest, kind, reliable, serious, sympathetic* and *trustworthy* (Table 8.38). For five of these nine traits (*friendly, honest, kind, reliable* and *serious*) the evaluations of the two versions are similar, neither the Mann-Whitney *U* test nor the Median test showing any significant difference between the two groups ($p > .05$). However, the

Table 8.38 Evaluations of Finnish-speaking informants listening to Speaker 3

	Tape 1 [a]: 'Good Pronunciation'		Tape 2 [b]: 'Poor Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	fair+	fair+	fair-	fair-	***/***
Educated	very+	very+	fair-	middle	***/***
Hardworking	very+	very+	fair+	fair+	***/**
Intelligent	very+	fair+	fair+	fair+	***/***
Rich	fair+	fair+	middle	middle	***/***
Self-confident	very+	fair+	fair-	fair-	***/***
Successful	fair+	fair+	fair-	middle	***/***
Solidarity/Benevolence					
Friendly	very+	very+	fair+	fair+	
Honest	very+	very+	fair+	fair+	
Kind	fair+	fair+	fair+	fair+	
Reliable	very+	very+	fair+	fair+	
Serious	fair+	fair+	fair+	fair+	
Sympathetic	very+	fair+	fair+	fair+	**/***
Trustworthy	very+	fair+	fair+	fair+	*/**

[a] $N = 113$ for all traits. [b] $N = 114$ for all traits.
*** $p < .001$; ** $p < .01$; * $p < .05$.

other four traits (*hardworking*, *intelligent*, *sympathetic* and *trustworthy*) lean more towards the 'positive' pole of the scale (and significantly so) when the Good Pronunciation is evaluated than when the Poor Pronunciation is evaluated.

One trait (*rich–poor*) yielded a 'middle' (but statistically significant) response for the Poor Pronunciation, but a moderately positive response ('fair+') for the Good Pronunciation, a result which would seem to support the evidence provided by the evaluation of Speaker 2 and suggest that a Fair or Poor Pronunciation offers Finnish-speaking listeners little clue as to the supposed wealth of a male speaker. The Fair Pronunciation of Speaker 1 (a woman), it should be noted, was evaluated as being 'middle-but-tending-to-negative,' a woman perhaps therefore being easier to classify for wealth than a man.

Four further traits—all concerned with status/competence—are evaluated either in a positive light or in a negative light depending on the version heard: the Good Pronunciation suggests that this male speaker (Speaker 3) is *ambitious*, *educated*, *self-confident* and *successful*, but the Poor Pronunciation suggests the speaker is *unambitious*, *uneducated*, *hesitant/unsure* and *unsuccessful*.

It is again striking that all the work value traits, but only two of the social value traits, are differentially evaluated depending on the version heard. In other words, the level of phonemic accuracy affects the impression of status/competence but has little effect on the impression of solidarity/benevolence.

The general impression of Speaker 3 recorded by the English-speaking listeners (for both versions of the speaker that are evaluated) is very similar to that of the Finnish-speaking listeners here, though there are a couple of important differences (Table 8.39, P. 234).

The speaker is heard as being not only *hardworking*, *intelligent*, *friendly*, *honest*, *kind*, *reliable*, *serious*, *sympathetic* and *trustworthy* but also (unlike the Finnish-speaking listeners) *educated* and *successful*. For seven of these traits (*hardworking*, *friendly*, *honest*, *kind*, *serious*, *sympathetic* and *trustworthy*) the evaluations of the Good Pronunciation and the Poor Pronunciation are similar: there is no significant difference between the two groups ($p > .05$). However, four traits (*educated*, *intelligent*, *successful* and *reliable*) are evaluated more positively with the Good Pronunciation than with the Poor Pronunciation.

Two traits lean in different directions depending on the version heard: the Good Pronunciation suggests the speaker is *ambitious* and *self-confident*, but the Poor Pronunciation suggests he is *unambitious* and *hesitant/unsure*.

As for the matter of wealth, that trait remains inscrutable for the Poor Pronunciation. Notice, however, that the wealth trait is only mildly positive for the Good Pronunciation: 'middle-tending-to-positive.' Thus judging from the data for the first three speakers, it would seem that as far as the English-speaking listeners are concerned the higher the number of phonemic errors, the more opaque the trait of wealth.

The impression afforded the English-speaking listeners by the Good Pronunciation is positive, and by the Poor Pronunciation (apart from the *unambitious* and *hesitant/unsure* traits) broadly so: there is broad congruity here in the leaning towards the positive ends of the scales. However, as with Speaker 2 (also a man),

Table 8.39 Evaluations of English-speaking informants listening to Speaker 3

	Tape 1 [a]: 'Good Pronunciation'		Tape 2 [b]: 'Poor Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	fair+	fair+	fair-	middle	***/***
Educated	fair+/very+ [c]	fair+/very+ [d]	fair+	fair+	***/***
Hardworking	fair+	fair+	fair+	fair+	
Intelligent	very+	very+	fair+	fair+	***/**
Rich	middle	fair+	middle	middle	***/***
Self-confident	fair+	fair+	fair-	fair-	***/***
Successful	fair+	fair+	fair+	middle	***/**
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	fair+	
Honest	very+	fair+	fair+	fair+	
Kind	very+	fair+	very+	fair+	
Reliable	very+	fair+	fair+	fair+	**/*
Serious	very+	fair+	fair+	fair+	
Sympathetic	very+	fair+	fair+/very+ [c]	fair+	
Trustworthy	fair+	fair+	fair+	fair+	

[a] $N = 70$ for all traits. [b] $N = 73$ for all traits. [c] The result is bi-modal. [d] The median falls between the two values given.
*** $p < .001$; ** $p < .01$; * $p < .05$.

it is clear that the traits that are significantly more positive in the evaluation of the better of the two versions are almost exclusively those traits concerned with status/competence. The solidarity/benevolence traits—apart from *reliable*—remain unaffected by the level of phonemic accuracy.

As regards the differences between the Finnish-speaking listeners and the English-speaking listeners in their perception of Speaker 3, it is striking that there are no statistically significant differences between the two sets of informants for the Good Pronunciation: they both see the speaker in the same way (Table 8.40, P. 235). Even for the Poor Pronunciation differences are limited to just three of the fourteen traits: *friendly*, *reliable* and *sympathetic*—all traits of solidarity/benevolence. For *friendly* and *reliable* the Finnish-speaking listeners are more positive than the English-speaking listeners (in-group support), whereas for *sympathetic* the reverse is the case. Such broad congruity in the responses of the two groups of informants comes as a surprise after the data gathered from Speakers 1 and 2 (Table 8.34 and Table 8.37). Is it possible that when the difference in level of mispronunciation is more marked, then the two groups of informants (native Finnish listeners and native English listeners) are more similar in their responses than when the difference is smaller? Can one legitimately

Table 8.40 Significant contrasts on various traits of Speaker 3

Trait	Finnish-speaking listeners		English-speaking listeners		Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Mode	Median	Mode	Median		
	Tape 1: 'Good Pronunciation'					
	No statistically significant differences					
	Tape 2: 'Poor Pronunciation'					
	(*N* = 114)		(*N* = 73)			
Solidarity/Benevolence						
Friendly	fair+	fair+	fair+	fair+	*/–	-
Reliable	fair+	fair+	fair+	fair+	***/*	-
Sympathetic	fair+	fair+	fair+/very+ [b]	fair+	***/***	+

[a] The English-speaking listeners' perceptions of the speaker are more positive (+) or more negatove (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

claim here that the starker the contrast in the mispronunciation of the stimuli speakers, the closer the two groups in their reactions and evaluations?

Thus, in brief, neither the Finnish-speaking listeners nor the English-speaking listeners—separately—view this speaker in the same light when he is speaking with a Good Pronunciation and a Poor Pronunciation: both sets of listeners upgrade the speaker on work value traits with the phonemically more accurate pronunciation, but fail to react differentially (broadly speaking) on social value traits. Moreover, there is only a slight discrepancy *between* the evaluations of the two groups, Finnish-speaking v. English-speaking listeners, since there are only three significant differences between them, all of which involve solidarity/benevolence traits.

Speaker 4

20.1-year-old woman
Tape 1: 47 mistakes 'Poor Pronunciation'
Tape 2: 11 mistakes 'Good Pronunciation'

All the evaluations of the 14 traits made by the two groups of Finnish-speaking listeners (Poor Pronunciation and Good Pronunciation) were non-uniform, and highly significantly so at the $p < .001$ level (Chi Square test). Results for the English-speaking listeners were similarly non-uniform, all but two being highly significant at the $p < .001$ level (*ambitious* and *serious* for the Poor Pronunciation were significant at the $p < .01$ level).

The Finnish-speaking listeners have the overall general impression that Speaker 4—regardless of the version heard—is: *educated, hardworking, intelligent, friendly, honest, kind, reliable, serious, sympathetic* and *trustworthy* (Table 8.41).

Table 8.41 Evaluations of Finnish-speaking informants listening to Speaker 4

	Tape 1 [a]: 'Poor Pronunciation'		Tape 2 [b]: 'Good Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	fair-	fair-	fair+	fair+	***/***
Educated	fair+	middle	very+	very+	***/***
Hardworking	fair+	fair+	very+	very+	***/***
Intelligent	fair+	middle	very+	fair+	***/***
Rich	fair-	middle	fair+	fair+	***/***
Self-confident	fair-	fair-	fair+	fair+	***/***
Successful	middle	middle	very+	fair+	***/***
Solidarity/Benevolence					
Friendly	very+	fair+	very+	fair+	
Honest	fair+	fair+	very+	fair+	
Kind	very+	very+	fair+	fair+	**/–
Reliable	fair+	fair+	very+	fair+	
Serious	fair+	fair+	fair+	fair+	**/–
Sympathetic	fair+/very+ [c]	fair+	fair+	fair+	
Trustworthy	fair+	fair+	very+	fair+	***/**

[a] $N = 113$ for all traits. [b] $N = 114$ for all traits. [c] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

For four of these traits (the social traits *friendly, honest, reliable* and *sympathetic*) the evaluations of the two versions are the same, for neither the Mann-Whitney *U* test nor the Median test reveal any statistically significant difference between them ($p > .05$).

For five of the other six traits (*educated, hardworking, intelligent, serious* and *trustworthy*) the evaluations lean more to the positive pole for the Good Pronunciation than for the Poor Pronunciation.

For *kind*, however, it is the Poor Pronunciation that yields a more positive response than the Good Pronunciation (though the two statistical tests do not confirm each other). The fact that the three traits *friendly, kind* and *sympathetic* do not run in parallel with each other for this female speaker is noteworthy, though it is not altogether surprising after the similar non-parallelism found with Speaker 2 (a man): see Table 8.35, P. 227. At least here the *kind* trait remains on the positive side of the scale with the Poor Pronunciation, whereas with Speaker 2 the Fair Pronunciation prompted a negative *unsympathetic* response.

One trait (*successful*) prompts a 'middle' (but statistically significant) response with the Poor Pronunciation, but a positive response for the Good Pronunciation. In the ears of native Finnish listeners a high level of mispronunciation in English is no indication of success or failure. On the other hand, relatively good pronunciation rules out lack of success.

Three traits—all indicators of status/competence—are seen in a negative light when the Poor Pronunciation is evaluated and in a positive light when the Good Pronunciation is evaluated: depending on the version heard, the speaker is thought to be, respectively, *unambitious*, *poor* and *hesitant/unsure*, or else *ambitious*, *rich* and *self-confident*. That a Poor Pronunciation is associated with *unambitious* and *hesitant/unsure* is to be expected, for the Fair Pronunciation of Speaker 2 and the Poor Pronunciation of Speaker 3 provide similar negative results. However, this is the first and only time that native Finnish listeners give a negative rating for wealth: Poor Pronunciation (47 mistakes) in the mouth of a woman speaker would seem to indicate poverty.

Once again all the status/competence traits provoke responses which are significantly affected by the level of phonemic accuracy (Mann-Whitney *U* test and Median test: $p < .001$). By contrast only three of the seven solidarity/benevolence traits (only one if one relies on the Median test) provoke similar differential responses.

The overall impression of Speaker 4 gained by the English-speaking listeners —regardless of the version heard—is that she is: *educated*, *hardworking*, *intelligent*, *successful*, *friendly*, *honest*, *kind*, *reliable*, *sympathetic* and *trustworthy* (Table 8.42, P. 238). For seven of these traits (all with one exception being traits of solidarity/benevolence: *hardworking*, *friendly*, *honest*, *kind*, *reliable*, *sympathetic* and *trustworthy*) the evaluation is the same regardless of the version heard (Mann-Whitney *U* test and Median test: $p > .05$). For the three other traits (*educated*, *intelligent*, *successful)* the Good Pronunciation prompts a more positive response than the Poor Pronunciation.

One trait (*ambitious*) prompts a bimodal 'middle/fair+' response with the Poor Pronunciation, but a positive response ('fair+') with the Good Pronunciation. Thus for native English listeners, to a certain extent the worse a foreigner's pronunciation the more opaque the trait of *ambition* would appear to be.

Another trait (*self-confident*) could also be interpreted as being 'middle' for the Poor Pronunciation while being positive for the Good Pronunciation: the mode for the Poor Pronunciation is bi-modal, split between 'fair-' and 'fair+,' *self-confident* and *hesitant/unsure*. The difference between the two evaluations is clearly significant: Mann-Whitney *U* test and Median test, $p < .001$. The phonemic inaccuracy of this woman is not enough here, by itself, to prompt a firm impression of *hesitant/unsure*, though the same level of inaccuracy in Speaker 3 (a man) was. Neither the Poor Pronunciation of Speaker 4 nor that of Speaker 3 was dotted with the stumbles, false starts, stuttering, and so on that are perhaps more readily recognised as giving the impression of *hesitant/unsure*, and yet the evaluation of this level of phonemic inaccuracy (47 mistakes) would perhaps

Table 8.42 Evaluations of English-speaking informants listening to Speaker 4

	Tape 1 [a]: 'Poor Pronunciation'		Tape 2 [b]: 'Good Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	middle/fair+ [c]	middle	fair+	fair+	**/–
Educated	fair+	fair+	very+	very+	***/**
Hardworking	fair+	fair+	fair+	fair+	
Intelligent	fair+	fair+	very+	fair+	**/*
Rich	middle	middle	fair+	fair+	**/*
Self-confident	fair-/+ [c]	middle	very+	fair+	***/***
Successful	fair+	middle	very+	fair+	**/**
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	fair+	
Honest	middle/fair+ [c]	fair+	fair+	fair+	
Kind	fair+	fair+	fair+	fair+	
Reliable	fair+	fair+	fair+	fair+	
Serious	fair-	middle	fair+	fair+	
Sympathetic	fair+	fair+	fair+	fair+	
Trustworthy	fair+	fair+	fair+	fair+	

[a] $N = 70$ for all traits. [b] $N = 73$ for all traits. [c] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

seem to be mediated by gender of voice, a male voice being down-graded relative to a female voice.

One trait, and only one, *serious–fun-loving*, would seem to have the scales tipped mildly towards the negative pole for the Poor Pronunciation but towards the positive pole for the Good Pronunciation. However, the contrast between the two evaluations is not statistically significant.

As for the matter of wealth (*rich–poor*), which was already prominent enough to catch my attention with Speakers 1, 2 and 3, it would truly seem to be a more opaque characteristic the more the phonemic accuracy deteriorates. Here with Speaker 4 being evaluated by native English listeners, the characteristic is not negative (i.e. *poor*) despite the bad pronunciation, but is evaluated as middle: i.e. it is indecipherable.

It is again striking that all but one (two if one relies on the Median test) of the seven status/competence traits are evaluated differentially according to the version heard: the Good Pronunciation provokes a more positive evaluation than the Poor Pronunciation. None of the solidarity/benevolence traits occasion similarly dichotomous results: English-speaking listeners do not vary in their evaluations of the 'social values' of a speaker with a stronger or weaker Finnish accent.

The contrast between the evaluations of the Finnish-speaking and English-speaking listeners reveals some striking similarities and discrepancies (Table 8.43).

Table 8.43 Significant contrasts on various traits of Speaker 4

Trait	Finnish-speaking listeners		English-speaking listeners		Mann-Whitney *U* test / Median test	GB listeners +/-[a]
	Mode	Median	Mode	Median		
Tape 1: 'Poor Pronunciation'						
	(*N* = 113)		(*N* = 70)			
Status/Competence						
Ambitious	fair-	fair-	middle/fair+[b]	middle	*/**	+
Educated	fair+	middle	fair+	fair+	**/*	+
Intelligent	fair+	middle	fair+	fair+	***/**	+
Solidarity/Benevolence						
Kind	very+	very+	fair+	fair+	*/–	-
Tape 2: 'Good Pronunciation'						
	(*N* = 114)		(*N* = 73)			
Status/Competence						
Hardworking	very+	very+	fair+	fair+	**/**	-
Solidarity/Benevolence						
Friendly	very+	fair+	fair+	fair+	–/**	-

[a] The English-speaking listeners' perceptions of the speaker are more positive (+) or more negatove (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

Where Speaker 3's Good Pronunciation prompted no differences at all between the reactions of the Finnish-speaking listeners and those of the English-speaking listeners, here Speaker 4's Good Pronunciation prompts two differences, involving the traits of *hardworking* and *friendly*. The Finnish-speaking listeners grade the woman speaker as more *hardworking* and more *friendly* than the English-speaking listeners (in-group support). Why both groups of informants (native Finnish listeners and native English listeners) respond to the Good Pronunciations of Speakers 3 and 4 in practically the same way is a mystery: why does this almost perfect (but not quite) level of pronunciation (with 11 phonemic mistakes) prompt identical responses? The broad congruity of responses towards the two speakers is striking but puzzling (Table 8.40 and Table 8.43).

As regards the Poor Pronunciation, differences are seen among the informants on four traits: *ambitious*, *educated*, *intelligent* and (according to the Mann-Whitney *U* test) *kind*. Curiously, unlike the situation with Speaker 3, the first three traits are status/competence traits and only the fourth one is a solidarity/benevolence trait. Is it stretching the evidence to conclude that there is a diffe-

rence based on gender of voice here? Where a *male* speaker is concerned, Poor Pronunciation is viewed differentially by native Finnish listeners and native English listeners on some aspects of solidarity and benevolence but viewed identically for status/competence. However, where a *female* speaker is concerned the reverse is the case, i.e. some status/competence traits are viewed differentially whereas solidarity/benevolence traits are not. In other words, when a Finn speaks English with a bad Finnish accent, everybody—both native Finnish listeners and native English listeners—would seem to pick up on the same 'work image' (status/competence) *if the speaker is a man,* but everybody would seem to pick up on the same 'social image' (solidarity/benevolence) *if the speaker is a woman.* More experiments are of course necessary to confirm or disprove this interpretation.

It would seem that it is only when the pronunciation is perfect (i.e. Excellent), or moderately worse than perfect (i.e. Fair), but not when it is slightly worse than perfect or plain awful (i.e. Good or Poor), that the difference between the two groups of informants (Finnish-speaking v. English-speaking listeners) comes into focus and covers a greater number of traits. That is certainly the impression given when comparing the results here with the results from the first two speakers (Excellent Pronunciation and Fair Pronunciation: Table 8.34 and Table 8.37). With the Good Pronunciation and the Poor Pronunciation the two groups of informants would seem to have few differences between them.

In brief the Finnish-speaking listeners do not see this Speaker 4 in the same light, broadly speaking, when she speaks with a Poor Pronunciation and when she speaks with a Good Pronunciation: the different image extends over all the status/competence traits and some of the solidarity/benevolence traits. The English-speaking listeners, on the other hand, only see this speaker in a different light where the status/competence traits are concerned, for no matter how she speaks English (Good or Poor), she is still basically the same socially attractive (*friendly, honest, kind, reliable, serious, sympathetic, trustworthy*) person. As for the discrepancy between the native Finnish listeners and the native English listeners, it is restricted almost exclusively to reactions to a few of the status/competence traits with the Poor Pronunciation.

Speaker 5

20.75-year-old woman		
Tape 1:	0 mistakes	'Excellent Pronunciation'
Tape 2:	33 mistakes	'Fair Pronunciation'

All the evaluations of the 14 traits made by the two groups of Finnish-speaking listeners (Excellent Pronunciation and Fair Pronunciation) were non-uniform, and highly significantly so at the $p < .001$ level (Chi Square test). Results for the English-speaking listeners were similarly non-uniform, and highly significant so at the $p < .001$ level (also Chi Square test), though one result (*serious* for the Fair Pronunciation) was significant at the $p < .01$ level).

The overall impression of Speaker 5 that the Finnish-speaking listeners had—regardless of the version heard—was that she is 'positive' on all 14 traits right across the board: she is *ambitious*, *educated*, *hardworking*, *intelligent*, *rich*, *self-confident*, *successful*, *friendly*, *honest*, *kind*, *reliable*, *serious*, *sympathetic* and *trustworthy* (Table 8.44).

Table 8.44 Evaluations of Finnish-speaking informants listening to Speaker 5

	Tape 1 [a]: 'Excellent Pronunciation'		Tape 2 [b]: 'Fair Pronunciation'		
Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test
Status/Competence					
Ambitious	very+	very+	fair-/+ [c]	fair+	***/***
Educated	very+	very+	fair+	fair+	***/***
Hardworking	very+	very+	very+	fair+	***/**
Intelligent	very+	very+	fair+	fair+	***/***
Rich	fair+/very+ [c]	fair+	fair+	middle	***/***
Self-confident	very+	very+	fair+	fair+	***/***
Successful	very+	very+	fair+	fair+	***/***
Solidarity/Benevolence					
Friendly	very+	very+	very+	fair+	**/–
Honest	very+	fair+	fair+	fair+	
Kind	very+	fair+	very+	fair+	
Reliable	very+	fair+	fair+	fair+	***/**
Serious	fair+	fair+	fair+	fair+	*/*
Sympathetic	fair+	fair+	fair-/very+ [c]	fair+	**/–
Trustworthy	very+	very+	fair+	fair+	***/***

[a] $N = 112$ for all traits. [b] $N = 113$ for all traits except *ambitious* where $N = 113$. [c] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

Two of the traits, *ambitious* and *sympathetic*, are bi-modal for the Fair Pronunciation, with one peak on the negative side of the scale, but since the median is 'fair+' in both cases, these traits can safely be interpreted as being positive. Moreover, the impression Speaker 5 gives with the Excellent Pronunciation is significantly more positive for twelve (ten according to the Median test) of the fourteen traits than that she gives with the Fair Pronunciation. Only the *honest* and *kind* traits are evaluated similarly for the two versions (Mann-Whitney *U* test and Median test, $p > .05$).

The striking feature of Table 8.44 is the large number of traits where the discrepancy in the phonemic accuracy of the versions evaluated (Excellent Pronunciation v. Fair Pronunciation) leads to a difference in the impression of this speaker: all the status/competence traits are differentially evaluated, as also five of the seven solidarity/benevolence traits. This is more than for any of the other speakers. The fact that the overall impression remains positive for both versions

detracts nothing from the importance of this finding: how positive an impression the native Finnish listeners gain of this speaker is clearly dependent on her level of phonemic accuracy.

The two traits that are 'odd-men-out' are *honest* and *kind*. That *honest* is evaluated in the same way for both versions here comes as no surprise at all since the evaluation of this trait proved to be equally unaffected by the phonemic accuracy/inaccuracy of the previous four speakers. However, the fact that *kind* prompts no dichotomous response either is surprising after Speaker 4, but expected inasmuch as the trait is similarly undifferentiated for Speakers 1, 2 and 3. Honesty and kindness are obviously inscrutable traits for native Finnish listeners.

For the first time among the five speakers, the results for *friendly* would appear to reveal dichotomous evaluations (though the two tests fail to confirm each other). The trait *sympathetic* would also seem to prompt dichotomous responses, though this result, too, is ambiguous, the two tests once again failing to confirm each other. The findings for this speaker would seem to suggest that the three traits of *friendly*, *kind*, and *sympathetic*, which one might expect to be inter-linked or inter-dependent, are in fact not so for native Finnish listeners.

The Fair Pronunciation here prompts a more or less positive reaction for the *ambitious* trait ('fair-/+' / 'fair+'). Of all the speakers, and of the five versions which are Fair or Poor, this is the only one which does *not* produce a negative response. Bad mispronunciation would seem to be equated by Finnish-speaking listeners with *unambitious*, though this is not the case for this woman speaker here.

The Fair Pronunciation also inspires a positive reaction for the *self-confident* trait, which—if one treats Speaker 1 as a mild aberration resulting from initial uncertainty and cold feet—is again unusual. At least for the Fair Pronunciation and Poor Pronunciations of Speakers 2, 3 and 4 the *self-confident* trait is rated negatively.

The Fair Pronunciation of this speaker also creates—unusually—a mild impression of wealth: the response to *rich* is 'middle-but-tending-to-positive.' The Fair Pronunciations and Poor Pronunciations of Speakers 1–4 provoke a 'middle' or 'middle-but-tending-to-negative' reaction.

There are several unusual things happening here with this last speaker and it is possible that some intervening variable (or variables) is affecting the result. Voice quality or pitch, smoothness of articulation, or some other feature may be at play here, creating a more positive image of the speaker than might be expected on the basis of degree of phonemic inaccuracy. Alternatively, even though one might have expected fatigue to have started exerting a negative influence on the reactions of these Finnish-speaking informants, the influence of task experience may be having quite the opposite effect, exerting a *positive* sway on their reactions: the informants may feel more confident in performing their task, and their critical acumen may be sharper.

The general impression that the English-speaking listeners form of Speaker 5—regardless of the version heard—is that she is positive for thirteen of the four-

teen traits, the exception being *rich*, which although being positive ('very+') for the Excellent Pronunciation is only 'middle'—and significantly so—for the Fair Pronunciation (Table 8.45).

Table 8.45 Evaluations of English-speaking informants listening to Speaker 5

Trait	Tape 1 [a]: 'Excellent Pronunciation'		Tape 2 [b]: 'Fair Pronunciation'		Mann-Whitney *U* test / Median test
	Mode	Median	Mode	Median	
Status/Competence					
Ambitious	very+	very+	fair+	fair+	***/***
Educated	ex+	ex+	fair+	fair+	***/***
Hardworking	very+	very+	fair+	fair+	***/***
Intelligent	ex+	very+	fair+	fair+	***/***
Rich	very+	very+	middle	middle	***/***
Self-confident	ex+	very+	fair+	middle	***/***
Successful	very+	very+	fair+	fair+	***/***
Solidarity/Benevolence					
Friendly	fair+	fair+	fair+	fair+	
Honest	very+	very+	fair+	fair+	
Kind	very+	fair+	fair+	fair+	
Reliable	very+	very+	fair+/very+ [c]	fair+	***/***
Serious	fair+	fair+	fair+	fair+	*/–
Sympathetic	fair+	fair+	fair+	fair+	
Trustworthy	very+	fair+	fair+	fair+	

[a] $N = 70$ for all traits. [b] $N = 73$ for all traits. [c] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

The fact that the trait *rich–poor* is opaque ('middle') for the Fair Pronunciation is noteworthy, as also the fact that for the Excellent Pronunciation it is comfortably positive: 'very+.' The native English listeners are loathe to commit themselves here for the Fair Pronunciation, but are more than willing to say 'Yes, rich,' for the Excellent Pronunciation.

It is unique that with the Excellent Pronunciation the mode and/or median for three traits (*educated, intelligent* and *self-confident*) are 'ex+,' i.e. the evaluation is at the extreme positive end of the scale, a result not found for any other traits with either informant group listening to any version of the other speakers. A further six traits (*ambitious, hardworking, rich, successful, honest* and *reliable*) have mode and median as 'very+,' thus giving an extraordinary 9 traits with mode and median at 'very+' or 'ex+.' The Excellent Pronunciations of Speakers 1 and 2 come nowhere close, with only three traits each (both including *educated* and *intelligent*) at such a level (modes and medians at 'very+'). Either informants have become so attuned to the exercise that they are now at last prepared to

utilise the extreme positive end of the scale, or else this speaker has some non-phonemic feature (or features) in her speech which encourages an extreme positive response. However, why the extreme positive response should affect these three traits and no others is unclear: the speaker seems to be a 'clever girl,' but is this really a direct result of phonemic accuracy?

For all seven status/competence traits and for two of the solidarity/benevolence traits the impression gained by the English-speaking listeners is more positive with the Excellent Pronunciation than with the Fair Pronunciation. Five solidarity/benevolence traits (*friendly*, *honest*, *kind*, *sympathetic* and *trustworthy*) are evaluated in the same way by both groups of informants ($p > .05$), the stark contrast in phonemic accuracy of the two versions (0 mistakes v. 33 mistakes) apparently having no effect on the evaluation of the speaker here. The conspicuous exception is *reliability*, more positive for the better pronunciation than for the worse one. This situation mirrors that with Speaker 2 and Speaker 3, suggesting that this trait is the only one of the 'social value' traits which is sometimes—though not always—affected by phonemic accuracy. Overall one can say that for native English listeners the status/competence traits *are* coloured by the phonemic accuracy of this speaker, but the overwhelming majority of solidarity/benevolence traits are not.

The difference between the evaluations of the Finnish-speaking listeners and the English-speaking listeners is almost non-existent as far as the Fair Pronunciation is concerned (Table 8.46, P. 245). There is only one difference according to the Mann-Whitney *U* test ($p < .05$), and that is with *trustworthy*, the native English listeners rating the speaker more positively for this trait than do the native Finnish listeners. However, the Median test fails to confirm this.

As for the Excellent Pronunciation, the discrepancies between the two groups vary slightly according to which statistical test one uses.

When relying on the results of the Mann-Whitney *U* test, the two groups of informants rate six traits significantly differently: *educated*, *intelligent*, *rich*, *self-confident*, *friendly* and *reliable*. For all these traits except *friendly*, the English-speaking listeners are significantly more positive in their evaluation than the Finnish-speaking listeners; the Finnish-speaking listeners are more negative, less supportive, less tolerant (in-group rejection). However, the finding that the Finnish-speaking listeners are more positive for the *friendly* trait confirms the results with Speaker 2 (Table 8.37, P. 230) and Speaker 3 (Table 8.40, P. 235), though one can only speculate as to why this is so: are the Finnish-speaking listeners merely betraying in-group (peer group) support (assuming that they recognised the speakers to be Finns), and the English-speaking listeners revealing some sort of inherent sub-conscious atavistic suspicion of foreigners? The data of course provides no answers to the question 'Why?'

When relying on the Median test the levels of significance are slightly different for some of these differentiated traits, but more importantly the test adds *ambitious* and *successful* to the list of traits evaluated more positively by the native English listeners than by the native Finnish listeners.

Table 8.46 Significant contrasts on various traits of Speaker 5

Trait	Finnish-speaking listeners		English-speaking listeners		Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Mode	Median	Mode	Median		
	Tape 1: 'Excellent Pronunciation'					
	(*N* = 112)		(*N* = 70)			
Status/Competence						
Ambitious	very+	very+	very+	very+	–/*	+
Educated	very+	very+	ex+	ex+	***/**	+
Intelligent	very+	very+	ex+	very+	***/***	+
Rich	fair+/ very+ [b]	fair+	very+	very+	**/*	+
Self-confident	very+	very+	ex+	very+	***/***	+
Successful	very+	very+	very+	very+	–/*	+
Solidarity/Benevolence						
Friendly	very+	very+	fair+	fair+	**/**	-
Reliable	very+	fair+	very+	very+	*/**	+
	Tape 2: 'Fair Pronunciation'					
	(*N* = 113)		(*N* = 73)			
Solidarity/Benevolence						
Trustworthy	fair+	fair+	fair+	fair+	*/– [a]	+

[a] The English-speaking listeners' perceptions of the speaker are more positive (+) or more negatove (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.
*** $p < .001$; ** $p < .01$; * $p < .05$.

In brief, therefore, one can say that the Finnish-speaking listeners as a group differentiate between the two versions (Excellent Pronunciation and Fair Pronunciation) across a wider range of traits than the English-speaking listeners do as a group: both groups make a distinction on all the status/competence traits, but the native Finnish listeners make a distinction on a wider range of solidarity/ benevolence traits than the native English listeners. As for the between-group contrasts, the native English listeners are more willing to give more positive ratings for the Excellent Pronunciation than the native Finnish listeners, and this greater support is focused on the status/competence traits. There is hardly any difference at all between the evaluations of the two groups when they listen to the Fair Pronunciation.

When looking at the results of the English-speaking listeners with all five speakers, it is perhaps significant that the level of mispronunciation always affects the *intelligent* trait whereas it never affects the *friendly* trait, the former supporting and the latter disconfirming the results obtained with variously foreign-accented Swedish by Cunningham-Andersson (1997).

8.3.1. Gender of informants

The gender of informants would seem to act only in a very limited way as a moderating variable on the evaluations made *within* the two groups of Finnish-speaking listeners and English-speaking listeners. I shall examine the two groups separately.

Finnish-speaking listeners

Within the native Finnish listener informant group there is a limited degree of gender differentiation, but nothing startling (Table 8.47 through to Table 8.51). Where a difference is discernible, the women are typically more positive in their evaluations than the men. However, most of these significant results are only mildly significant (at the $p < .05$ level), and of these one of the two tests usually disconfirms the other. Only with two evaluations, both with Speaker 2 (*self-confident* for the Fair Pronunciation, and *intelligent* for the Excellent Pronunciation), do both the tests indicate a highly significant result ($p < .001$).

Most of the traits which show any gender differentiation are concerned with solidarity/benevolence, and five of the ten voice guises presented for evaluation show differentiation on only those traits (Speaker 1 Tapes 1 & 2; Speaker 4 Tape 2; Speaker 5 Tapes 1 & 2). In other words when Finnish men and women have different perceptions of these speakers, the difference generally lies in the social values (solidarity/benevolence) and not in the work values (status/competence).

Table 8.47 Speaker 1: Finnish-speaking listeners

Trait	SF Women Mode	SF Women Median	SF Men Mode	SF Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Excellent Pronunciation'					
	(N = 59)		(N = 54)			
Solidarity/Benevolence						
Honest	very+	very+	fair+/very+ [b]	fair+	*/*	+
Kind	very+	fair+	fair+	fair+	*/–	+
Reliable	very+	fair+	fair+	fair+	**/*	+
Trustworthy	very+	fair+	fair+	fair+	*/*	+
	Tape 2: 'Fair Pronunciation'					
	(N = 60)		(N = 54)			
Solidarity/Benevolence						
Friendly	very+	fair+	fair+	fair+	*/–	+
Reliable	very+	fair+	fair+	fair+	–/*	+

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.

[b] The result is bi-modal.

** $p < .01$; * $p < .05$.

Table 8.48 Speaker 2: Finnish-speaking listeners

Trait	SF Women Mode	SF Women Median	SF Men Mode	SF Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Fair Pronunciation'					
	(*N* = 59)		(*N* = 54)			
Status/Competence						
Intelligent	fair-	middle	fair-	fair-	*/–	+
Self-confident	fair+	fair+	fair-	fair-	***/***	+
Successful	fair-/ fair+ [b]	middle	fair-	fair-/ middle [c]	*/–	+
Solidarity/Benevolence						
Honest	very+	fair+	fair+	fair+	–/*	+
Serious	fair-	middle	fair+	fair+	*/**	-
	Tape 2: 'Excellent Pronunciation'					
	(*N* = 60)		(*N* = 54)			
Status/Competence						
Educated	very+	very+	very+	fair+	*/–	+
Hardworking	fair+	fair+	fair-	middle	*/–	+
Intelligent	very+	very+	fair+	fair+	***/***	+
Self-confident	very+	very+	very+	fair+	*/–	+
Successful	fair+	fair+	fair+	fair+	**/**	+
Solidarity/Benevolence						
Friendly	fair+	fair+	fair+	fair+	**/–	+
Honest	fair+	fair+	fair+	fair+	**/*	+
Reliable	fair+	fair+	fair+	fair+	*/–	+
Sympathetic	fair+	fair+	fair+	fair+	*/–	+
Trustworthy	very+	fair+	fair+	fair+	**/*	+

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.

[b] The result is bi-modal.

[c] The median falls between the two values.

*** $p < .001$; ** $p < .01$; * $p < .05$.

Table 8.49 Speaker 3: Finnish-speaking listeners

Trait	SF Women Mode	SF Women Median	SF Men Mode	SF Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Good Pronunciation'					
	(*N* = 59)		(*N* = 54)			
Status/Competence						
Self-confident	fair-/ very+ [b]	middle	very+	fair+	**/–	-
Solidarity/Benevolence						
Friendly	very+	very+	very+	fair+	*/**	+
Kind	fair+	fair+	fair+	fair+	*/–	+
	Tape 2: 'Poor Pronunciation'					
	(*N* = 60)		(*N* = 54)			
	No statistically significant differences					

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The result is bi-modal. ** $p < .01$; * $p < .05$.

Table 8.50 Speaker 4: Finnish-speaking listeners

Trait	SF Women Mode	SF Women Median	SF Men Mode	SF Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Poor Pronunciation'					
	(*N* = 59)		(*N* = 54)			
Status/Competence						
Ambitious	fair-	fair-	fair-	fair-	*/–	+
Solidarity/Benevolence						
Friendly	very+	very+	fair+	fair+	*/*	+
Honest	very+	very+	fair+	fair+	*/**	+
Reliable	very+	fair+	fair+	fair+	*/–	+
Trustworthy	fair+	fair+	fair-/fair+ [b]	middle	*/–	+
	Tape 2: 'Good Pronunciation'					
	(*N* = 60)		(*N* = 54)			
Solidarity/Benevolence						
Friendly	very+	very+	very+	fair+	***/*	+
Sympathetic	fair+	fair+	fair+	fair+	*/–	+

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The result is bi-modal. *** $p < .001$; ** $p < .01$; * $p < .05$.

Table 8.51 Speaker 5: Finnish-speaking listeners

Trait	SF Women Mode	SF Women Median	SF Men Mode	SF Men Median	Mann-Whitney *U* test / Median test	Women listeners +/-[a]
	Tape 1: 'Excellent Pronunciation'					
	(*N* = 58)		(*N* = 54)			
Solidarity/Benevolence						
Kind	very+	very+	fair+	fair+	*/–	+
Reliable	fair+	very+	very+	fair+	*/–	+
	Tape 2: 'Fair Pronunciation'					
	(*N* = 60)		(*N* = 53)			
Solidarity/Benevolence						
Kind	fair+/ very+[b]	fair+	very+	fair+	*/–	+
Sympathetic	very+	fair+	fair-	middle	*/–	+
Trustworthy	fair+	fair+	fair+	middle	*/–	+

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The result is bi-modal. * $p < .05$.

The traits which lead to a measure of consistent gender-prompted differentiation among the Finnish-speaking listeners are listed in Table 8.52.

Table 8.52 Frequency of appearance (>= 3) of main traits

Trait	Frequency[a]
Friendly	5
Reliable	5
Honest	4
Kind	4
Trustworthy	4
Sympathetic	3
Self-confident	3

[a] Maximum possible = 10.

Only *friendly* and *reliable* appear as many as five times out of the possible total of ten (the ten versions for evaluation). Traits concerned with status/competence seem hardly prone to gender differentiation at all, for only *self-confident* appears as many as three times. Of these three times one is extremely unusual (that for Speaker 3's Good Pronunciation) in that for once the men are more positive in their evaluations than the women. The results for Speaker 2's Excellent Pronunciation are clearly exceptional in having five of the status/competence traits revealing gender differentiation.

The reverse side of this Table 8.52 is not presented but is just as important: the seven traits not listed betray *no* gender differentiation for eight or more of the ten versions presented for evaluation. It is this conspicuous 'absence' that forms the basis of my initial claim that there is only a limited degree of gender differentiation within the group of Finnish-speaking listeners.

An examination of which speakers are differentiated by informant gender for the same traits with both versions they produce reveals no consistency. Speaker 2, with his unusually wide range of differentiated traits, has four traits which are differentiated with both versions: *intelligent, self-confident, successful, honest.* Speakers 1, 4 and 5 have only one trait each being differentiated for both the versions they produce (*reliable, friendly* and *kind* respectively). It is possible that in these cases the differentiation is dependent more on the voice quality and voice set and/or the suprasegmental features which are common to both versions rather than on the phonemic accuracy which is unique to each version.

English-speaking listeners

Within the native English listener informant group no general pattern of gender differentiation is detectable (Table 8.53 through to Table 8.57). This seemingly non-result is important: English-speaking men and English-speaking women perceive Finnish-accented speakers in the same way.

All the results which show any significant difference with either the Mann-Whitney *U* test or the Median test (except *serious* for Speaker 2's Fair Pronunciation, $p < .01$, and *hardworking* for Speaker 5's Excellent Pronunciation, $p < .001$) are only mildly significant, and most often one of the two tests disconfirms the other. Where there are differentiations, there is perhaps a slight tendency to differentiation on the solidarity/benevolence scales rather than on the status/competence scales (though the results for Speaker 5 go against this trend),

Table 8.53 Speaker 1: English-speaking listeners

Trait	GB Women Mode	GB Women Median	GB Men Mode	GB Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Excellent Pronunciation'					
	(*N* = 48)		(*N* = 22)			
Solidarity/Benevolence						
Honest	very+	fair+/ very+ [b]	middle/ fair+ [c]	middle/ fair+ [b]	*/–	+
	Tape 2: 'Fair Pronunciation'					
	(*N* = 50)		(*N* = 23)			
	No statistically significant differences					

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The median falls between the two values. [c] The result is bi-modal. * $p < .05$.

and generally the women are more positive in their evaluations than the men (the one exception being the *rich* trait with Speaker 5's Fair Pronunciation).

Table 8.54 Speaker 2: English-speaking listeners

Trait	GB Women Mode	GB Women Median	GB Men Mode	GB Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Fair Pronunciation'					
	($N = 48$)		($N = 22$)			
Solidarity/Benevolence						
Serious	very+	very+	fair+	fair+	**/**	+
	Tape 2: 'Excellent Pronunciation'					
	($N = 50$)		($N = 23$)			
Solidarity/Benevolence						
Serious	very+/ex+ [b]	very+	fair+	fair+	*/–	+

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The result is bi-modal. ** $p < .01$; * $p < .05$.

Table 8.55 Speaker 3: English-speaking listeners

Trait	GB Women Mode	GB Women Median	GB Men Mode	GB Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Good Pronunciation'					
	($N = 48$)		($N = 22$)			
Status/Competence						
Ambitious	fair+	fair+	fair+	middle/fair+ [b]	*/–	+
Solidarity/Benevolence						
Honest	very+	very+	middle	fair+	*/–	+
Kind	very+	very+	fair+	fair+	*/*	+
Reliable	very+	very+	fair+	fair+	*/*	+
Serious	very+	fair+/very+ [b]	fair+	fair+	–/*	+
	Tape 2: 'Poor Pronunciation'					
	($N = 50$)		($N = 23$)			
	No statistically significant differences					

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The median falls between the two values. * $p < .05$.

Table 8.56 Speaker 4: English-speaking listeners

Trait	GB Women Mode	GB Women Median	GB Men Mode	GB Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Poor Pronunciation'					
	(*N* = 48)		(*N* = 22)			
Solidarity/Benevolence						
Friendly	fair+	fair+	fair+	fair+	–/*	+
Trustworthy	fair+	fair+	middle	middle	*/–	+
	Tape 2: 'Good Pronunciation'					
	(*N* = 50)		(*N* = 23)			
	No statistically significant differences					

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
* $p < .05$.

Table 8.57 Speaker 5: English-speaking listeners

Trait	GB Women Mode	GB Women Median	GB Men Mode	GB Men Median	Mann-Whitney *U* test / Median test	Women listeners +/- [a]
	Tape 1: 'Excellent Pronunciation'					
	(*N* = 48)		(*N* = 22)			
Status/Competence						
Ambitious	ex+	very+	very+	very+	*/*	+
Hardworking	ex+	very+	very+	very+	***/***	+
Rich	very+	very+	very+	fair+	*/–	+
Successful	ex+	very+	very+	fair+/ very+ [b]	*/–	+
Solidarity/Benevolence						
Honest	very+	very+	middle	fair+	*/–	+
Reliable	ex+	very+	very+	very+	–/*	+
	Tape 2: 'Fair Pronunciation'					
	(*N* = 50)		(*N* = 23)			
Status/Competence						
Rich	middle	middle	fair+	fair+	*/**	-

[a] Women's perceptions of the speaker are more positive (+) or more negative (-) than those of the men.
[b] The median falls between the two values. *** $p < .001$; ** $p < .01$; * $p < .05$.

It is noteworthy that no traits consistently lead to gender differentiation: English-speaking men and women do not differ from each other in the evaluation

of any particular trait. Only two traits appear with any regularity (*honest* and *serious*), but even these appear in connection with only three of the ten versions. One might have expected a gender-provoked dichotomy to appear on traits closely connected with both social values (solidarity/benevolence) and work values (status/competence), but this is not the case.

It is interesting to note that a specific level of phonemic accuracy/inaccuracy does not seem to increase the salience of any particular trait. If it did, one would expect particular traits to appear *consistently* with only one particular level of pronunciation (Excellent Pronunciation, Good Pronunciation, Fair Pronunciation or Poor Pronunciation) rather than with several or all, but this does not occur. *Honest* appears with the Excellent Pronunciations of Speakers 1 and 5 but not Speaker 2, and then it also appears with the Good Pronunciation of Speaker 3.

One final point worth noting is that Speakers 2 and 5 both have one trait, *serious* and *rich* respectively, which is differentiated with both versions. Here one is tempted to conclude that the voice quality and/or voice set and/or suprasegmentals (shared by both versions) are prompting the gender differentiation rather than the level of phonemic accuracy (unique to each version.)

Finnish-speaking women v. English-speaking women, and Finnish-speaking men v. English-speaking men

A further breakdown of the data, separating all the men from all the women, to see whether there is any difference in the evaluations of the Finnish-speaking women and the English-speaking women, and of the Finnish-speaking men and the English-speaking men, leads to some totally unexpected results (Table 8.58 through to Table 8.63). The null hypothesis is of course that all the women, and all the men, separately, evaluate the voice guises in the same way, regardless of their mother tongue.

– Excellent Pronunciation

Starting with an examination of the guises which exhibit Excellent Pronunciation, it is immediately clear that the Finnish-speaking women do not react to the guises in the same way as the English-speaking women, whereas the Finnish-speaking men and the English-speaking men generally do (Table 8.58, P. 254, and Table 8.59, P. 255).

For Speaker 1 the different evaluation of the Finnish-speaking listeners and English-speaking listeners (seen in Table 8.34) for three status/competence traits is solely dependent on the evaluation of the women, for the Finnish-speaking men and the English-speaking men do not differ at all. The English-speaking women are more positive in their evaluations than the Finnish-speaking women.

A similar though not so clear-cut situation is found with Speaker 2: the women are basically the ones who disagree in their evaluation of the Excellent Pronunciation (seen in Table 8.37), not the men. The women differ for nine traits which seem to fall into two groups for upgrading and downgrading: the English-speaking women upgrade three status/competence traits relative to the Finnish-speaking women, but downgrade five solidarity/benevolence traits (leaving one odd-man-out, the upgrading with *serious*). In other words, the Finnish-speaking

Table 8.58 Excellent Pronunciation (Speakers 1 and 2)

Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Speaker 1: Tape 1					
	SF Women (*N* = 59)		GB Women (*N* = 48)			
Status/Competence						
Educated	very+	very+	ex+	very+	–/*	+
Intelligent	very+	fair+	very+	very+	***/**	+
Rich	middle	middle	fair+/very+ [b]	fair+	***/**	+
	SF Men (*N* = 54)		GB Men (*N* = 22)			
	No statistically significant differences					
	Speaker 2: Tape 2					
	SF Women (*N* = 60)		GB Women (*N* = 50)			
Status/Competence						
Educated	very+	very+	very+	very+	*/*	+
Intelligent	very+	very+	very+	very+	*/*	+
Self-confident	very+	very+	very+	very+	*/–	+
Solidarity/Benevolence						
Friendly	fair+	fair+	fair+	middle	***/**	-
Honest	fair+	fair+	middle	fair+	**/*	-
Kind	fair+	fair+	middle/fair+ [b]	middle	*/–	-
Serious	fair+	fair+	very+/ex+ [b]	very+	**/*	+
Sympathetic	fair+	fair+	fair-	fair-	***/*	-
Trustworthy	very+	fair+	fair+	fair+	**/*	-
	SF Men (*N* = 54)		GB Men (*N* = 23)			
Status/Competence						
Educated	very+	fair+	very+	very+	*/–	+
Intelligent	fair+	fair+	very+	very+	*/*	+
Rich	fair+	fair+	middle/fair+ [b]	fair+	*/**	+

[a] The English-speaking listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.

*** $p < .001$; ** $p < .01$; * $p < .05$.

women are overall more negative than the English-speaking women for the work traits (in-group rejection), more positive for the social traits (in-group support).

The two groups of men differ for only three status/competence traits (including *rich*, which the women do not differ on at all), the English-speaking men being more positive than the Finnish-speaking men.

Table 8.59 Excellent Pronunciation (Speaker 5)

Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Speaker 5: Tape 1					
	SF Women (*N* = 58)		GB Women (*N* = 48)			
Status/Competence						
Ambitious	very+	very+	ex+	very+	*/**	+
Educated	very+	very+	ex+	ex+	***/**	+
Intelligent	very+	very+	ex+	ex+	***/***	+
Rich	fair+	fair+	very+	very+	***/**	+
Self-confident	very+	very+	ex+	ex+	***/**	+
Successful	very+	very+	ex+	very+	–/*	+
Solidarity/Benevolence						
Friendly	very+	very+	fair+	fair+	**/*	-
Kind	very+	very+	very+	fair+	–/*	-
Reliable	fair+	very+	ex+	very+	–/*	+
	SF Men (*N* = 54)		GB Men (*N* = 22)			
Status/Competence						
Hardworking	very+	very+	very+	very+	*/–	-

[a] The English-speaking listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. *** $p < .001$; ** $p < .01$; * $p < .05$.

It also becomes apparent from Table 8.58 that the women differ in their perception of trustworthiness, a feature which did not appear in the basic contrast between Finnish-speaking listeners and English-speaking listeners (seen in Table 8.37).

A similar startling pattern is seen with Speaker 5 (Table 8.59). The original dichotomy observed between the Finnish-speaking listeners and the English-speaking listeners (seen in Table 8.46) is once more seen to rest almost exclusively upon the women, not upon the men. The men disagree for *hardworking*, nothing else, whereas the women disagree for nine traits, and these predominantly 'work' traits. The women also differ in their perception of the speaker's kindness, a feature not apparent from the basic contrast between Finnish-speaking listeners and English-speaking listeners (seen in Table 8.46).

The general pattern of upgrading and downgrading would again seem to suggest that overall the English-speaking listeners (particularly the women) are more positive in their evaluations of status/competence traits than Finnish-speaking listeners, but more negative in their evaluation of solidarity/benevolence traits.

Where Excellent Pronunciation is concerned, it seems as if the Finnish-speaking men and the English-speaking men share the same critical standards, but this is not the case for the women.

– Good Pronunciation

Progressing to an examination of the data with the Good Pronunciations, the surprise rests in the fact that the same imbalance is not found to the same degree (Table 8.60).

Table 8.60 Good Pronunciation

Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Speaker 3: Tape 1					
	SF Women ($N = 59$)		GB Women ($N = 48$)			
Status/Competence						
Self-confident	fair-/very+ [b]	middle	fair+	fair+	*/–	+
	SF Men ($N = 54$)		GB Men ($N = 22$)			
Status/Competence						
Self-confident	very+	fair+	fair+	fair+	–/*	-
	Speaker 4: Tape 2					
	SF Women ($N = 60$)		GB Women ($N = 50$)			
Status/Competence						
Hard-working	very+	very+	fair+	fair+	***/**	-
Solidarity/Benevolence						
Friendly	very+	very+	fair+	fair+	***/***	-
Honest	very+	very+	fair+	fair+	*/–	-
	SF Men ($N = 54$)		GB Men ($N = 23$)			
	No statistically significant differences					

[a] The English-speaking listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.

*** $p < .001$; ** $p < .01$; * $p < .05$.

The complete congruity in the evaluations made by the two language groups of Speaker 3's Good Pronunciation (seen in Table 8.40) is more or less confirmed by the results both within the female group and within the male group. The significance value for the *self-confident* trait just tips into mildly significant ($p < .05$) for either the Mann-Whitney *U* test or the Median test for one or the other group. No other traits show any differentiation.

For Speaker 4, however, the original indication of a differentiation on two traits (*hard-working* and *friendly*, seen in Table 8.43) is confirmed *only* within the female group, and at a highly significant level. The Finnish-speaking women and the English-speaking women have a different perception of the speaker on the traits of *hardworking*, *friendly* and—not apparent before—*honest*, but the

Finnish-speaking men and the English-speaking men do not: the men agree with each other.

Thus for the second level of phonemic accuracy, Good Pronunciation, there is only the mildest of indications that there might be a within gender group differentiation based on mother tongue, but only within the group of women.

– Fair Pronunciation

Progressing down the scale of phonemic inaccuracy to the three Fair Pronunciations, the picture here is somewhat confusing (Table 8.61 below and Table 8.62, P. 259): there is relative wealth of detail for Speakers 1 and 2, for both the set of women and the set of men, yet Speaker 5 prompts hardly any differentiation at all.

For Speaker 1's Fair Pronunciation with 31 phonemic mistakes (Table 8.61) the traits first revealed as being differently evaluated by the Finnish-speaking listeners and the English-speaking listeners overall (seen in Table 8.34) widely

Table 8.61 Fair Pronunciation (Speaker 1)

Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	GB listeners +/-[a]
	Speaker 1: Tape 2					
	SF Women ($N = 60$)[b]		GB Women ($N = 50$)			
Status/Competence						
Ambitious	fair-	fair-	fair-	middle	*/*	+
Educated	fair+	fair+	very+	very+	***/***	+
Intelligent	fair+	middle	very+	very+	***/***	+
Rich	middle	middle	fair+	middle	**/**	+
Successful	fair-	middle	fair+	fair+	**/**	+
Solidarity/Benevolence						
Sympathetic	fair+	fair+	very+	fair+	*/–	+
Trustworthy	fair-	middle	fair+	fair+	**/–	+
	SF Men ($N = 54$)		GB Men ($N = 23$)			
Status/Competence						
Ambitious	fair-	fair-	middle	middle	–/*	+
Educated	fair+	fair+	very+	very+	***/**	+
Intelligent	fair+	fair+	very+	very+	***/***	+
Rich	fair-	fair-	fair+	middle	**/**	+
Solidarity/Benevolence						
Sympathetic	fair+	middle	fair+	fair+	***/***	+
Trustworthy	fair-	middle	fair+	fair+	**/*	+

[a] The English-speaking listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. [b] $N = 59$ for *ambitious*.

*** $p < .001$; ** $p < .01$; * $p < .05$.

appear as being identically differentiated by the two groups of women and the two groups of men (*ambitious, educated, intelligent, rich, sympathetic, trustworthy*), the English-speaking listeners in both paired groups being significantly more positive than the Finnish-speaking listeners for both the status/competence traits and the solidarity/benevolence traits.

However, two traits which had originally just scraped onto the level of statistical significance fall off from the list, *hardworking* and *self-confident*, and one work trait, *successful*, now appears as being differentiated by the women but not by the men, the English-speaking women following the pattern of being more positive than the Finnish-speaking women for status/competence traits. In other words, for this speaker, at this level of phonemic inaccuracy, it would seem that the difference between the Finnish-speaking listeners and the English-speaking listeners overall does not break down under gender analysis in the same way as it does at the two higher levels of phonemic accuracy: both the men and the women differ among themselves for certain traits according to mother tongue, and this time the English-speaking listeners are always more positive than the Finnish-speaking listeners for both work and social traits.

For Speaker 2's Fair Pronunciation with 31 phonemic mistakes (Table 8.62, P. 259) one finds a similar pattern of conformity: the group of Finnish-speaking and English-speaking women, and the group of Finnish-speaking and English-speaking men, both disagree with each other for most of the traits originally found as being differentiated by the two language groups overall (Table 8.37). However, for this speaker the uniform pattern of more positive evaluations being given by the English-speaking listeners than by the Finnish-speaking listeners no longer holds true. Where the men are concerned, the English-speaking men upgrade five status/competence traits in comparison with the Finnish-speaking men, but downgrade two solidarity/benevolence traits. An almost similar pattern is found with the women, but with two notable exceptions: *self-confident* and *serious*: with the work trait *self-confident* the Finnish-speaking women are unexpectedly more *positive* than the English-speaking women, and with the social trait *serious* they are unexpectedly more *negative*. The fact that the Finnish-speaking women and the Finnish-speaking men evaluate the *self-confident* work trait in opposite directions (the women positive, the men negative) is striking, and is sufficient to explain why this trait failed to appear on the original table contrasting the two language groups overall (Table 8.37): the men and the women cancel each other out, reducing their overall evaluation to the same as that overall of the English-speaking listeners.

Once again the men distinguish themselves by evaluating wealth differently, this time the English-speaking men being more positive than the Finnish-speaking men. This is a feature not seen in (Table 8.37), as also the men's differentiation on another work trait, *successful*.

For Speaker 5's Fair Pronunciation with 33 phonemic mistakes (Table 8.62) the almost identical responses of the Finnish-speaking listeners and the English-speaking listeners overall as seen in the original analysis (Table 8.46) is more or

Table 8.62 Fair Pronunciation (Speakers 2 and 5)

Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Speaker 2: Tape 1					
	SF Women ($N = 59$)		GB Women ($N = 48$)			
Status/Competence						
Ambitious	fair-	fair-	fair-/ middle [b]	middle	*/**	+
Hard-working	very-/ fair- [b]	fair-	fair-/ fair+ [b]	middle	*/*	+
Self-confident	fair+	fair+	fair-	fair-	**/**	-
Solidarity/Benevolence						
Honest	very+	fair+	fair+	middle	*/*	-
Reliable	middle	fair+	fair-	middle	*/–	-
Serious	fair-	middle	very+	very+	***/***	+
	SF Men ($N = 54$)		GB Men ($N = 22$)			
Status/Competence						
Hard-working	fair-	fair-	fair-	middle	*/*	+
Intelligent	fair-	fair-	fair+	fair+	*/–	+
Self-confident	fair-	fair-	fair-	middle	*/–	+
Rich	middle	middle	middle	middle	*/–	+
Successful	fair-	fair-/ middle [c]	middle	middle/ fair+ [c]	*/–	+
Solidarity/Benevolence						
Honest	fair+	fair+	fair-/ middle [b]	middle	**/–	-
Kind	fair+	fair+	middle	middle	**/–	-
	Speaker 5: Tape 2					
	SF Women ($N = 60$)		GB Women ($N = 50$)			
Status/Competence						
Rich	fair+	fair+	middle	middle	***/***	-
	SF Men ($N = 53$)		GB Men ($N = 23$)			
Status/Competence						
Intelligent	fair+	fair+	fair+	fair+	*/–	+

[a] The English-speaking listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the Finnish-speaking listeners.

[b] The result is bi-modal.

[c] The median falls between the two values.

*** $p < .001$; ** $p < .01$; * $p < .05$.

less confirmed here. The two groups of men would appear to disagree with each other on *intelligent* (Mann-Whitney U test $p < .05$, but unconfirmed by the Median test), while the women clearly disagree on *rich* ($p < .001$), the Finnish-speaking women unusually being more positive for this work trait than the English-speaking women. Basically this speaker, with this level of mispronunciation, prompts no within group gender differentiation at all: she is clearly unusual when compared to Speakers 1 and 2. However, her Excellent Pronunciation also prompted minimal differentiation between the two groups of men (Table 8.59).

Thus, at the level of Fair Pronunciation, one is forced to admit that a breakdown of the data on the basis of Finnish-speaking women v. English-speaking women, and Finnish-speaking men v. English-speaking men, reveals little more than what was already learned from the straight contrasts of Finnish-speaking listeners v. English-speaking listeners.

– Poor Pronunciation

Finally examining the bottom level of phonemic inaccuracy, the Poor Pronunciations with 47 phonemic mistakes, another surprise is in store (Table 8.63, P. 261).

For Speaker 3's Poor Pronunciation one would expect on the basis of Table 8.40 that the two groups of women and the two groups of men would distinguish themselves according to mother-tongue for the social traits of *friendly*, *reliable* and *sympathetic*. This is indeed the case for the women but not quite so for the men: the men are not differentiated for *friendly*, a trait which was in any case only just significant in the overall contrast of Finnish-speaking listeners v. English-speaking listeners.

For Speaker 4's Poor Pronunciation the dichotomy originally found between Finnish-speaking listeners and English-speaking listeners overall for *ambitious*, *educated*, *intelligent* and *kind* (Table 8.43) is shown to be dependent almost exclusively upon a difference in the evaluations not of the women but of the *men*. The women only betray a difference for *intelligent*, but the men for all four expected traits, three of which, it should be noted, are status/competence traits.

To summarise this survey through the levels of phonemic accuracy/inaccuracy, there is some evidence here of a swing from Finnish-speaking women differentiating themselves from, and being more positive than, English-speaking women at the top level of pronunciation for solidarity/benevolence traits through to Finnish-speaking men differentiating themselves from, and being more negative than, English-speaking men at the bottom level of pronunciation for status/competence traits. The differentiation is clearer at the top end of the scale than at the bottom end, but there is enough evidence to suspect that there is a real swing here.

Finally, there would appear to be no traits which the Finnish-speaking women and the English-speaking women consistently disagree on across the five speakers and the four different levels of phonemic accuracy (Table 8.58 through to Table 8.63). The most frequent trait to appear as differentiated in the ten versions is *intelligent* (Table 8.64, P. 262). The Finnish-speaking men and the

Table 8.63 Poor Pronunciation

Trait	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	GB listeners +/- [a]
	Speaker 3: Tape 2					
	SF Women (*N* = 60)		GB Women (*N* = 50)			
Solidarity/Benevolence						
Friendly	fair+	fair+	fair+	fair+	*/–	-
Reliable	fair+	fair+	fair+	fair+	**/–	-
Sympathetic	fair+	fair+	very+	fair+	*/*	+
	SF Men (*N* = 54)		GB Men (*N* = 23)			
Solidarity/Benevolence						
Reliable	fair+/ very+ [b]	fair+	fair+	middle	**/–	-
Sympathetic	fair+	fair+	fair+	fair+	**/*	+
	Speaker 4: Tape 1					
	SF Women (*N* = 59)		GB Women (*N* = 48)			
Status/Competence						
Intelligent	fair+	middle	fair+	fair+	**/–	+
	SF Men (*N* = 54)		GB Men (*N* = 22)			
Status/Competence						
Ambitious	fair-	fair-	fair+	fair+	**/***	+
Educated	fair-/ fair+ [b]	middle	very+	fair+	*/–	+
Intelligent	fair+	middle	fair+	fair+	*/*	+
Solidarity/Benevolence						
Kind	fair+/ very+ [b]	fair+	fair+	fair+	*/–	-

[a] The English-speaking listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the Finnish-speaking listeners. [b] The result is bi-modal.

*** $p < .001$; ** $p < .01$; * $p < .05$.

English-speaking men disagree on an even narrower range of traits than the women (Table 8.65, P. 262), and there is never any disagreement on *friendly*.

In other words, taking the reverse side of these last two tables (Table 8.64 and Table 8.65), all the women and all the men broadly agree amongst themselves, separately, regardless of their mother-tongue, on the vast majority of traits, and what differences do occur would seem to occur with no consistency. On the basis of the current evidence, no traits are particularly prone to differentiation within the two gender groups.

Table 8.64 Frequency of appearance (>= 3) of main traits among the female informants

Trait	Frequency
Intelligent	5
Educated	4
Friendly	4
Rich	4
Self-confident	4
Ambitious	3
Honest	3
Reliable	3
Sympathetic	3

Table 8.65 Frequency of appearance (>= 3) of main traits among the male informants

Trait	Frequency
Intelligent	5
Educated	3
Rich	3

8.3.2. Age of informants

The data was analysed to discover whether the age of the informants acted as a mediating variable to influence the perception the listeners had of the speakers.

(N.B. The age distribution of the informants, and the problems involved in setting the dividing-line for age at 22/23, were presented at the beginning of section 8.1.2.)

For the Finnish-speaking listeners there is basically no difference between the younger and the older listeners in their evaluations of the speakers: there is generally no age-dependent dichotomy among the native Finnish listeners.

Nevertheless, there are a few exceptions, mostly with low significance levels and/or with one or other of the two statistical tests contradicting the other. In every case the older listeners are more positive in their evaluations than the younger ones (Table 8.66, P. 263).

Speaker 5 clearly prompts more age-dependent variation in the informants' responses than the other speakers: two traits with the Excellent Pronunciation (Tape 1) and six (of the fourteen) traits with the Fair Pronunciation (Tape 2). Curiously the traits involved are different for each version, and the two with the Excellent Pronunciation are social traits, whereas five of the six for the Fair Pronunciation are work traits: the older native Finnish listeners upgrade the work traits at this level of mispronunciation, but not at the top level of phonemic accuracy. This conspicuous predominance of Speaker 5 may of course result not from the phonemic accuracy or inaccuracy of the versions she produces, but from the fact that she is the last speaker in the set to be evaluated, or from the possibility that other non-phonemic vocal factors are distorting the perception that the listeners have of her, favourably in the ears of the older listeners, less favourably in the ears of the younger listeners.

Table 8.66 Significant contrasts in perception of speakers (influence of SF informant's age)

Speaker	Version	Trait	sc/sb [a]	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	Older listeners +/- [b]
				Tape 1					
				SF <=22 (*N* = 71)		SF >=23 (*N* = 42) [c]			
1	1 – 'Excellent'	Reliable	sb	fair+	fair+	very+	very+	*/*	+
2	3 – 'Fair'	Hard-working	sc	fair-	fair-	fair-	fair-/middle [d]	*/*	+
2	3 – 'Fair'	Intelligent	sc	fair-	fair-	fair-/fair+ [e]	middle	**/*	+
2	3 – 'Fair'	Rich	sc	middle	middle	middle	middle	*/–	+
4	5 – 'Poor'	Rich	sc	fair-	fair-	fair+	middle	–/*	+
5	1 – 'Excellent'	Honest	sb	very+	fair+	very+	very+	*/–	+
5	1 – 'Excellent'	Trustworthy	sb	very+	very+	very+	very+	–/*	+
				Tape 2					
				SF <=22 (*N* = 70)		SF >=23 (*N* = 43) [f]			
4	2 – 'Good'	Self-confident	sc	fair+	fair+	fair+	fair+	*/–	+
5	4 – 'Fair'	Educated	sc	fair+	middle/fair+ [d]	fair+/very+ [e]	fair+	***/–	+
5	4 – 'Fair'	Hard-working	sc	very+	fair+	very+	very+	*/–	+
5	4 – 'Fair'	Intelligent	sc	fair+	fair+	fair+	fair+	*/–	+
5	4 – 'Fair'	Rich	sc	fair-	middle	middle/fair+ [e]	fair+	*/–	+
5	4 – 'Fair'	Successful	sc	fair-/fair+ [e]	middle	very+	fair+	**/*	+
5	4 – 'Fair'	Sympathetic	sb	fair-	middle	very+	fair+	*/–	+

[a] Trait classification: sc = Status/Competence; sb = Solidarity/Benevolence. [b] The older listeners' perceptions of the speakers are more positive (+) or more negative (-) than the younger listeners. [c] For Speaker 5, *N* = 41. [d] The median falls between the two values. [e] The result is bi-modal. [f] For Speaker 4, *N* = 44.

*** $p < .001$; ** $p < .01$; * $p < .05$.

For the English-speaking listeners, too, there is also—broadly speaking—no difference between the younger and older informants in their perceptions of the speakers.

However, there are some exceptions and more than among the Finnish-speaking listeners: Table 8.67, P. 265 and Table 8.68, P. 266. Most of these exceptions have low significance values, and/or one or other of the two statistical tests contradicts the other. Sometimes the older informants are more positive than the younger ones (as is always the case with the native Finnish listeners), sometimes more negative, but the positive/negative values are always the same for each version (i.e. voice guise).

The most surprising feature here is that the older English-speaking listeners are more *negative* than the younger ones for certain traits involving Speaker 1's Fair Pronunciation (Tape 2), Speaker 3's Good Pronunciation (Tape 1), and Speaker 5's Excellent Pronunciation (Tape 1) and Fair Pronunciation (Tape 2). All the traits for Speaker 5's two guises which display a dichotomy based on age are *downgraded* by the older English-speaking listeners, whereas all such dichotomous traits are *upgraded* by the older Finnish-speaking listeners. The contrast between the Finnish-speaking listeners and the English-speaking listeners is stark here. Once again Speaker 5 presents results which are difficult to interpret.

It is also noteworthy that on only three occasions do the same traits appear as being differentiated according to the age of the listener with both guises that each speaker produces. Speaker 1's Excellent Pronunciation (Tape 1) gains a more positive evaluation for *sympathetic* from the older English-speaking listeners than the younger English-speaking listeners, but her Fair Pronunciation (Tape 2) gains a more negative evaluation for the same trait *sympathetic*. The situation is reversed for Speaker 3: his Good Pronunciation (Tape 1) earns a more negative evaluation for *self-confident* and *rich* from the older English-speaking listeners than the younger English-speaking listeners, whereas his Poor Pronunciation (Tape 2) earns a more *positive* evaluation for the same traits.

One trait, *educated*, appears with a certain—though limited—measure of consistency: with Speaker 2's Fair Pronunciation (Tape 1), Speaker 3's Poor Pronunciation (Tape 2), Speaker 4's Poor Pronunciation (Tape 1), and Speaker 5's Excellent Pronunciation (Tape 1), i.e. on four out of ten possible occasions. The older English-speaking listeners upgrade Speakers 2, 3 and 4 for this trait relative to the younger English-speaking listeners, but downgrade Speaker 5. The trait most sensitive to being mediated by listener age would seem to be *educated*, though the sensitivity is only moderate.

It is striking that few traits here are the same traits as are differentiated according to the informant age of the Finnish-speaking listeners: *intelligent* for Speaker 2's Fair Pronunciation (Tape 1), where the older listeners in both language groups are more positive in their evaluations than the younger ones, and *honest* for Speaker 5's Excellent Pronunciation (Tape 1), where the older Finnish-speaking listeners are more positive but the older English-speaking listeners more negative than their younger counterparts. The Finnish-speaking listeners

Table 8.67 Significant contrasts in perception of speakers, influence of GB informants' age (Tape 1)

Speaker	Version	Trait	sc/sb [a]	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	Older listeners +/- [b]
				Tape 1					
				GB <=22 (*N* = 36)		GB >=23 (*N* = 34)			
1	1 – 'Excellent'	Friendly	sb	fair+	fair+	fair+	fair+	*/–	+
1	1 – 'Excellent'	Self-confident	sc	fair+	fair+	very+	very+	*/**	+
1	1 – 'Excellent'	Sympathetic	sb	fair+	fair+	fair+	fair+	*/–	+
1	1 – 'Excellent'	Trustworthy	sb	fair+	fair+	very+	very+	*/–	+
2	3 – 'Fair'	Educated	sc	fair-	fair-	fair+	fair+	**/*	+
2	3 – 'Fair'	Honest	sb	fair-	fair-/middle [c]	fair+	fair+	***/**	+
2	3 – 'Fair'	Intelligent	sc	fair+	midle	fair+	fair+	*/**	+
3	2 – 'Good'	Rich	sc	fair+	fair+	middle	middle	**/–	-
3	2 – 'Good'	Self-confident	sc	fair+	fair+	fair+	fair+	*/–	-
4	5 – 'Poor'	Educated	sc	fair+	fair+	very+	fair+	**/**	+
5	1 – 'Excellent'	Educated	sc	ex+	ex+	very+	very+	*/–	-
5	1 – 'Excellent'	Honest	sb	very+	very+	very+	fair+	–/*	-
5	1 – 'Excellent'	Intelligent	sc	ex+	ex+	very+	very+	*/–	-
5	1 – 'Excellent'	Rich	sc	very+	very+	middle/fair+ [d]	fair+	**/–	-
5	1 – 'Excellent'	Successful	sc	very+	ex+	very+	fair+/very+ [c]	**/**	-

[a] Trait classification: sc = Status/Competence; sb = Solidarity/Benevolence. [b] The older listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the younger listeners. [c] The median falls between the two values. [d] The result is bi-modal.

*** $p < .001$; ** $p < .01$; * $p < .05$.

Table 8.68 Significant contrasts in perception of speakers, influence of GB informants' age (Tape 2)

Speaker	Version	Trait	sc/sb [a]	Mode	Median	Mode	Median	Mann-Whitney *U* test / Median test	Older listeners +/- [b]
				Tape 2					
				GB <=22 (*N* = 30)		GB >=23 (*N* = 43)			
1	3 - 'Fair'	Honest	sb	fair+	fair+/very+ [c]	fair+	fair+	**/*	-
1	3 – 'Fair'	Reliable	sb	very+	very+	fair+	fair+	–/**	-
1	3 – 'Fair'	Sympathetic	sb	very+	very+	very+	fair+	*/–	-
2	1 – 'Excellent'	Sympathetic	sb	fair-	fair-	fair-	middle	*/–	+
3	5 – 'Poor'	Educated	sc	fair-	middle	fair+	fair+	**/–	+
3	5 – 'Poor'	Rich	sc	fair-	fair-	middle	middle	*/–	+
3	5 – 'Poor'	Self-confident	sc	very-	fair-	fair-	fair-	*/–	+
5	4 – 'Fair'	Ambitious	sc	very+	fair+	fair+	middle	–/*	-
5	4 – 'Fair'	Friendly	sb	very+	very+	fair+	fair+	*/**	-
5	4 – 'Fair'	Self-confident	sc	fair+	fair+	very-/fair-/middle [d]	middle	*/*	-

[a] Trait classification: sc = Status/Competence; sb = Solidarity/Benevolence. [b] The older listeners' perceptions of the speakers are more positive (+) or more negative (-) than those of the younger listeners. [c] The median falls between the two values. [d] The result is multi-modal.

** $p < .01$; * $p < .05$.

and the English-speaking listeners agree overwhelmingly in *not* displaying any age-dependent variation in evaluations, but they disagree whenever they do display age-dependent variations.

To summarise briefly, age would seem to exert hardly any influence at all on the evaluations made by the Finnish-speaking and English-speaking listeners: there are only a few occasions when the older and younger Finnish-speaking listeners differ in their evaluations of the speakers and when the older and younger English-speaking listeners do so. Moreover, there is hardly any overlap between the two language groups either in the traits or in the versions of phonemic inaccuracy which prompt the few divergent evaluations.

8.3.3. The 'shy' (*ujo*) trait

The attempt to capture the so-called typically Finnish trait of 'shyness' by introducing the two trait label pairs *self-confident–hesitant/unsure* (a work trait) and *serious-minded–fun-loving* (a social trait) warrants careful examination. The null hypothesis would claim that a Finnish speaker's pronunciation when speaking English has no effect on a listener's perception of that person's shyness, that accuracy/inaccuracy of pronunciation in English fails to reveal shyness. Thus one would expect no differences in the evaluation of the two versions of my five speakers. However, the results again show a division between work-perceived traits and socially perceived traits. The results of the Finnish-speaking listeners are presented again in Table 8.69, P. 268 and those of the English-speaking listeners in Table 8.70, P. 269.

The Finnish-speaking listeners are clearly influenced by the level of mispronunciation to make a distinction on the *self-confident–hesitant/unsure* adjective pair ($p < .001$ for all but the first speaker), but the distinction they make for the *serious-minded–fun-loving* pair is non-existent for Speaker 3 and borderline for the four other speakers (the two statistical tests fail to support each other for Speakers 1, 2 and 4, and reach only the $p < .01$ level for Speaker 5).

The English-speaking listeners are similarly influenced by the level of mispronunciation, though the statistical tests here give more clear-cut results: the *self-confident–hesitant/unsure* adjective pair is again clearly influenced by the degree of phonemic mispronunciation (except for Speaker 1), but the *serious-minded–fun-loving* adjective pair is not (no significant contrast for Speakers 2, 3 and 4, and the two statistical tests contradict each other for Speakers 1 and 5).

These results with the *serious-minded–fun-loving* trait would seem to suggest that a change in phonemic accuracy makes no difference to a listener's impression of a speaker's serious-mindedness—neither a Finnish-speaking listener's nor an English-speaking listener's. When shyness is interpreted as a social trait, therefore, its perception remains unaffected by the degree of phonemic accuracy.

The same is not true for the *self-confident–hesitant/unsure* trait: the perception of a Finn's self-confidence *is* influenced—to some extent, for both Finnish-speaking and English-speaking listeners—by the phonemic accuracy of the speaker. Here it should be remembered that the speed of delivery, the voice

Table 8.69 Evaluations of Finnish-speaking informants on the 'shyness' traits

Trait	Tape 1 Mode	Tape 1 Median	Tape 2 Mode	Tape 2 Median	Mann-Whitney *U* test/Median test
	Speaker 1				
	'Excellent Pronunciation'		'Fair Pronunciation'		
Self-confident	very+	fair+	fair+	middle	**/**
Serious	very+	very+	very+	fair+	*/–
	Speaker 2				
	'Fair Pronunciation'		'Excellent Pronunciation'		
Self-confident	fair-	middle	very+	fair+	***/***
Serious	fair+	middle	fair+	fair+	*/–
	Speaker 3				
	'Good Pronunciation'		'Poor Pronunciation'		
Self-confident	very+	fair+	fair-	fair-	***/***
Serious	fair+	fair+	fair+	fair+	–/–
	Speaker 4				
	'Poor Pronunciation'		'Good Pronunciation'		
Self-confident	fair-	fair-	fair+	fair+	***/***
Serious	fair+	fair+	fair+	fair+	**/–
	Speaker 5				
	'Excellent Pronunciation'		'Fair Pronunciation'		
Self-confident	very+	very+	fair+	fair+	***/***
Serious	fair+	fair+	fair+	fair+	*/*

*** $p < .001$; ** $p < .01$; * $p < .05$.

quality, the stress patterning, etc. were as identical as possible for the two versions of each speaker that were evaluated. In addition there was an identical absence of stumbling, stuttering and non-verbal vocal hesitations and fillers in both versions. Thus, quite simply, when shyness is interpreted as a work trait, then a higher degree of phonemic accuracy in and of itself both increases the impression of self-confidence and decreases the impression of hesitancy and uncertainty (and hence, by extension, diffidence and self-effacement).

This finding is surprising in that most people would probably associate the vocal signalling of lack of self-confidence not with phonemic accuracy but with disfluency, i.e. false-starts, repetitions, voice quality and fillers like 'um' and 'er,' etc. However, phonemic accuracy would indeed appear to be a key ingredient in the perception of self-confidence. What remains to be discovered is whether various degrees of phonemic accuracy in other varieties of foreign-accented English also affect the perception of a speaker's self-confidence, for then we could answer the question of whether this 'special' Finnish trait of shyness as being dependent on the level of phonemic accuracy is peculiar to Finnish-accented English or not.

Table 8.70 Evaluations of English-speaking informants on the 'shyness' traits

Trait	Tape 1 Mode	Tape 1 Median	Tape 2 Mode	Tape 2 Median	Mann-Whitney *U* test/Median test
	Speaker 1				
	'Excellent Pronunciation'		'Fair Pronunciation'		
Self-confident	fair+	fair+	NA	NA	–/–
Serious	very+	very+	very+	very+	*/–
	Speaker 2				
	'Fair Pronunciation'		'Excellent Pronunciation'		
Self-confident	fair-	fair-	very+	very+	***/***
Serious	fair+	fair+	fair+	fair+	–/–
	Speaker 3				
	'Good Pronunciation'		'Poor Pronunciation'		
Self-confident	fair+	fair+	fair-	fair-	***/***
Serious	very+	fair+	fair+	fair+	–/–
	Speaker 4				
	'Poor Pronunciation'		'Good Pronunciation'		
Self-confident	fair-/+[a]	middle	very+	fair+	***/***
Serious	fair-	middle	fair+	fair+	–/–
	Speaker 5				
	'Excellent Pronunciation'		'Fair Pronunciation'		
Self-confident	ex+	very+	fair+	middle	***/***
Serious	fair+	fair+	fair+	fair+	*/–

[a] The result is bi-modal. *** $p < .001$; * $p < .05$.

The dividing of my selected traits into so-called status/competence traits ('work traits') and solidarity/benevolence traits ('social traits') is clearly artificial (though informal non-rigorous non-extensive pre-test trials did confirm the two groupings), but my splitting of 'shy' into two label pairs with the first in one camp and the second in the other was intentional, motivated, admittedly, by nothing more than a general feeling of dissatisfaction with the pole labels *shy–not shy*. However, the results of this split now allow me to suggest that the perception of stereotypical Finnish shyness—when understood as a work trait—is influenced by the degree of phonemic accuracy in the Finnish-accented English that Finns use. This conclusion is as equally valid for Finnish-speaking listeners as English-speaking listeners: Finns listening to compatriots perceive shyness through a filter of phonemic accuracy, and English-speaking listeners listening to unidentified and unidentifiable foreigners also perceive shyness through a filter of phonemic accuracy. However, when shyness is understood as a social trait, then phonemic inaccuracy plays no part in the perception of that shyness. The interesting thing to do now would be to forget my original reluctance for the bipolar adjective pair *shy–not shy* and see whether the use of this very pair would

yield neutral results with no statistically significant preference for (or towards) one pole or the other, i.e. indicate a confusion among listeners as to how to interpret the trait label 'shy.'

8.3.4. Summary

When the Finnish-speaking listeners evaluate the fourteen traits of the five speakers (Tables 8.33, 8.36, 8.39, 8.42 and 8.45), their perception of the seven status/competence traits is consistently affected across all five speakers by the level of phonemic accuracy, though only the *trustworthy* solidarity/benevolence trait is similarly so affected. The social traits thus remain opaque: they are affected only sporadically, the *honest* trait not being affected at all, *friendly* and *kind* only once each; *friendly*, *kind* and *sympathetic* are apparently not interlinked, the evaluations not always going in parallel. Where there is a dichotomy in the evaluations, the better of the two versions produced by the speakers is always given a more positive rating. Only the Fair Pronunciation and Poor Pronunciation versions earn any evaluations which fall on the negative side of the scale, but this negative bias is not regular (e.g. *ambitious* is affected across all five speakers, but *self-confident* and *successful* across only three speakers).

When the English-speaking listeners evaluate the fourteen traits (Tables 8.34, 8.37, 8.40, 8.43 and 8.46), hesitation and uncertainty would seem to reign for the first speaker (with only the *rich* trait being differentiated with any clarity), though for the remaining four speakers an even clearer division emerges between the status/competence traits (nearly always differentiated) and the solidarity/benevolence traits (generally not differentiated) than that seen with the Finnish-speaking listeners. *Friendly*, *kind* and *sympathetic* are never differentiated and are thus clearly associated with each other; and *trustworthy* or *serious* can be added to these three depending on which test one relies on, the Mann-Whitney *U* test or the Median test respectively. Once again where there is differentiation, the better of the two versions always prompts a more positive rating. Occasionally the informants fail to take a clear stand one way or the other for a trait, but there is no regular pattern here except for the *rich* trait which would seem to be particularly inscrutable at the Fair Pronunciation and Poor Pronunciation level of phonemic inaccuracy. Only infrequently do evaluations fall on or lean towards the negative side of the scales, and there is no regular pattern.

For a comparison of the evaluations of the Finnish-speaking listeners and the English-speaking listeners there are of course ten versions being evaluated (five speakers and two versions each) and hence seventy scales for status/competence traits (seven traits per version x ten versions) and seventy scales for solidarity/benevolence traits (seven traits per version x ten versions). For the majority of the scales there is overall *no* difference between the Finnish-speaking listeners and the English-speaking listeners.

However, there *are* some clear patterns of disagreement between the two mother-tongue groups. The main focus of differentiation is clearly on the status/competence traits, and a concomitant lesser focus on the solidarity/benevolence traits. On the one hand, for twenty-five of the seventy status/competence scales,

the English-speaking listeners are more *positive* in their evaluations than the Finnish-speaking listeners (for one trait they are more negative; and for one trait the contrast is not really applicable but could also be interpreted as being similarly negative). On the other hand, for nineteen of the seventy solidarity/benevolence scales the English-speaking listeners are more *negative* in their evaluations than the Finnish-speaking listeners. In other words the English-speaking listeners tend to upgrade the work traits of the foreign-accented speakers they hear, whereas the Finnish-speaking listeners are less positive towards these same speakers, their compatriots, when evaluating these same traits. However, the Finnish-speaking listeners would seem to be slightly more supportive than the English-speaking listeners when judging the social traits of their compatriots: for the English-speaking listeners the speakers are just foreigners and this fact perhaps tinges their evaluations in a slightly negative way.

There is little trace of gender differentiation either within the Finnish-speaking group of listeners or within the English-speaking group of listeners (though within the Finnish-speaking group there are some mild indications that the women are marginally more positive than the men on social traits). However, when examining all the women together and all the men together, one finds slight signs of differentiation based on mother tongue, there being some differences between the Finnish-speaking women and the English-speaking women at the top level of pronunciation (notably for the social traits, the Finnish-speaking women being more positive), and a few differences between the Finnish-speaking men and the English-speaking men at the bottom level of pronunciation (for the work traits, the Finnish-speaking men being more negative).

Analyses to discover variations based on the age of informants (older v. younger) within each mother-tongue group yield only irregular patterns of association.

9. CONCLUSION

> Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him . . .
>
> —Judges Ch. 12 v. 6.

This study was undertaken as a continuation of the vast number of studies into attitudes and reactions towards accented English, but focuses on foreign-accented English, an area which has received relatively little attention. A trawl of the literature reveals only a few studies in this field, the majority of these looking at Spanish-accented or French-accented English. Studies of Finnish-accented English can be counted on the fingers of one hand.

The aim of the empirical part of the thesis was twofold: to investigate evaluations of certain consonantal mispronunciations of Finnish-accented English, and to study the perception of certain personality traits of Finnish people speaking English with various degrees of Finnish-accentedness. Two experiments were conducted with the use of tape-recordings of Finnish-accented English, the first one being a straight contrast between two ways of saying (and mispronouncing) the same sentence, and the second one requiring the evaluation of the personality traits suggested by the speech of five different Finns reading the same text.

The method used in the first experiment—the presentation of two versions of a sentence made by the same speaker; the dichotomous choice between the two versions imposed on informants (Which is better?)—was specifically chosen to lead to a hierarchy of consonant error and to the subsequent deduction of a priority list for teaching and learning purposes.

The method used in the second experiment was a modification of the matched-guise technique, a widely used and well-tried technique for accessing people's attitudes towards speakers using various accents. The merits and weaknesses of the technique—both acknowledged in Chapter 4—are widely recognised. The overriding advantage of the technique for the present experiment was that it allowed the strict control of variables affecting listeners' perceptions of speaker personality. However, I need to review the MGT's drawbacks in brief here, seeing how they impinge both on my second uniquely designed MGT experiment and on my first *non*-MGT experiment, before summarising the main findings of the two experiments, presenting suggestions for future research and examining the implications for the teaching and learning of pronunciation.

9.1. Problems associated with the MGT in relation to the present studies

First, the ability of a speaker to produce one 'normal' accent plus an 'authentic' alternative accent (or accents) for experimental purposes has been doubted by some critics, for this switch in accents requires multiple and complex segmental and suprasegmental changes, such as changes in consonantal articulations, vowel qualities and durations, coarticulations and assimilations, rhythm, tempo, pitch movement and pitch level. Doubts as to speakers' abilities has led to accusations of artificiality. However, the differences between the two guises that the speakers produced for the present study involved only limited consonantal alterations. Such modest atomistic variation is of course typical of the many interlanguage varieties of foreign-accented speech, since the features to be acquired to transform foreign-accented speech into native-sounding speech are probably taught step by step by most teachers (that is if the teacher devises structured pronunciation lessons with targeted goals) and usually learned step by step by learners (if the learner is actively involved in the pronunciation learning process, consciously self-monitoring production).[1] Any possible doubts about the 'normality' and 'authenticity' of the guises in the present study, therefore, are of a different quality to those that have been expressed for the native accent guises presented in other MGT studies.

The question of phoney accents is further clouded by the fact that the speakers here were speaking a *foreign* language, and neither of the guises produced is the 'normal' voice of the speakers. In addition, some (but not all) of the speakers needed considerable coaching and rehearsing in order to produce some of the required mispronunciations, and this lack of spontaneity could also be taken as evidence for the lack of 'authenticity' in the present guises (even though such preparation leads to no similar worries with, for example, actors, newsreaders, weathermen, preachers, barristers, politicians or lecturers). However, some speakers found rehearsals and repetitions necessary in order to produce *only* the mispronunciations I wanted: for some people holding everything absolutely constant *except for* the required consonantal changes was difficult. For example, when other than the target phonemes became distorted and 'contaminated,' or when identical rhythm was lost, stress placement was changed, pitch movement was altered or smacking of lips was introduced, then either the recording had to be taken again, or else the particular phrase was abandoned and attempts continued with the next phrase (as in Experiment 1), or else the speaker abandoned the project completely (as in Experiment 2). Significantly it was the more *accurate* version (the more native-accent sounding version) that was always the more 'spontaneous' of the two recordings that were finally accepted, and it was the *less* accurate version (the more Finnish-accented sounding version) which required more rehearsing and repetition: the speakers normally spoke English with excellent near-native pronunciation of English, but they could—on demand and relatively easily—revert to, or adopt, specific features of typically Finnish-

[1] Classroom practice and learner behaviour may vary, of course, but see *inter alia*: Prator (1971), Parish (1977), Stevick (1978), Catford (1987), Crawford (1987), Morley (1987, 1994), Celce-Murcia et al. (1996), Pennington (1996).

accented English. It was the isolation of the consonants and the control of everything else that they found tricky. In my opinion the final recordings that were taken for the master tapes were no less valid or authentic as examples of Finnish-accented English than any other 'spontaneous' recordings. The crucial factor was the control of all the variables.

Second, MGT studies have been censured for using subterfuge with their informants, presenting two (or more) guises of the same speaker as if they were two (or more) different speakers. No deception was employed in the present studies. In Experiment 1 the two guises of each phrase played to informants for evaluation were clearly made by the same speaker. In Experiment 2 informants were randomly assigned to two different groups and each group heard one and only one guise of each speaker; thus deception was avoided and the research design focused not on within-group variance but on between-group variance.

Third, the deficient statistical methods employed by many MGT studies have been censured. However, this is a criticism not of the MGT methodology itself but of its implementation. Here in the present studies, considerable pains were taken to keep the samples of informants as large as possible, and non-parametrical statistical tests were specifically chosen for the analyses. Nevertheless, two stochastic shortcomings remain in the present study and mildly weaken the confidence to be placed in the validity of the conclusions I draw: the use of (a) convenience samples for the informants, and (b) a non-random sample of language (mispronunciations) for the recordings (see sections 2.2.3 and 6.1.3). Unfortunately, for practical reasons, these sampling problems were unavoidable.

Fourth, the repetition of the evaluation task which informants have to perform in most MGT studies has prompted concern for the influence of halo effects, response sets and informant fatigue (see e.g. Lee 1971), not to mention concern for unnaturalness. The present studies deliberately used repetition of evaluation tasks, and the influence of this repetition on the results remains obscure. However, the results of Experiment 2 highlight the fact that listeners who have no initial evaluation 'frame,' and no point of reference for making judgements, may need a few 'tries' or 'voices' to establish their own baseline. It could be argued, therefore, that a certain amount of repetition is essential for attitude studies, the issue being how much repetition should be tolerated.

Finally, critics of the MGT condemn the absence of a locus of discourse in most MGT studies. The same shortcoming is present here: the experiments are sanitised 'laboratory' experiments with disembodied voices heard in a 'vacuum,' with no attempt to explain to the informants evaluating the speech or speakers who is talking to whom, where, why or when. It is possible, therefore, that in extended 'real' discourse situations other factors (such as grammar, lexis, idiom, style, topic, purpose, place, gestures, clothing) will sometimes influence the initial perception of both speech and speakers, confirming or cancelling out the perception created by the consonantal mispronunciations under investigation here. This redundancy and defeasibility introduced by other discourse factors needs to be borne in mind when trying to evaluate the significance of the present results, but it should be remembered that even though listeners depend on many

different signals for evaluating speech or a speaker, we have little information to tell us whether the mispronunciation of foreign-accented speech is of major or minor significance in the overall process of impression formation taking place in any particular discourse situation.[2] In some situations the effect of foreign accent on person perception may be of little import (e.g. some emergency situations, some public address system announcements, some encounters between supermarket checkout staff and customers): the information imparted may be far more important than the vehicle which conveys it. But in other situations it may be of enormous import (e.g. job interviews, media interviews, business meetings, formal presentations, lectures, jury service, TV advertisements).[3] Unfortunately it is not possible at the moment to predict which discourse situations lead to greater *or* lesser attention being attached to foreign-accented mispronunciation (e.g. conversations with greater *or* lesser levels of dynamic interaction between speaker and listener; dialogues with equal levels of interlocutant participation *or* monologues with zero involvement of the listener; discourse events with face-to-face *or* non face-to-face talk). Moreover, it is not yet possible to say what the 'life' of accent perception is: how long do impressions based on accent survive alongside other discourse signals and other acquired information? Given this uncertainty, it is impossible to say how the lack of a locus of discourse in the present study affects the wider interpretation of the results, but locus of discourse is of course an issue which needs to be addressed in future studies.

9.2. The overall findings

As stated at the outset of this thesis, research design imposes restrictions on the validity of the conclusions and the generalisations that can be drawn. However, within the constraints of the two different research frameworks adopted for the present complementary experiments the answers to the research questions posed in Chapter 5 can confidently be summarised as follows:

Experiment 1:

1. Mispronounced RP / v /, when articulated as [ʋ], a voiced labiodental frictionless continuant, is not liked by British people, though Finnish-speaking people and Swedish-speaking Finns think that the mistake is not at all serious. By contrast fronted / tʃ / and non-aspirated / p / are relatively well tolerated by British people.
2. The mispronunciation tolerance hierarchies of the Finnish-speaking, Swedish-speaking and English-speaking listeners are not the same, the first group being slightly more divergent from the English-speaking listeners than the second group.

---

[2] One study with native English accents (see Reid 1977:27–8; also Wells 1982:15) suggests that when trying to guess a compatriot's class affiliation, British informants place more value on speech (and this presumably means primarily accent) than on place of abode, friends, job, school attended, money-spending habits, income, clothing and car, though Reid stresses that there are problems in interpreting the results of this study.

[3] See the report in *The Times* 4.4.2000 concerning an Iranian-born woman who claimed she was ousted from her job of manageress at Harrods because she did not speak 'BBC English.' She presumably spoke English with a foreign (Farsi) accent.

3. Neither the gender nor the age of the English-speaking listeners has a moderating effect on the evaluations of mispronunciation, though there is some indication that both the English-speaking men and the younger informants are reluctant to be critical where *female* voices are concerned. Similarly the gender and age of the Finnish-speaking listeners has no moderating effect on the evaluations of mispronunciation, though some evidence exists for claiming that the Finnish-speaking men are overall relatively more indecisive than the Finnish-speaking women, and that the younger informants are indecisive where *male* voices are concerned.
4. Despite the impossibility of isolating gender of voice as a variable, it would seem that gender of voice probably has only a limited effect on listeners' evaluations of mispronunciation. Nevertheless, English-speaking listeners appear to be less tolerant of mispronounced / ʃ / in female voices than male voices.

Experiment 2:

1. Both English-speaking listeners and Finnish-speaking listeners find it difficult to estimate the age of Finns speaking English, a foreign language, with a Finnish accent. Generally speakers with higher phonemic accuracy are thought to be older than they actually are, and those with lower phonemic accuracy younger than they are.
2. Both mother-tongue informant groups appear to have an established preconception of what constitutes 'good' and 'bad' pronunciation. However, the Finnish-speaking listeners would seem to be more demanding than the English-speaking listeners for they are more negative in their evaluations.
3. Both the Finnish-speaking listeners separately as a group and the English-speaking listeners separately as a group perceive the status/competence traits of a speaker in different ways according to the speaker's level of phonemic accuracy, rating these traits more positively with better pronunciations. The solidarity/benevolence traits are only sporadically differentiated in a similar fashion.
4. There are few differences overall between the two mother-tongue groups in their evaluations of the traits, both broadly perceiving the speakers in the same ways, though the English-speaking listeners are slightly more positive than the Finnish-speaking listeners for some of the status/competence traits, and very slightly more negative for the solidarity/benevolence traits.
5. Informant gender generally has no effect on the perception of the speakers, apart from slight signs of Finnish-speaking women being more positive than English-speaking women for social traits at the top levels of pronunciation, and Finnish-speaking men being more negative than English-speaking men for work traits at the bottom levels of mispronunciation.
6. Informant age only has an infrequent and very irregular effect upon the perception of Finnish-accented speakers of English.

9.3. Future research

These answers to my research questions and the conclusions I have drawn still need to be confirmed with follow-up studies (preferably conducted in the same and different ways): the fact that a single instantiation of accent X produces result Y with informant sample Z is merely suggestive and not conclusive evidence for conclusion C. This need for corroborative results—common to all studies of attitudes—is underlined by the small number of self-contradictory results that were obtained for Experiment 1 (see e.g. Table 7.3 in section 7.2.1). However, the number of such results was very small, and these contradictions should not be confused with those results where the statistical significance levels suggest that one mispronunciation, or one level of overall mispronunciation, has no effect on the listeners' perceptions of the speech or the speaker (see e.g. Table 8.33 in section 8.3).

In addition these results naturally prompt further research questions and suggest further lines of enquiry for future sociophonetic studies into foreign-accented speech and foreign speaker perception. In particular, does the position of a mispronounced phoneme in word, syllable and sentence structure affect its salience? What results would be obtained with other Finnish-accented consonants or vowels as the independent variables? How do other combinations of number, frequency and co-occurrence of tokens of mispronounced phonemes affect listeners' perceptions of a speaker? Do infelicities in Weak Forms, juncture, tempo, rhythm, intonation and voice quality outweigh those in consonantal accuracy? Are there just a few clearly marked 'steps' in reactions to Finnish-accented English, or is there a steady cline of innumerable slowly ever-changing 'states,' i.e. is the perceived broadness–mildness of a foreign accent limited to three, or four, or more recognisable steps, or is it a continuum?

More studies are essential, both for corroboration and furtherance of the present results, but the problems involved in undertaking sociolinguistic surveys, particularly in recruiting informants (see section 6.2.2. above), should not be underestimated. In Finland I was fortunate in benefiting from the willing support of both my own students and those of my colleagues. In Britain, however, the finding of informants for this study was a horrendous nightmare—no exaggeration—for which nothing and no one had prepared me. I have always found students (generally speaking, 99.9% of the time) to be extremely pleasant young people, unrecognisable as the hooligans and yobs repeatedly discussed and frequently vilified in the press. However, the giving of 20–25 minutes of time to help me conduct a sociophonetic survey seemed to be a huge sacrifice for nearly all my potential informants in Britain and *many* people refused to help me, presumably because it was *too* huge a sacrifice. Without captive audiences, where students are asked to participate in an experiment during class time, and/or where the experiment is presented as practical experience with obvious relevance to their own studies, prospective investigators should sit down and take a long careful think before embarking on surveys similar to the present ones. The research design and statistical procedures to be used will determine the size of the sample necessary for a study, but whatever the minimum sample size aimed

at, a study of attitudes presupposes the availability of willing informants: while I had such informants in Finland, I did not have them in Britain. Nor did I always have the full-hearted support of teachers, lecturers and professors, though to those who did help me I convey my sincerest heartfelt thanks. Experiences such as I had easily lead to severe depression, and prospective students in this field should be warned: I strongly advise everyone proposing to undertake sociolinguistic attitudinal survey work to think again. Or else make sure you have a wide network of supporters, enormous resources of time, energy and money, and above all determination and stamina, what Finns call *sisu*.

Two decisions that I took concerning my research designs necessarily meant that I would need large numbers of informants. First I refused to treat my data as interval scale data (there is no zero, and the intervals of measurement are neither equal nor describable), preferring to consider my measuring instruments as nominal scales (Experiment 1) and nominal/ordinal scales (Experiment 2). Second, unlike the vast majority of previous sociophonetic studies using the MGT for examining the perception of accented speakers, I decided for Experiment 2 to exploit a between-subjects design rather than a within-subjects design: the voice guises being compared with each other were evaluated by two different groups of people—*not* by the same people—and the null hypothesis was that these two groups who listened to the different versions of the same speaker would not differ in their responses to that speaker. These two decisions together—although well-founded—inevitably meant that large numbers of informants were required for the non-parametric statistical tests to provide results of any significance. The problems were enormous, but my research designs are more solid theoretically, and my results more robust statistically, than those—I would claim—of many previous studies. This point should weigh heavily when attempting to gauge the credence, and hence value, to be afforded the present findings.

9.4. Implications for the teaching and learning of pronunciation

More work needs to be done in the field of attitudes towards foreign accents, focusing, as Pennington (1994:104) says, on those features of a non-native's phonology that draw most attention from, and cause the most negative reactions in, native listeners. However, the evidence of the present empirical studies is sufficient for teachers of English pronunciation to Finnish-speaking and Swedish-speaking learners in Finland (perhaps even to foreign learners generally) to decide on some of the priorities to set in the classroom for English phoneme proficiency. The unexpected fact is that mispronounced / v / (when articulated as [ʋ]) is not liked by British people at all, even though the learners themselves think that the mistake is not at all serious. Thus while common sense and practical experience promotes Finns to teach and learn e.g. the alveolar and palatoalveolar RP sibilants before the RP palatoalveolar affricates, and the alveolar (not dental) articulation of RP / t / before the nasal release of RP / t / in such words as [ˈbʌtˡn̩], so, for perceptual and impression-forming reasons, the present results suggest that mispronounced / v / should be prioritised before the mispronounced fronting of / tʃ / and the mispronounced non-aspiration of / p /.

In addition the findings are clear enough to warrant warning Finnish-speaking learners that even a small degree of Finnish-accentedness in their English leads to a noticeable deterioration in certain of their personality traits as perceived by British listeners, notably in their status/competence traits: the perception of an unidentified foreigner's intellectual, professional and work skills and abilities varies according to their standard of pronunciation. In other words, British people evaluate a foreigner's friendliness, kindness, honesty and so on in a constant manner—the evaluation being independent of the person's standard of phonemic pronunciation—whereas they find it impossible to conceive that a foreigner who pronounces English *badly* can be as intelligent, educated, successful and so on, or as potentially as good a business partner, as someone who pronounces English *well*.

While the perception of their social traits may seem to be of more immediate importance to students than the perception of their work traits, this discovery is potentially of enormous significance in the post-school and post-university work environment. The discovery perhaps also justifies a minor adjustment to my theoretical model of foreign speaker perception (Figure 4.1, P. 102) such that 'solidarity/benevolence' in the evaluation column is in brackets, thus making sure that the variation in the perception of these traits is understood as being less than that of the status/competence traits.

The present results should also prompt writers of textbooks on English phonetics to re-examine their views on error gravity and to reconsider their suggestions for the priorities to set in pronunciation for minimum general intelligibility, minimum irritation and high acceptability. It is time in L2 pronunciation teaching to talk about the *cosmetic significance* of foreign accents: a foreign accent may be comprehensible and intelligible, may also be acceptable, but it also signals speaker personality traits, and these traits are perceived by listeners more positively or more negatively in the same cosmetic way as clothing, general appearance and personal behaviour.

The tools available to teachers to sensitise students generally to accent are limited—as also is the time available. 'Do this because I say so!' is not likely to produce very good results. A better way may be to inform learners, particularly those of around twenty years of age, of one major unexpected result in the present study and let them draw their own conclusion: bad pronunciation prompts serious under-aging, and good pronunciation over-aging, i.e. impressions, respectively, of immaturity and maturity. There is also perhaps room for greater reference to, and analysis of, the use of accents for character and personality portrayal in TV programmes, plays, films and commercials—both in the L1 and in the L2. For example, how do producers and actors manipulate us, the audience, with their uses of native and foreign accents? What sort of voice does a newsreader have in one's L1 and in the L2, or a weather forecaster, or the good guys and the bad guys in a film, or the person trying to persuade us to buy a particular brand of soap-powder? Why—in Britain—is a deviant 's' acceptable in an actor like Sean Connery, but totally unacceptable in an actress like Judi Dench? Why—in Finland—does Neil Hardwick's English-accented Finnish

apparently appeal to so many TV viewers?[4] And—most importantly—why does one foreigner being interviewed on the news in Britain in English (or in Finland in Finnish) *sound* so clever, honest and successful while another foreigner *sounds* so untrustworthy, lazy and stupid?

Since teaching usually presupposes testing, the ultimate sanction to help instil the message of the importance of accent is presumably pronunciation testing. Indeed the implications of the present results for the testing of pronunciation, for the evaluation of mispronunciation in general (error gravity), are considerable: / v, dʒ / and / θ / (and / ʃ / in women) are more serious perceptually than fronted / tʃ / and unaspirated / p /. Unfortunately the present experiments leave open the position, for example, of non-aspiration in the other fortis plosives, of mispronounced / ð / and of dentalised / t / in the phoneme hierarchy, of 'strongly' pronounced Weak Forms, of rhythm and intonation, of mispronunciation combinations and frequency, and more studies are necessary to map out the whole area in more detail. The present studies also offer no information as to whether particular phonemic inaccuracies are always considered by native English listeners as more or less tolerable regardless of the mother tongue of the speaker, and regardless of any other accompanying inaccuracies. However, the present results clearly suggest that the tolerances of various features of mispronunciation are not always predictable, are not always what we might expect, however experienced and knowledgeable we are as teachers and linguists. They also indicate that the viewpoint of the evaluator—native listener v. non-native listener—may be significant: there would appear to be no universal hierarchy of errors (see McCretton & Rider 1993).

Testing usually goes hand in hand with teaching, but we are still a long way from being able to say what minimum standards of pronunciation proficiency the teachers of English in Finland—and elsewhere around the world—should demand of their students for particular purposes in particular real-life situations. Nor, surprisingly, do we know what minimum standards of pronunciation teachers should demand of native speakers of English when *they* (English speakers) speak French, German or whatever. Little work has been done in this area. Churchill's accent when he spoke French is a legend, notoriously dreadful, as stylised as his accent in English, but how did the French react to it? The present British Prime Minister, Tony Blair, also accepts the chance to speak French when in France, but how well does his accent go down? And as for the NATO Spokesman who handled the media during the Kosovo crisis early in 1999, Jamie Shea, remembering that his native London accent was not always appreciated by his compatriots and other native English listeners, how did the Francophone world respond to his English-accented French?

While teachers may demand certain minimum pronunciation standards, learners may of course remain unconvinced of the need for such standards, feel despair at ever achieving them, or refuse to worry about them at all. The initial

---

[4] Neil Hardwick, an Englishman, is a well-known popular TV personality in Finland, with his own late-night chat show (in Finnish) and his own series of travel programmes (in Finnish). He has also written autobiographical works (in Finnish), describing his fits of depression.

struggle involved in using a different set of muscles to articulate the TL's phoneme inventory is huge, and the effort required to learn and automatise all the allophonic variations (to say nothing of all the suprasegmental details) is enormous. Moreover, the move towards a more native-like accent is psychologically difficult, entailing a loss of (or change in) self, social and national identity: the learner literally has to acquire a new voice, and this requires a new perception of oneself and a new approach to self-presentation. This establishing of a new second identity for use within the TL is hindered by the fact that the set of rules for appraising accents in the TL are obviously different from the set in the mother tongue and yet completely unknown: the rules for the perception of personality through accent in different languages are not the same; the value systems encoded in the accents of different languages are not the same; and yet neither teachers nor learners have a handbook to refer to to discover what the new rules and value systems are. Most learners, therefore, presumably find it impossible to ensure the congruence of their accent with the self-image which they wish to present, and realising that they lack accent flexibility in the TL, they are probably also frustrated at not being able to accommodate their accent (whether convergently or divergently) according to speech register, subject and social situation.

Most learners—except for those students with a genuine deep interest in languages—probably couldn't care less about their accent in the L2: the acquisition of a near-native accent is thought to be impossible; motivation is lacking; goals are set at the minimum for comprehension (Pennington 1994:104–5). Nevertheless, foreign language learners are well aware that in their mother tongues an accent is noticeable and identifiable, and that in certain situations it may be an advantage or a stigma:

> The way we say a thing, or fail to say it, fail to learn to say it, has an importance in life that it is impossible to overstate—a far-reaching importance, as the very hinge of the relation of man to man. . . . [Pronunciation] destitute of any approach to an emission of the consonant . . . becomes a mere helpless slobber of disconnected vowel noises. (Henry James. Speech delivered 8.6.1905, printed in *Harper's Bazaar*, Pp. 40–1.)

School children in particular are well-known for sometimes going to extremes with accents, taking an accent as a signal for inclusion or exclusion:

> An accent is as striking as outlandish dress or physical deformity. Language is an instrument of both communication and ex-communication. It makes possible intimacy, also exclusiveness. (Waller 1987:2)

For example, in Finland people recognise and react to Rauma-accented, Lapland-accented and Helsinki-accented Finnish, to Finnish Swedish and Swedish Swedish, though curiously enough few Finns seem ready to admit openly that their reactions are anything but neutral, a process of simple regional identification. Research into the attitudes of Finns towards Finnish accents is still basically in its infancy.

In the English-speaking world, on the other hand, sociophonetic awareness and research has a longer history. In England, for example, it has long been

known that people respond positively towards Welsh, Irish and Scottish accented Englishes, two hundred years ago James Boswell stating that

> a small intermixture of provincial peculiarities may, perhaps, have an agreeable effect, as the notes of different birds concur in the harmony of the grove, and please more than if they were all exactly alike. I could name some gentlemen of Ireland, to whom a slight proportion of the accent and recitative of that country is an advantage. The same observation will apply to the gentlemen of Scotland. (Boswell. *Life of Johnson*. Everyman's Library edition 1967. Vol. 1, P. 417)

More recently a Scottish accent has been recognised as the customer-friendly voice admirably suited to anonymous telephone help-line speakers.[5] Fronting of the RP alveolar sibilant / s / to [θ] or [s̪] is thought to be camp when produced by men (Wells 1982:22), and consistent retraction of the RP alveolar sibilants, particularly in women, is interpreted as drunkenness:

> 'Sho kind,' said the woman genteelly. 'Sho shorry to cauzhe all thish trouble.' She had evidently been tanking up at home before opening time. . . . 'Thezhe thingzh,' she said, 'happen in the besht of familiezh.' (Burgess 1982:24)

Glottal reinforcement or replacement of syllable-closing plosives is thought to be ugly, even 'uneducated' (Wells 1982:35, 261). Rhotic British accents are generally identified by English people as being 'rural' and 'quaint' (Trudgill 1974a: 20–1; Wells 1982:30–1). Among English people persistent dentalisation of the RP alveolar fortis plosive / t / is widely recognised as a lisp, or in men as a sign of effeminacy (witness the long tradition for the portrayal of effeminacy in the cinema), and among Americans the RP alveolar plosive / t / instead of the General American alveolar tap [ɾ] is considered to be 'artificial, prissy, effeminate' (Wells 1982:250).

Everyone notices and responds to accents, and with nothing but the voices of speakers to help them, listeners from the same language and cultural background regularly reach the same conclusions concerning the personality of the speakers they hear, deluding themselves as to their own infallibility in recognising the personality of speakers from purely phonetic detail (Laver 1968:50–1). But neither Finnish nor British (nor any) foreign-language learners have any way of knowing how attitudes towards both native accents and their speakers 'translate' into the FL situation (a point first stressed by Sapir in 1927). No learners can put themselves in the position of the TL listener, can even begin to understand the complex combination of affective, belief and conative reactions that their foreign accent stirs in the listener. They have little to tell them what degree of foreign accent will, for example, prompt a better or worse response from TL listeners, provoke a more or less helpful and co-operative reaction in their interlocutors, lead to, or not lead to, a job offer in the target language country, or suggest to TL listeners that they are hard-working, intelligent and trustworthy, or unambitious, poor and cruel.

In other words language learners may well realise that foreign language pronunciation/mispronunciation has repercussions on how they are viewed, on how

---

[5] See e.g. the report in *The Times* 24.4.96 announcing Tesco's decision to move its customer service centre from Hertfordshire—where the employees speak 'Estuary English'—to Dundee.

they are treated and on how well they themselves can influence people's reactions to them, but because they have no information to help them, or feel the job of accent improvement is impossible, or feel the rewards are not worth the investment, they prefer to ignore the whole issue.

The present study offers some valuable information to both teachers and learners in Finland concerning prioritisation in pronunciation teaching and learning, and concerning person perception in Finnish-accented English. It also provides some new insight into how English-speaking listeners perceive speakers with a non-identifiable foreign accent. However, much work remains to be done. Far more information is necessary. We still need more facts.

APPENDIX A : PHONETIC TRANSCRIPTIONS OF THE RECORDINGS FOR EXPERIMENT 1

Below is a transcription of the sets of pairs of sentences for Tape 1 and Tape 2. The transcription is basically phonemic (using Cruttenden's set of symbols), though the misarticulations are presented as underlined symbols which, strictly speaking, are neither phonemic nor phonetic. These underlined symbols should be interpreted in the following way:

/ υ / = voiced labiodental frictionless continuant.
/ s̲ / = voiceless postalveolar sibilant.
/ p̲ / = voiceless unaspirated bilabial plosive.
/ t̲ / = voiceless unaspirated dental plosive.
/ t̲s̲ / = voiceless postalveolar affricate (with no lip-rounding).
/ d̲z̲ / = voiced postalveolar affricate (with no lip-rounding).

Occasionally a speaker may use a Strong Form where a native speaker would probably have used a Weak Form, a rhotacised non-RP vowel may sometimes intrude, and the word-stress placement may sometimes be unsatisfactory. Such features are indicated, though other features, such as the intrusion of an unexpected glottal stop, jerky sentence rhythm and slightly non-RP vowel qualities, are *not* transcribed.

Tape 1

♠ = Male speaker; ♥ = Female speaker

1. — / v (♠) / ðə ˈvæn hæz ˈpɑːst ðə ˈvɪlɪdʒ ˈpʌb /
/ ðə ˈʋ̲æn hæz ˈpɑːst ðə ˈʋ̲ɪlɪdʒ ˈpʌb /

2. dʒ / tʃ (♥) / ˈd̲z̲ɒn eɪt ə ˈlʌntʃ ðət wəz ˈmʌtʃ tuː ˈlɑːd̲z̲ /
/ ˈdʒɒn eɪt ə ˈlʌnt̲s̲ ðət wəz ˈmʌt̲s̲ tuː ˈlɑːdʒ /

3. θ / p (♠) / ˈt̲auzn̩dz əv ˈpiːpl̩ əd ˈt̲ɔːt ɪt ˈpɜːfɪkt /[1]
/ ˈθauzn̩dz əv ˈp̲iːpl̩ əd ˈθɔːt ɪt ˈp̲ɜːfɪkt /

---

[1] The recordings for this pair are not entirely satisfactory, and the data for this pair should perhaps have been discarded.

4. tʃ / ʃ (♠) / ðə ˈkəʊts hæd te ˈriːts ɪts destɪˈneɪʃn̩ bɪfɔː ðə ˈʃəʊ ˈstɑːtɪd /
/ ðə ˈkəʊtʃ hæd tə ˈriːtʃ ɪts destɪˈneɪsn bɪfɔː ðə ˈsəʊ ˈstɑːtɪd /

5. θ / — (♥) / ðə wəz ˈnʌtɪŋ ðə ˈtʃæmpjən kəd ˈduː | ˈnʌtɪŋ tə ˈtʃeɪndʒ /
/ ðə wəz ˈnʌθɪŋ ðə ˈtʃæmpjən kəd ˈduː | ˈnʌθɪŋ tə ˈtʃeɪndʒ /

6. tʃ / dʒ (♥) / ˈfɔːtsnətlɪ ˈnɒt mʌts ˈdæmɪdʒ wəz ˈdʌn tə ðə ˈhedʒ /
/ ˈfɔːtʃnətlɪ ˈnɒt mʌtʃ ˈdæmɪdz wəz ˈdʌn tə ðə ˈhedz /

7. — / p (♠) / ðɪ ɪmˈpreʃn̩ ðə ˈpeɪntɪŋ ˈgeɪv ɒv ðə ˈʃip ɒn ðə ˈwɔːtə wəz ˈpɜːfɪkt /
/ ðɪ ɪmˈpreʃn̩ ðə ˈpeɪntɪŋ ˈgeɪv ɒv ðə ˈʃip ɒn ðə ˈwɔːtə wəz ˈpɜːfɪkt /

8. v / tʃ (♥) / ðə ˈwɪərɪ ˈʊɪkə ˈentəd ɪz ˈtʃɜːtʃ /
/ ðə ˈwɪərɪ ˈvɪkə ˈentəd ɪz ˈtsɜːts /

9. θ / ʃ (♠) / hɪ ˈtɔːt ðə ˈʃeɪp əv ðə ˈtɪŋ wəz ˈʃɒkɪŋ /
/ hɪ ˈθɔːt ðə ˈseɪp əv ðə ˈθɪŋ wəz ˈsɒkɪŋ /

10. tʃ / p (♥) / ðə ˈpɪŋk ˈwɒts wəz ˈmʌts tuː ɪkˈspensɪv /
/ ðə ˈpɪŋk ˈwɒtʃ wəz ˈmʌtʃ tuː ɪkˈspensɪv /

11. dʒ / θ (♠) / ˈjes | ðə wəz ˈsʌmθɪŋ ɒn ðə ˈbrɪdz | ˈsʌmθɪŋ ˈlɑːrdz /[2]
/ ˈjes | ðə wəz ˈsʌmtɪŋ ɒn ðə ˈbrɪdʒ | ˈsʌmtɪŋ ˈlɑːrdʒ /

12. v / ʃ (♠) / ðə ˈʊəʊəl ˈsaʊndz əv ˈɪŋglɪʃ ɑː ˈʊerɪ ˈspeʃl̩ /
/ ðə ˈvəʊəl ˈsaʊndz əv ˈɪŋglɪs ɑː ˈverɪ ˈspesl̩ /

13. p / dʒ (♥) / ðə ˈlɑːdʒ ˈhelpɪŋ əv ˈpɒrɪdʒ wəz ˈpaɪpɪŋ ˈhɒt /
/ ðə ˈlɑːdz ˈhelpɪŋ əv ˈpɒrɪdz wəz ˈpaɪpɪŋ ˈhɒt /

14. tʃ / θ (♥) / ðə wəz ˈnʌθɪŋ ðə ˈtsæmpjən kəd ˈduː | ˈnʌθɪŋ tə ˈtseɪndʒ /
/ ðə wəz ˈnʌtɪŋ ðə ˈtʃæmpjən kəd ˈduː | ˈnʌtɪŋ tə ˈtʃeɪndʒ /

15. ʃ / v (♠) / ðə ˈseɪp əv ðə ˈvest wəz ˈsɒkɪŋ n̩d ˈvʌlgə /
/ ðə ˈʃeɪp əv ðə ˈʊest wəz ˈʃɒkɪŋ n̩d ˈʊʌlgə /

16. — / dʒ (♥) / ˈjes | ðə wəz ˈsʌmθɪŋ ɒn ðə ˈbrɪdʒ | ˈsʌmθɪŋ ˈlɑːdʒ /
/ ˈjes | ðə wəz ˈsʌmθɪŋ ɒn ðə ˈbrɪdz | ˈsʌmθɪŋ ˈlɑːdz /

17. θ / tʃ (♠) / ðə ˈtʃæmpjən tʊk ə ˈbret | ˈðen sed ˈsʌmtɪŋ tə ˈtʃeɪndʒ ðə ˈsʌbdʒɪkt /
/ ðə ˈtsæmpjən tʊk ə ˈbreθ | ˈðen sed ˈsʌmθɪŋ tə ˈtseɪndʒ ðə ˈsʌbdʒɪkt /

18. — / ʃ (♠) / ðə ˈvaʊəl saʊndz əv ˈɪŋglɪʃ ə ˈverɪ ˈspeʃl̩ /
/ ðə ˈvaʊəl saʊndz əv ˈɪŋglɪs ə ˈverɪ ˈspesl̩ ‖

---

[2] The speaker noticeably rhotacises the / ɑː / in large here, but as he does so in both versions, I am not too worried about it.

19. p / θ (♥) / aɪ ˈθɪŋk ðeə wə ˈθɜːtɪfɔː ˈpəʊstəz ɪn ðə ˈpʌb /
/ aɪ ˈtɪŋk ðeə wə ˈtɜːtɪfɔː ˈpəʊstəz ɪn ðə ˈpʌb /

20. tʃ / v (♥) / ðə ˈtsɜːts ˈvɪkə wəz ˈnɒt ˈrɪəlɪ verɪ ˈrɪts ət ˈɔːl /
/ ðə ˈtʃɜːtʃ ˈʊɪkə wəz ˈnɒt ˈrɪəlɪ ʊerɪ ˈrɪtʃ ət ˈɔːl /

21. p / ʃ (♠) / ðɪ ɪmˈpreʃn̩ ðə ˈpeɪntɪŋ ˈgeɪv əv ðə ˈʃip ɒn ðə ˈwɔːtə wɒz ˈpɜːfɪkt /
/ ðɪ ɪmˈpresn̩ ðə ˈpeɪntɪŋ ˈgeɪv əv ðə ˈsip ɒn ðə ˈwɔːtə wɒz ˈpɜːfɪkt /

22. dʒ / θ (♥) / ðə wəz ˈsʌmθɪŋ ˈlɑːdz ɒn ðə ˈsteɪdz əv ðə ˈθɪətə /
/ ðə wəz ˈsʌmtɪŋ ˈlɑːdʒ ɒn ðə ˈsteɪdʒ əv ðə ˈtɪətə /

23. v / p (♠) / ðə ˈʊæn hæz ˈpɑːst ðə ˈʊɪlɪdʒ ˈpʌb /
/ ðə ˈvæn hæz ˈpɑːst ðə ˈvɪlɪdʒ ˈpʌb /

24. dʒ / ʃ (♠) / hɪ ˈpɒlɪʃt ðɪ ˈedz əv ɪz ˈʃerɪfs ˈbædz /
/ hɪ ˈpɒlɪst ðɪ ˈedʒ əv ɪz ˈserɪfs ˈbædʒ /

25. p / tʃ (♥) / ðə ˈbʊtʃə ˈpaundɪd ðə ˈpɔːk ˈtʃɒp /
/ ðə ˈbʊtsə ˈpaundɪd ðə ˈpɔːk ˈtsɒp /

26. θ / v (♥) / ðə ˈtɜːtɪ ˈvɪlənz əd ˈtʌmpt ɒn ðə ˈvæn /
/ ðə ˈθɜːtɪ ˈʊɪlənz əd ˈθʌmpt ɒn ðə ˈʊæn /

27. p / — (♠) / ˈθauzn̩dz əv ˈpiːpl̩ hæd ˈθɔːt ɪt ˈpɜːfɪkt /
/ ˈθauzn̩dz əv ˈpiːpl̩ hæd ˈθɔːt ɪt ˈpɜːfɪkt /

28. dʒ / v (♥) / hɪ ɪnˈdzɔɪd ˈdraɪvɪŋ ɪz ˈvæn θruː ðə ˈvɪlɪdz /
/ hɪ ɪnˈdʒɔɪd ˈdraɪʊɪŋ ɪz ˈʊæn θruː ðə ˈʊɪlɪdʒ /

29. ʃ / p (♠) / ðə ˈsɒp hæd tə peɪ fər ˈɔːl ðə ˈpɑːsl̩ z əv ˈsuːz /[3]
/ ðə ˈʃɒp hæd tə ˈpeɪ fər ɔːl ðə ˈpɑːsl̩ z əv ˈʃuːz /

30. tʃ / — (♥) / ðə ˈkəʊts hæd tə ˈriːts ɪts destɪˈneɪʃn̩ bɪfɔː ðə ˈʃəʊ ˈstɑːtɪd /
/ ðə ˈkəʊtʃ hæd tə ˈriːtʃ ɪts destɪˈneɪʃn̩ bɪfɔː ðə ˈʃəʊ ˈstɑːtɪd /

Tape 2

♠ = Male speaker; ♥ = Female speaker

1. — / tʃ (♠) / ðə ˈkəʊtʃ hæd tə ˈriːtʃ ɪts destɪˈneɪʃn̩ bɪfɔː ðə ˈʃəʊ ˈstɑːtɪd /
/ ðə ˈkəʊts hæd tə ˈriːts ɪts destɪˈneɪʃn̩ bɪfɔː ðə ˈʃəʊ ˈstɑːtɪd /

2. p / ʃ (♥) / ðə ˈʃɒp hæd tə ˈpeɪ fər ɔːl ðə ˈpɑːsl̩ z əv ˈʃuːz /
/ ðə ˈsɒp hæd tə peɪ fər ˈɔːl ðə ˈpɑːsl̩ z əv ˈsuːz /

---

[3] The rhythm of these two sentences is not quite the same, and therefore the / ʃ - p / contrast is not the only factor here which may affect the response of informants.

3. v / dʒ (♠) / hɪ ɪnˈdʒɔɪd ˈdraɪʋ̞ɪŋ hɪz ˈʋ̞æn θruː ðə ˈʋ̞ɪlɪdʒ /
/ hɪ ɪnˈdz̞ɔɪd ˈdraɪvɪŋ hɪz ˈvæn θruː ðə ˈvɪlɪdz̞ /

4. — / p (♥) / ˈθauzn̩dz əv ˈpiːpl̩ hæd ˈθɔːt ɪt ˈpɜːfɪkt /
/ ˈθauzn̩dz əv ˈp̞iːpl̩ hæd ˈθɔːt ɪt ˈp̞ɜːfɪkt /

5. v / θ (♠) / ðə ˈθɜːtɪ ˈʋ̞ɪlənz əd ˈθʌmpt ɒn ðə ˈʋ̞æn /
/ ðə ˈt̪ɜːtɪ ˈvɪlənz əd ˈt̪ʌmpt ɒn ðə ˈvæn /

6. tʃ / p (♠) / ðə ˈbʊt̪sə ˈpaundɪd ðə ˈpɔːk ˈt̪sɒp /
/ ðə ˈbʊtʃə ˈp̞aundɪd ðə ˈp̞ɔːk ˈtʃɒp /

7. ʃ / dʒ (♥) / hɪ ˈpɒlɪs̪t ðɪ ˈedʒ əv ɪz ˈs̪erɪfs ˈbædʒ /
/ hɪ ˈpɒlɪʃt ðɪ ˈedz̞ əv ɪz ˈʃerɪfs ˈbædz̞ /

8. p / v (♥) / ðə ˈvæn hæz ˈp̞ɑːst ðə ˈvɪlɪdʒ ˈp̞ʌb /
/ ðə ˈʋ̞æn hæz ˈpɑːst ðə ˈʋ̞ɪlɪdʒ ˈpʌb /

9. θ / dʒ (♠) / ðeə wəz ˈsʌmt̪ɪŋ ˈlɑːdʒ ɒn ðə ˈsteɪdʒ əv ðə ˈt̪ɪətə /
/ ðeə wəz ˈsʌmθɪŋ ˈlɑːdz̞ ɒn ðə ˈsteɪdz̞ əv ðə ˈθɪətə /

10. ʃ / p (♥) / ðɪ ɪmˈpres̪n̩ ðə ˈpeɪntɪŋ ˈgeɪv əv ðə ˈs̪ip ɒn ðə ˈwɔːtə wɒz ˈpɜːfɪkt /
/ ðɪ ɪmˈpreʃn̩ ðə ˈp̞eɪntɪŋ ˈgeɪv əv ðə ˈʃip ɒn ðə ˈwɔːtə wɒz ˈp̞ɜːfɪkt /

11. v / tʃ (♠) / ðə ˈtʃɜːtʃ ˈʋ̞ɪkə wəz nɒt ˈrɪəlɪ ˈʋ̞erɪ ˈrɪtʃ ət ˈɔːl /[4]
/ ðə ˈtʃɜːt̪s ˈvɪkə wəz ˈnɒt ˈrɪəlɪ ˈverɪ ˈrɪt̪s ət ˈɔːl /

12. θ / p (♠) / aɪ ˈt̪ɪŋk ðeə wə ˈt̪ɜːtɪfɔː ˈpəʊstəz ɪn ðə ˈpʌb /
/ aɪ ˈθɪŋk ðeə wə ˈθɜːtɪfɔː ˈp̞əʊstəz ɪn ðə ˈp̞ʌb /

13. ʃ / — (♥) / ðə ˈvaʊəl ˈsaʊndz əv ˈɪŋglɪs̪ ɑː ˈverɪ ˈspes̪l̩ /
/ ðə ˈvaʊəl ˈsaʊndz əv ˈɪŋglɪʃ ɑː ˈverɪ ˈspeʃl̩ /

14. tʃ / θ (♥) / ðə ˈt̪sæmpjən tʊk ə ˈbreθ | ðen ˈsed ˈsʌmθɪŋ tə ˈt̪seɪndʒ ðə ˈsʌbdʒɪkt /[5]
/ ðə ˈtʃæmpjən tʊk ə ˈbret̪ | ˈðen sed ˈsʌmt̪ɪŋ tə ˈtʃeɪndʒ ðə ˈsʌbdʒɪkt /

15. dʒ / — (♠) / ˈjes | ðə wəz ˈsʌmθɪŋ ɒn ðə ˈbrɪdz̞ | ˈsʌmθɪŋ ˈlɑːdz̞ /
/ ˈjes | ðə wəz ˈsʌmθɪŋ ɒn ðə ˈbrɪdʒ | ˈsʌmθɪŋ ˈlɑːdʒ /

16. v / ʃ (♥) / ðə ˈʃeɪp əv ðə ˈʋ̞est wəz ˈʃɒkɪŋ n̩d ˈʋ̞ʌlgə /
/ ðə ˈs̪eɪp əv ðə ˈvest wəz ˈs̪ɒkɪŋ n̩d ˈvʌlgə /

17. θ / tʃ (♠) / ðə wəz ˈnʌt̪ɪŋ ðə ˈtʃæmpjən kəd ˈduː | ˈnʌt̪ɪŋ tə ˈtʃeɪndʒ /
/ ðə wəz ˈnʌθɪŋ ðə ˈt̪sæmpjən kəd ˈduː | ˈnʌθɪŋ tə ˈt̪seɪndʒ /

---

[4] The rhythm for the two versions is not quite the same here.
[5] The rhythm for the two versions is again not quite the same.

18. dʒ / p (♠) / ðə ˈlɑːdz ˈhelpɪŋ əv ˈpɒrɪdz wəz ˈpaɪpɪŋ ˈhɒt / [6]
/ ðə ˈlɑːrdʒ ˈhelpɪŋ əv ˈpɒrɪdʒ wəz ˈpaɪpɪŋ ˈhɒt /

19. ʃ / v (♥) / ðə ˈvaʊəl ˈsaundz əv ˈɪŋglɪs ɑː ˈverɪ ˈspesl̩ /
/ ðə ˈʊaʊəl ˈsaundz əv ˈɪŋglɪʃ ɑː ˈʊerɪ ˈspeʃl̩ /

20. θ / dʒ (♥) / ˈjes | ðə wəz ˈsʌmtɪŋ ɒn ðə ˈbrɪdʒ | ˈsʌmtɪŋ ˈlɑːdʒ /
/ ˈjes | ðə wəz ˈsʌmθɪŋ ɒn ðə ˈbrɪdz | ˈsʌmθɪŋ ˈlɑːdz /

21. p / tʃ (♠) / ðə ˈpɪŋk ˈwɒtʃ wəz ˈmʌtʃ tuː ɪkˈspensɪv /
/ ðə ˈpɪŋk ˈwɒts wəz ˈmʌts tuː ɪkˈspensɪv /

22. ʃ / θ (♥) / hɪ ˈθɔːt ðə ˈseɪp əv ðə ˈθɪŋ wəz ˈspɒkɪŋ /
/ hɪ ˈtɔːt ðə ˈʃeɪp əv ðə ˈtɪŋ wəz ˈʃɒkɪŋ /

23. tʃ / v (♠) / ðə ˈwɪərɪ ˈvɪkə ˈentəd hɪz ˈtsɜːts /
/ ðə ˈwɪərɪ ˈʊɪkə ˈentəd hɪz ˈtʃɜːtʃ /

24. p / — (♥) / ðɪ ɪmˈpreʃn̩ ðə ˈpeɪntɪŋ ˈgeɪv ɒv ðə ˈʃip ɒn ðə ˈwɔːtə wəz ˈpɜːfɪkt /
/ ðɪ ɪmˈpreʃn̩ ðə ˈpeɪntɪŋ ˈgeɪv ɒv ðə ˈʃip ɒn ðə ˈwɔːtə wəz ˈpɜːfɪkt /

25. dʒ / tʃ (♠) / ˈfɔːtʃnətlɪ ˈnɒt mʌtʃ ˈdæmɪdz wəz ˈdʌn tə ðə ˈhedz /
/ ˈfɔːtsnətlɪ ˈnɒt mʌts ˈdæmɪdʒ wəz ˈdʌn tə ðə ˈhedʒ /

26. — / θ (♠) / ðə wəz ˈnʌθɪŋ ðə ˈtʃæmpjən kəd ˈduː | ˈnʌθɪŋ tə ˈtʃeɪndʒ /
/ ðə wəz ˈnʌtɪŋ ðə ˈtʃæmpjən kəd ˈduː | ˈnʌtɪŋ tə ˈtʃeɪndʒ /

27. ʃ / tʃ (♥) / ðə ˈkəʊtʃ hæd tə ˈriːtʃ ɪts destɪˈneɪsn̩ bɪfɔː ðə ˈsəʊ ˈstɑːtɪd /
/ ðə ˈkəʊts hæd tə ˈriːts ɪts destɪˈneɪʃn̩ bɪfɔː ðə ˈʃəʊ ˈstɑːtɪd /

28. p / θ (♥) / ˈθauzn̩dz əv ˈpiːpl̩ əd ˈθɔːt ɪt ˈpɜːfɪkt /
/ ˈtauzn̩dz əv ˈpiːpl̩ əd ˈtɔːt ɪt ˈpɜːfɪkt/

29. tʃ / dʒ (♠) / ˈðʒɒn eɪt ə ˈlʌnts ðət wəz ˈmʌts tuː ˈlɑːdʒ /
/ ˈdzɒn eɪt ə ˈlʌntʃ ðət wəz ˈmʌtʃ tuː ˈlɑːdz /

30. v / — (♥) / ðə ˈʊæn hæz ˈpɑːst ðə ˈʊɪlɪdʒ ˈpʌb /
/ ðə ˈvæn hæz ˈpɑːst ðə ˈvɪlɪdʒ ˈpʌb /

---

[6] The 'b' version unfortunately has noticeable rhotacisation of the / ɑː / in large.

APPENDIX B :
TEXTS OF THE QUESTIONNAIRE FOR EXPERIMENT 1

For each of the three language groups used as informants (English, Finnish and Finnish-Swedish) two versions of the questionnaire were used: Questionnaire 1 and Questionnaire 2. Questionnaire 2 used the same sentences as Questionnaire 1 but the order of the sentences was reversed: Sentence 30 on Questionnaire 1 became Sentence 1 on Questionnaire 2, Sentence 29 became Sentence 2, etc. Note that the presentation on Tape 2 of the two versions of the sentences was also reversed, and in addition the gender of the speakers for the presentation of a particular sentence on Tape 1 was also changed.

Below are presented just one of the two Questionnaires for each language group.

English version - Questionnaire 1

ALTERNATIVE PRONUNCIATIONS

Instructions

The purpose of this study is to discover what preferences people have concerning foreign speakers' possible pronunciations of English.

On the tape you are about to listen to, you will hear several different speakers, both men and women, saying the sentences listed below.

Each sentence is said twice by the same speaker: version A and version B. Which version do you prefer? Which pronunciation sounds better?

You will hear each sentence four times:
A - B, A - B.

By putting a cross in Box A or Box B, indicate which version you feel is **better** English. Do not worry about the meaning of the sentences. Just listen to the voice and to the pronunciation, and say which version **sounds better.**

Which is better?

1. The van has passed the village pub. A ☐ B ☐

2. John ate a lunch that was much too large. A ☐ B ☐

3. Thousands of people had thought it perfect. A ☐ B ☐

4. The coach had to reach its destination before the show started. A ☐ B ☐

5. There was nothing the champion could do, nothing to change. A ☐ B ☐

6. Fortunately not much damage was done to the hedge. A ☐ B ☐

7. The impression the painting gave of the ship on the water was perfect. A ☐ B ☐

8. The weary vicar entered his church. A ☐ B ☐

9. He thought the shape of the thing was shocking. A ☐ B ☐

10. The pink watch was much too expensive. A ☐ B ☐

11. Yes, there was something on the bridge, something large. A ☐ B ☐

12. The vowel sounds of English are very special. A ☐ B ☐

Which is better?

13. The large helping of porridge was piping hot. A ☐ B ☐

14. There was nothing the champion could do, nothing to change. A ☐ B ☐

15. The shape of the vest was shocking and vulgar. A ☐ B ☐

16. Yes, there was something on the bridge, something large. A ☐ B ☐

17. The champion took a breath, then said something to change the subject. A ☐ B ☐

18. The vowel sounds of English are very special. A ☐ B ☐

19. I think there were 34 posters in the pub. A ☐ B ☐

20. The church vicar was not really very rich at all. A ☐ B ☐

21. The impression the painting gave of the ship on the water was perfect. A ☐ B ☐

22. There was something large on the stage of the theatre. A ☐ B ☐

23. The van has passed the village pub. A ☐ B ☐

24. He polished the edge of his sheriff's badge. A ☐ B ☐

Which is better?

25. The butcher pounded the pork chop. A ☐ B ☐

26. The thirty villains had thumped on the van. A ☐ B ☐

27. Thousands of people had thought it perfect. A ☐ B ☐

28. He enjoyed driving his van through the village. A ☐ B ☐

29. The shop had to pay for all the parcels of shoes. A ☐ B ☐

30. The coach had to reach its destination before the show started. A ☐ B ☐

That is the end of the presentation of the pairs of sentences. Please stop the tape and go on to fill in the questionnaire concerning your personal background. Thank you.

Finnish version - Questionnaire 1

VAIHTOEHTOISET ÄÄNTÄMYKSET

Ohjeet

Tämän kyselyn avulla pyritään saamaan selville, mitä suomalaiset ajattelevat englannin kielen ääntämisestä.

Kuulet nauhoituksen, jossa muutamat naiset ja miehet sanovat seuraavilla sivuilla olevat lauseet kahdella eri tavalla. Pidätkö enemmän versiosta A vai versiosta B? Kumpi versio äännetään mielestäsi paremmin? Kumpi kuulostaa paremmalta?

Kuulet jokaisen lauseen neljä kertaa, järjestyksessä A - B, A - B.

Ole hyvä ja merkitse rastilla versio, joka vastaa sinusta parempaa ääntämistä. **Älä kiinnitä huomiota lauseen merkitykseen, kuuntele vain puhujan ääntä ja ääntämistä, ja valitse mielestasi parempi versio.**

	Kumpi on parempi?
1. The van has passed the village pub.	1 ☐ 2 ☐
2. John ate a lunch that was much too large.	1 ☐ 2 ☐
3. Thousands of people had thought it perfect.	1 ☐ 2 ☐
4. The coach had to reach its destination before the show started.	1 ☐ 2 ☐
5. There was nothing the champion could do, nothing to change.	1 ☐ 2 ☐
6. Fortunately not much damage was done to the hedge.	1 ☐ 2 ☐
7. The impression the painting gave of the ship on the water was perfect.	1 ☐ 2 ☐
8. The weary vicar entered his church.	1 ☐ 2 ☐
9. He thought the shape of the thing was shocking.	1 ☐ 2 ☐
10. The pink watch was much too expensive.	1 ☐ 2 ☐
11. Yes, there was something on the bridge, something large.	1 ☐ 2 ☐
12. The vowel sounds of English are very special.	1 ☐ 2 ☐

Kumpi on parempi?

13. The large helping of porridge was piping hot. 1 ☐ 2 ☐

14. There was nothing the champion could do, nothing to change. 1 ☐ 2 ☐

15. The shape of the vest was shocking and vulgar. 1 ☐ 2 ☐

16. Yes, there was something on the bridge, something large. 1 ☐ 2 ☐

17. The champion took a breath, then said something to change the subject. 1 ☐ 2 ☐

18. The vowel sounds of English are very special. 1 ☐ 2 ☐

19. I think there were 34 posters in the pub. 1 ☐ 2 ☐

20. The church vicar was not really very rich at all. 1 ☐ 2 ☐

21. The impression the painting gave of the ship on the water was perfect. 1 ☐ 2 ☐

22. There was something large on the stage of the theatre. 1 ☐ 2 ☐

23. The van has passed the village pub. 1 ☐ 2 ☐

24. He polished the edge of his sheriff's badge. 1 ☐ 2 ☐

Kumpi on parempi?

25. The butcher pounded the pork chop. 1 ☐ 2 ☐

26. The thirty villains had thumped on the van. 1 ☐ 2 ☐

27. Thousands of people had thought it perfect. 1 ☐ 2 ☐

28. He enjoyed driving his van through the village. 1 ☐ 2 ☐

29. The shop had to pay for all the parcels of shoes. 1 ☐ 2 ☐

30. The coach had to reach its destination before the show started. 1 ☐ 2 ☐

Äskeinen oli viimeinen lausepari.

Ole hyvä ja pysäytä nauhuri ja täytä seuraava kyselylomake, joka käsittelee sinun taustatietojasi.
tausttaustatietojasi.
Kiitos!

Swedish version - Questionnaire 2

ALTERNATIVA UTTAL

Instruktioner

Avsikten med denna studie är att avslöja hur finländare reagerar på engelskt uttal med olika brytningar.

Du ska snart höra ett antal talare, både manliga och kvinnliga, uttala meningarna som finns på följande sidorna. Varje mening uttalas på två olika sätt. Vilken version gillar du mera, A eller B? Vilket uttal låter bättre, A eller B?

Du kommer att höra båda versionerna två gånger. Ordningen är:
A - B,
A - B.

Varsågod, sätt ett kryss i endera lådan, A eller B, beroende på vilken version du tycker är bättre engelska. Du behöver inte bry dig om meningarnas betydelse, koncentrera dig bara på rösten och uttalet, och bestäm vilken version låter bättre, A eller B.

	Vilketdera är bättre?
1. The coach had to reach its destination before the show started.	A ☐ B ☐
2. The shop had to pay for all the parcels of shoes.	A ☐ B ☐
3. He enjoyed driving his van through the village.	A ☐ B ☐
4. Thousands of people had thought it perfect.	A ☐ B ☐
5. The thirty villains had thumped on the van.	A ☐ B ☐
6. The butcher pounded the pork chop.	A ☐ B ☐
7. He polished the edge of his sheriff's badge.	A ☐ B ☐
8. The van has passed the village pub.	A ☐ B ☐
9. There was something large on the stage of the theatre.	A ☐ B ☐
10. The impression the painting gave of the ship on the water was perfect.	A ☐ B ☐
11. The church vicar was not really very rich at all.	A ☐ B ☐
12. I think there were 34 posters in the pub.	A ☐ B ☐

	Vilketdera är bättre?
13. The vowel sounds of English are very special.	A ☐ B ☐
14. The champion took a breath, then said something to change the subject.	A ☐ B ☐
15. Yes, there was something on the bridge, something large.	A ☐ B ☐
16. The shape of the vest was shocking and vulgar.	A ☐ B ☐
17. There was nothing the champion could do, nothing to change.	A ☐ B ☐
18. The large helping of porridge was piping hot.	A ☐ B ☐
19. The vowel sounds of English are very special.	A ☐ B ☐
20. Yes, there was something on the bridge, something large.	A ☐ B ☐
21. The pink watch was much too expensive.	A ☐ B ☐
22. He thought the shape of the thing was shocking.	A ☐ B ☐
23. The weary vicar entered his church.	A ☐ B ☐
24. The impression the painting gave of the ship on the water was perfect.	A ☐ B ☐

Vilketdera är bättre?

25. Fortunately not much damage was done to the hedge. A ☐ B ☐

26. There was nothing the champion could do, nothing to change. A ☐ B ☐

27. The coach had to reach its destination before the show started. A ☐ B ☐

28. Thousands of people had thought it perfect. A ☐ B ☐

29. John ate a lunch that was much too large. A ☐ B ☐

30. The van has passed the village pub. A ☐ B ☐

Det här var den sista frasen.
Var god och stanna bandet och fortsätt att fylla i informationen om din personliga bakgrund.

Tack!

APPENDIX C : PHONETIC TRANSCRIPTIONS OF THE RECORDINGS FOR EXPERIMENT 2

Below is a transcription of the various recorded versions of the text *Jeff and the Picture*. The transcription is basically phonemic (using Cruttenden's set of symbols), though for the sake of economy the misarticulations are presented as underlined symbols which, strictly speaking, are neither phonemic nor phonetic. These underlined symbols should be interpreted in the following way:

/ υ / = voiced labiodental frictionless continuant.
/ s̲ / = voiceless postalveolar sibilant.
/ p̲ / = voiceless unaspirated bilabial plosive.
/ t̲ / = voiceless unaspirated dental plosive.
/ t̲s̲ / = voiceless postalveolar affricate (with no lip-rounding).
/ d̲z̲ / = voiced postalveolar affricate (with no lip-rounding).

The occasional presence of jerky sentence rhythm and slightly non-RP vowel qualities is *not* transcribed. Alternative pronunciations are presented in brackets, one speaker using one version and one speaker the other.

Version 1 'Excellent' : 0 mistakes (Tape 1: Speakers 1 & 5; Tape 2: Speaker 2)

/ ˈdʒef pʊt əsaɪd ðə ˈpɪktʃə hɪd (hɪ həd) bɪn ˈstrʌglɪŋ wɪð ɪn ˈveɪn |
n̩ ˈset ɒf ɒn ə ˈpeɪntɪŋ trɪp tə ðə ˈleɪk |
tə duː ə ˈpɪktʃər əv ˈsuː həˈself ||
hɪ ˈwɒntɪd tə ˈpeɪnt hər əz (æz) ɪ ˈsɔː hər ɪn ɪz ɪmædʒɪˈneɪʃn̩ |
ə ˈleɪdɪ frəm ðə ˈpɑːst ||
ˈðʌs ʃɪ wəz ˈstændɪŋ bɪsaɪd ðə ˈwɔːtə | ˈveɪld | ˈθɔːtfʊl |
ˈgeɪzɪŋ təwɔːdz ðə ˈvɪlɪdʒ n̩ ðə ˈtʃɜːtʃ ||
ʃɪ lʊkt ˈstreɪndʒlɪ əʊld ˈfæʃn̩d ||
ˈθriː deɪz ˈleɪtə hɪ ˈfɪnɪʃt ɪt | n̩ ˈgeɪv ɪt tə hə |
ˈseɪɪŋ ɪn ə ˈkwaɪət ˈvɔɪs ||
ɪts ˈnʌθɪŋ verɪ ˈmʌtʃ ||
ʃɪ ˈstʌdɪd ðə ˈpɪktʃə | ˈtɜːnd tə ˈtʌtʃ (h)ɪz ˈhænd n̩ ˈsed | ˈθæŋk juː ||
aɪm ˈnəʊ ˈdʒʌdʒ | bət ˈaɪ θɪŋk (aɪ ˈθɪŋk) ðɪs ɪz ˈverɪ ˈbjuːtɪfʊl /

Version 2 'Good' : 11 mistakes (Tape 1: Speaker 3; Tape 2: Speaker 4)
(/ v / x 2; / θ / x 1; / ð / x 1; / ʃ / x 2; / tʃ / x 4; / dʒ / x 1)

/ ˈdʒef pʊt əsaɪd ðə ˈpɪktʃə hɪd (hɪ həd) bɪn ˈstrʌglɪŋ wɪð ɪn ˈʊeɪn |
n̩ ˈset ɒf (set ˈɒf) ɒn ə ˈpeɪntɪŋ trɪp tə ðə ˈleɪk |
tə duː ə ˈpɪktsər əv ˈsuː həˈself ||
hɪ ˈwɒntɪd tə ˈpeɪnt hər əz ɪ ˈsɔː hər ɪn ɪz ɪmædʒɪˈneɪsn̩ |
ə ˈleɪdɪ frəm ðə ˈpɑːst ||
ˈðʌs ʃɪ wəz ˈstændɪŋ bɪsaɪd də ˈwɔːtə | ˈveɪld | ˈθɔːtfʊl |
ˈgeɪzɪŋ təwɔːdz ðə ˈʊɪlɪdʒ n̩ ðə ˈtʃɜːts ||
ʃɪ lʊkt ˈstreɪndʒlɪ əʊld ˈfæsn̩d ||
ˈθriː deɪz ˈleɪtə hɪ ˈfɪnɪʃt ɪt | n̩ ˈgeɪv ɪt tə hə |
ˈseɪɪŋ ɪn ə ˈkwaɪət ˈvɔɪs ||
ɪts ˈnʌθɪŋ verɪ ˈmʌts ||
ʃɪ ˈstʌdɪd ðə ˈpɪktʃə | ˈtɜːnd tə tʌts (h)ɪz ˈhænd n̩ ˈsed | ˈθæŋk juː ||
aɪm ˈnəʊ ˈdʒʌdz | bət ˈaɪ tɪŋk ðɪs ɪz ˈverɪ ˈbjuːtɪfʊl /

Version 3(a) 'Fair' : 31 mistakes (Tape 1: Speaker 2; Tape 2: Speaker 1)
(/ v / x 3; / θ / x 4; / ð / x 5; / ʃ / x 6; / p / x 2/3; / tʃ / x 7/6; / dʒ / x 4)

/ ˈdʒef pʊt əsaɪd də ˈpɪktsə hɪd bɪn ˈstrʌglɪŋ wɪd ɪn ˈveɪn |
n̩ ˈset ɒf ɒn ə ˈpeɪntɪŋ trɪp tə ðə ˈleɪk |
tə duː ə ˈpɪktsər əv ˈsuː həˈself ||
hɪ ˈwɒntɪd tə ˈpeɪnt (ˈpeɪnt) hər əz ɪ ˈsɔː hər ɪn ɪz ɪmædzɪˈneɪsn̩ |
ə ˈleɪdɪ frəm ðə ˈpɑːst ||
ˈðʌs sɪ wəz ˈstændɪŋ bɪsaɪd də ˈwɔːtə | ˈʊeɪld | ˈθɔːtfʊl |
ˈgeɪzɪŋ təwɔːdz də ˈʊɪlɪdz n̩ də ˈtsɜːts (ˈtʃɜːts) ||
sɪ lʊkt ˈstreɪndzlɪ əʊld ˈfæsn̩d ||
ˈtriː deɪz ˈleɪtə hɪ ˈfɪnɪst ɪt | n̩ ˈgeɪv ɪt tə hə |
ˈseɪɪŋ ɪn ə ˈkwaɪət ˈʊɔɪs ||
ɪts ˈnʌtɪŋ verɪ ˈmʌts ||
sɪ ˈstʌdɪd ðə ˈpɪktsə | ˈtɜːnd tə tʌts ɪz ˈhænd n̩ ˈsed | ˈtæŋk juː ||
aɪm ˈnəʊ ˈdʒʌdz | bət ˈaɪ tɪŋk ðɪs ɪz ˈverɪ ˈbjuːtɪfʊl /

Version 3(b) 'Fair' : 33 mistakes (Tape 2: Speaker 5)
(/ v / x 4; / θ / x 4; / ð / x 5; / ʃ / x 6; / p / x 4; / tʃ / x 6; / dʒ / x 4)

/ ˈdʒef pʊt əsaɪd də ˈpɪktsə hɪd bɪn ˈstrʌglɪŋ wɪd ɪn ˈveɪn |
n̩ ˈset ɒf ɒn ə ˈpeɪntɪŋ trɪp tə ðə ˈleɪk |
tə duː ə ˈpɪktsər əv ˈsuː həˈself ||
hɪ ˈwɒntɪd tə ˈpeɪnt hər əz ɪ ˈsɔː hər ɪn ɪz ɪmædzɪˈneɪsn̩ |
ə ˈleɪdɪ frəm ðə ˈpɑːst ||
ˈðʌs sɪ wəz ˈstændɪŋ bɪsaɪd də ˈwɔːtə | ˈʊeɪld | ˈθɔːtfʊl |
ˈgeɪzɪŋ təwɔːdz ðə ˈʊɪlɪdz n̩ də ˈtʃɜːts ||
sɪ lʊkt ˈstreɪndzlɪ əʊld ˈfæsn̩d ||
ˈtriː deɪz ˈleɪtə hɪ ˈfɪnɪst ɪt | n̩ ˈgeɪv ɪt tə hə |
ˈseɪɪŋ ɪn ə ˈkwaɪət ˈʊɔɪs ||
ɪts ˈnʌtɪŋ verɪ ˈmʌts ||
sɪ ˈstʌdɪd də ˈpɪktsə | ˈtɜːnd tə tʌts ɪz ˈhænd n̩ ˈsed | ˈtæŋk juː ||
aɪm ˈnəʊ ˈdʒʌdz | bət ˈaɪ tɪŋk ðɪs ɪz ˈʊerɪ ˈbjuːtɪfʊl /

Version 4 'Poor' : 47 mistakes (Tape 1: Speaker 4; Tape 2: Speaker 3)
(/ v / x 6; / θ / x 5; / ð / x 10; / ʃ / x 6; / p / x 7; / tʃ / x 7; / dʒ / x 6)

/ ˈdzef pʊt əsaɪd də ˈpɪktsə hɪd (hɪ həd) bɪn ˈstrʌglɪŋ wɪd ɪn ˈʊeɪn |
n̩ ˈset ɒf ɒn ə ˈpeɪntɪŋ trɪp tə də ˈleɪk |
tə duː ə ˈpɪktsər əv ˈsuː həˈself ||
hɪ ˈwɒntɪd tə ˈpeɪnt hər əz ɪ ˈsɔː hər ɪn ɪz ɪmædzɪˈneɪsn̩ |
ə ˈleɪdɪ frəm də ˈpɑːst ||
ˈdʌs sɪ wəz ˈstændɪŋ bɪsaɪd də ˈwɔːtə | ˈʊeɪld | ˈtɔːtfʊl |
ˈgeɪzɪŋ təwɔːdz də ˈʊɪlɪdz n̩ də ˈtsɜːts ||
sɪ lʊkt ˈstreɪndzlɪ əʊld ˈfæsn̩d ||
ˈtriː deɪz ˈleɪtə hɪ ˈfɪnɪst ɪt | n̩ ˈgeɪv ɪt tə hə |
ˈseɪɪŋ ɪn ə ˈkwaɪət ˈʊɔɪs ||
ɪts ˈnʌtɪŋ ʊerɪ ˈmʌts ||
sɪ ˈstʌdɪd də ˈpɪktsə | ˈtɜːnd tə tʌts (h)ɪz ˈhænd n̩ ˈsed | ˈtæŋk juː ||
aɪm ˈnəʊ ˈdzʌdz | bət ˈaɪ tɪŋk dɪs ɪz ˈʊerɪ ˈbjuːtɪfʊl /

APPENDIX D : TEXTS OF THE QUESTIONNAIRE FOR EXPERIMENT 2

Two versions of the questionnaire for Experiment 2 were used for both the English-speaking listeners and the Finnish-speaking listeners: Questionnaire 1 and Questionnaire 2. Questionnaire 2 was the same as Questionnaire 1 except that the presentation on the page of all the questions except the first one was reversed. Thus the second question concerning age had the excellent–awful poles reversed so that it appeared as awful–excellent; and the fourteen characteristics were reversed both 'horizontally' and 'vertically': i.e. the left pole became the right pole, and the last of the fourteen became the first, the thirteenth became the second, etc.

Since the instructions are the same (in English and Finnish) for both versions of the Questionnaire, and since the forms for Speakers 2 to 5 are the same as the form for Speaker 1, only illustrative pages are reproduced here, not the whole Questionnaires.

English version : Instructions for both Questionnaire 1 and Questionnaire 2

Questionnaire

Background

The purpose of this study is to discover whether listeners are able to infer the personality and other characteristics of a speaker from merely listening to their voice. For example, does the voice of a speaker reveal or suggest to you how educated, or proud, or ambitious, etc. that person is?

Instructions

On the tape you are about to listen to, you will hear five different speakers reading the same text entitled "Jeff and the Picture". The recording of each speaker is repeated so that you will hear the same speakers twice:
"Speaker 1: First Reading: (... ... ...). Second Reading: (... ... ...)", etc.

Listen to each voice, and as you listen,**without paying any attention to the content of what they are saying**, answer the questions which follow.

The first question concerns age and is straightforward.

The second question involves a 7-point scale with the two ends marked with opposite extreme values: "excellent pronunciation" vs. "awful pronunciation". Indicate where you think the speaker is situated along this scale.

The third question involves fourteen of the same sort of 7-point scales with the ends again marked with opposite (bi-polar) adjectives. Please indicate where you would place the speaker on each of these scales.

You might like to consider the boxes of the 7-point scales as indicating a range going from extremely X, through neutral, to extremely Y. For example, the adjective pair "long" - "short" would have the following sort of range:

- extremely long
- long
- quite long
- in the middle (neither "long" nor "short")
- quite short
- short
- extremely short.

If you are unsure, make a guess!

You have a separate answer page to fill in for each of the five speakers.

English version : Questionnaire 1, Speaker 1

Speaker 1

Age

How old do you think the speaker is?

☐ 14-15 ☐ 16-17 ☐ 18-19 ☐ 20-21 ☐ 22-23 ☐ 24-25 ☐ 26-27

☐ 28-29 ☐ 30-31 ☐ 32-33 ☐ 34-35 ☐ 36-37 ☐ 38-39 ☐ 40-41

Accent

How good do you think this foreigner's pronunciation of English is?

excellent pronunciation	☐	☐	☐	☐	☐	☐	☐	awful pronunciation

Characteristics

Basing your judgement merely on the voice of the speaker, where would you place the speaker on the following 7-point scales?

educated	☐	☐	☐	☐	☐	☐	☐	uneducated
stupid	☐	☐	☐	☐	☐	☐	☐	intelligent
cruel	☐	☐	☐	☐	☐	☐	☐	kind
friendly	☐	☐	☐	☐	☐	☐	☐	unfriendly
unsympathetic	☐	☐	☐	☐	☐	☐	☐	sympathetic
untrustworthy	☐	☐	☐	☐	☐	☐	☐	trustworthy
reliable	☐	☐	☐	☐	☐	☐	☐	unreliable
hardworking	☐	☐	☐	☐	☐	☐	☐	lazy
poor	☐	☐	☐	☐	☐	☐	☐	rich
honest	☐	☐	☐	☐	☐	☐	☐	dishonest
unsuccessful	☐	☐	☐	☐	☐	☐	☐	successful
ambitious	☐	☐	☐	☐	☐	☐	☐	unambitious
serious	☐	☐	☐	☐	☐	☐	☐	fun-loving
hesitant / unsure	☐	☐	☐	☐	☐	☐	☐	self-confident

English version : Questionnaire 2, Speaker 1

Speaker 1

Age

How old do you think the speaker is?

☐ 14-15 ☐ 16-17 ☐ 18-19 ☐ 20-21 ☐ 22-23 ☐ 24-25 ☐ 26-27

☐ 28-29 ☐ 30-31 ☐ 32-33 ☐ 34-35 ☐ 36-37 ☐ 38-39 ☐ 40-41

Accent

How good do you think this foreigner's pronunciation of English is?

awful pronunciation	☐	☐	☐	☐	☐	☐	☐	excellent pronunciation

Characteristics

Basing your judgement merely on the voice of the speaker, where would you place the speaker on the following 7-point scales?

self-confident	☐	☐	☐	☐	☐	☐	☐	hesitant / unsure
fun-loving	☐	☐	☐	☐	☐	☐	☐	serious
unambitious	☐	☐	☐	☐	☐	☐	☐	ambitious
successful	☐	☐	☐	☐	☐	☐	☐	unsuccessful
dishonest	☐	☐	☐	☐	☐	☐	☐	honest
rich	☐	☐	☐	☐	☐	☐	☐	poor
lazy	☐	☐	☐	☐	☐	☐	☐	hard-working
unreliable	☐	☐	☐	☐	☐	☐	☐	reliable
trustworthy	☐	☐	☐	☐	☐	☐	☐	untrustworthy
sympathetic	☐	☐	☐	☐	☐	☐	☐	unsympathetic
unfriendly	☐	☐	☐	☐	☐	☐	☐	friendly
kind	☐	☐	☐	☐	☐	☐	☐	cruel
intelligent	☐	☐	☐	☐	☐	☐	☐	stupid
uneducated	☐	☐	☐	☐	☐	☐	☐	educated

Finnish version : Instructions for both Questionnaire 1 and Questionnaire 2

Kyselylomake

Taustatietoa

Tämän tutkimuksen tarkoituksena on selvittää, pystyvätkö kuulijat päättelemään tietyn puhujan persoonallisuuden ja muita luonteenpiirteitä pelkästään kuuntelemalla tämän ääntä. Kertooko puhujan ääni jotakin esimerkiksi siitä, miten koulutettu, ylpeä tai kunnianhimoinen hän on?

Ohjeet

Nauhalla, jonka tulet kohta kuuntelemaan, on viisi eri puhujaa, jotka kaikki lukevat saman tekstin - "Jeff and the Picture". Kukin puhuja lukee tekstin kahteen kertaan.

Kuuntele jokaisen puhujan ääntä ja kuunnellessasi vastaa erillisillä papereilla oleviin kysymyksiin **kiinittämättä minkäänlaista huomiota heidän puheensä sisältöön.**

Ensimmäinen kysymys koskee yksinkertaisesti puhujan ikää.

Toisessa kysymyksessa on seitsenportainen asteikko, jonka päät on nimetty vastakkaisilla ääriarvoilla: "erinomainen ääntäminen" ja "surkea ääntäminen". Merkitse asteikkoon se kohta, johon puhuja mielestäsi sijoittuu.

Kolmannen kysymyksen 14 kohtaa sisältävät samantyyppisiä seitsenportaisia asteikkoja, joiden ääripäät on myöskin nimetty vastakkaisilla adjektiiveilla. Merkitse kohta, johon sijoittaisit puhujan kullakin asteikolla.

Voit ajatella asteikkojen ilmaisevan tiettyä ominaisuutta siten, että vasemmalla on "äärimmäisen X", keskella neutraali ja oikealla "äärimmäisen Y". Esimerkiksi adjektiiviparin "pitka" - "lyhyt" voit ajatella jakautuvan seuraavasti:

- tosi pitkä
- pitkä
- melko pitkä
- keskiverto
- melko lyhyt
- lyhyt
- tosi lyhyt

Kullekin puhujalle on oma vastauskaavake.

Finnish version : Questionnaire 1, Speaker 1

PUHUJA 1

Ikä

Minkä ikäinen arvelet puhujan olevan?

☐ 14-15 ☐ 16-17 ☐ 18-19 ☐ 20-21 ☐ 22-23 ☐ 24-25 ☐ 26-27

☐ 28-29 ☐ 30-31 ☐ 32-33 ☐ 34-35 ☐ 36-37 ☐ 38-39 ☐ 40-41

Korostus

Miten hyvänä pidät tämän suomalaisen englannin kielen ääntämistä?

erinomainen	☐	☐	☐	☐	☐	☐	☐	surkea

Luonteenpiirteet

Mihin arvelisit puhujan sijoittuvan seuraavilla asteikoilla pelkän äänensä perusteella?

sivistynyt / koulutettu	☐	☐	☐	☐	☐	☐	☐	sivistymätön / kouluttamaton
tyhmä	☐	☐	☐	☐	☐	☐	☐	älykäs
julma	☐	☐	☐	☐	☐	☐	☐	lempeä
ystävällinen	☐	☐	☐	☐	☐	☐	☐	epäystävällinen
epämiellyttävä	☐	☐	☐	☐	☐	☐	☐	miellyttävä
luottamusta herättävä	☐	☐	☐	☐	☐	☐	☐	ei luottamusta herättävä
luotettava	☐	☐	☐	☐	☐	☐	☐	epäluotettava
rehellinen	☐	☐	☐	☐	☐	☐	☐	epärehellinen
ahkera	☐	☐	☐	☐	☐	☐	☐	laiska
epäonnistuva	☐	☐	☐	☐	☐	☐	☐	menestyvä
kunnianhimoinen	☐	☐	☐	☐	☐	☐	☐	ei kunnianhimoinen
itsevarma	☐	☐	☐	☐	☐	☐	☐	epävarma
vakava	☐	☐	☐	☐	☐	☐	☐	leikkimielinen
köyhä	☐	☐	☐	☐	☐	☐	☐	rikas

Finnish version : Questionnaire 2, Speaker 1

PUHUJA 1

Ikä

Minkä ikäinen arvelet puhujan olevan?

☐ 14-15 ☐ 16-17 ☐ 18-19 ☐ 20-21 ☐ 22-23 ☐ 24-25 ☐ 26-27

☐ 28-29 ☐ 30-31 ☐ 32-33 ☐ 34-35 ☐ 36-37 ☐ 38-39 ☐ 40-41

Korostus

Miten hyvänä pidät tämän suomalaisen englannin kielen ääntämistä?

surkea	☐	☐	☐	☐	☐	☐	☐	erinomainen

Luonteenpiirteet

Mihin arvelisit puhujan sijoittuvan seuraavilla asteikoilla pelkän äänensä perusteella?

rikas	☐	☐	☐	☐	☐	☐	☐	köyhä
leikkimielinen	☐	☐	☐	☐	☐	☐	☐	vakava
epävarma	☐	☐	☐	☐	☐	☐	☐	itsevarma
ei kunnianhimoinen	☐	☐	☐	☐	☐	☐	☐	kunnianhimoinen
menestyvä	☐	☐	☐	☐	☐	☐	☐	epäonnistuva
laiska	☐	☐	☐	☐	☐	☐	☐	ahkera
epärehellinen	☐	☐	☐	☐	☐	☐	☐	rehellinen
epäluotettava	☐	☐	☐	☐	☐	☐	☐	luotettava
ei luottamusta herättävä	☐	☐	☐	☐	☐	☐	☐	luottamusta herättävä
miellyttävä	☐	☐	☐	☐	☐	☐	☐	epämiellyttävä
epäystävällinen	☐	☐	☐	☐	☐	☐	☐	ystävällinen
lempeä	☐	☐	☐	☐	☐	☐	☐	julma
älykäs	☐	☐	☐	☐	☐	☐	☐	tyhmä
sivistymätön / kouluttamaton	☐	☐	☐	☐	☐	☐	☐	sivistynyt / koulutettu

APPENDIX E : BIOGRAPHICAL BACKGROUND QUESTIONNAIRE

All informants for both Experiment 1 and Experiment 2 filled in a background questionnaire about their personal biographical details. The only information from this wealth of detail that was eventually used in the present study was: gender, age and mother-tongue.

English version

Background Information

1. Are you . . . ☐ Male ☐ Female

2. What is your nationality? ☐ British ☐ Other
Please name: _____

3. What is your mother tongue? ☐ English ☐ Welsh
☐ Other Please name: _____

4. In what year were you born? ____

5. Where did you go to school?

5a. In which town(s)? _____

5b. In which county(ies) / district(s)? _____

6. Have you completed any of the following examinations?
Please cross 'yes' or 'no' for each exam.

6a. CSE ☐ Yes ☐ No

6b. GCSE ☐ Yes ☐ No

6c. 'O' Levels / Scottish Lowers ☐ Yes ☐ No

6d. 'A' Levels / Scottish Highers ☐ Yes ☐ No

6e. College qualifications (HND, SRN, etc.) ☐ Yes ☐ No

6f. First university degree (BA, BSc, etc.) ☐ Yes ☐ No

6g. Postgraduate university degree (MA, PhD, etc.) ☐ Yes ☐ No

7. Which if any foreign languages did you study at school, and for how many years did you study them?

7a. Languages at school		7b. Years
☐ None		
☐ French	.....	☐
☐ German	.....	☐
☐ Spanish	.....	☐
☐ Other	.....	☐

8. Which if any foreign languages did you (do you) study at college and/or university, and for how many years?

8a. Languages at College/University		8b. Years
☐ None		
☐ French	.....	☐
☐ German	.....	☐
☐ Spanish	.....	☐
☐ Other	.....	☐

9. How good do you think you are now in the languages you just crossed in Questions 7a and 8a compared to your friends or other people?

9a. French	9b. German	9c. Spanish
☐ Excellent	☐ Excellent	☐ Excellent
☐ Good	☐ Good	☐ Good
☐ Better than average	☐ Better than average	☐ Better than average
☐ Average	☐ Average	☐ Average
☐ Below average	☐ Below average	☐ Below average
☐ Weak	☐ Weak	☐ Weak
☐ Extremely weak	☐ Extremely weak	☐ Extremely weak

9d. Other - Please name : _____

☐ Excellent
☐ Good
☐ Better than average
☐ Average
☐ Below average
☐ Weak
☐ Extremely weak

10. Do you ever use whatever foreign language skills you have for pleasure and/or work?

10a. Pleasure		10b. Work
☐	Never	☐
☐	Rarely (e.g. once or twice a year)	☐
☐	Occasionally (e.g. about once a month)	☐
☐	Quite often (e.g. about once a week)	☐
☐	Regularly (e.g. every day, or almost every day)	☐

11. Have you ever travelled abroad on holiday in a non-English speaking country?

☐ No ☐ Yes

12. Have you ever worked (or lived) abroad?

☐ No ☐ Yes

13. Where? In what countries? _____

14. For how long did you work (live) abroad? ____

15. And when did you work (live) abroad? In what year? ____

WORK

EITHER 16. What do you do for a living at the moment?

_____

OR 17. If you are a student, . . .

17a. in which college/university do you study? _____

17b. in which department do you study? _____

17c. when did you start your studies? ____

18. How do you think **other people** would describe your income level at the moment?

- ☐ extremely high
- ☐ high
- ☐ average
- ☐ low
- ☐ no income or unemployed

FAMILY BACKGROUND

19. What did your **father** do for a living while you were at school?

_____

20. What did your **mother** do for a living while you were at school?

_____

21. Which of the following labels would you say best describes your parents' income while you were going to school?

21a. Father's income

- ☐ extremely high
- ☐ high
- ☐ average
- ☐ low
- ☐ no income or unemployed

21b. Mother's income

- ☐ extremely high
- ☐ high
- ☐ average
- ☐ low
- ☐ no income or unemployed

22. What education did your parents have?

22a. Father's Education

- ☐ Basic school education
- ☐ Job and/or skilled work training
- ☐ College level education
- ☐ University/Academic education

22b. Mother's Education

- ☐ Basic school education
- ☐ Job and/or skilled work training
- ☐ College level education
- ☐ University/Academic education

Finnish version

Kyselylomake

1. Sukupuoli ☐ Nainen ☐ Mies

2. Kansalaisuus ☐ Suomi ☐ Ruotsi ☐ Muu Mikä? _____

3. Äidinkieli ☐ Suomi ☐ Ruotsi ☐ Saame ☐ Muu

Mikä? _____

4. Syntymävuosi ____

5. Koulun käynti:

5a. Millä paikkakunnilla kävit koulua?

_____

5b. Missä läänissä (lääneissä)? _____

6. Kuinka monta vuotta opiskelit englantia koulussa? [.]

7. Kuinka monta vuotta olet opiskellut englantia korkeakoulussa? [.]

8. Kuinka hyvä englannin kielen taitosi on mielestäsi saman verran englantia opiskelleisiin kavereihin ja muihin henkilöihin verrattuna? Onko se . . .

erinomainen ☐
hyvä ☐
keskitasolla ☐
heikohko ☐
erittäin heikko ☐

9. Oletko koskaan käynyt maassa, missä englantia puhutaan äidinkielenä (esim. Iso-Britanniassa, Yhdysvalloissa) ?

☐ En koskaan ☐ Kyllä

10. Jos vastaus on kyllä:

10a. - Kuinka usein olet käynyt maassa, missä englantia puhutaan äidinkielenä? ☐ Kerran ☐ Kahdesti ☐ Enemmän

10b. - Kuinka monta viikkoa olet asunnut maassa, missä englantia puhutaan äidinkielenä?

☐ 1-4 viikkoa ☐ 5-8 viikkoa ☐ 9 viikkoa tai enemmän

11. Mitä kieliä (englannin kieltä lukuun ottamatta) olet opiskellut koulussa ja/tai korkeakoulussa?

	11a. Koulussa	**11b. Korkeakoulussa**
Ruotsi	☐	☐
Saksa	☐	☐
Venäjä	☐	☐
Ranska	☐	☐
Muu	☐	☐

12. Kuinka hyvä kielitaitosi edellä mainituissa kielissä on muihin ihmisiin verrattuna?

12a. Ruotsi	12b. Saksa	12c. Venäjä
☐ erinomainen	☐ erinomainen	☐ erinomainen
☐ hyvä	☐ hyvä	☐ hyvä
☐ keskitasoa parempi	☐ keskitasoa parempi	☐ keskitasoa parempi
☐ keskitasolla	☐ keskitasolla	☐ keskitasolla
☐ keskitason alapuolella	☐ keskitason alapuolella	☐ keskitason alapuolella
☐ heikohko	☐ heikohko	☐ heikohko
☐ erittäin heikko	☐ erittäin heikko	☐ erittäin heikko

12d. Ranska	12e. Muu	Mikä? _____
☐ erinomainen	☐ erinomainen	
☐ hyvä	☐ hyvä	
☐ keskitasoa parempi	☐ keskitasoa parempi	
☐ keskitasolla	☐ keskitasolla	
☐ keskitason alapuolella	☐ keskitason alapuolella	
☐ heikohko	☐ heikohko	
☐ erittäin heikko	☐ erittäin heikko	

13. Luetko englannin- vai suomenkielisiä sanomalehtiä, aikakauslehtiä, ammattilehtiä, kirjoja?

- ☐ Aina suomenkielisiä.
- ☐ Suomenkielisiä useammin kuin englanninkielisiä.
- ☐ Yhtä paljon molempia.
- ☐ Englanninkielisiä useammin kuin suomenkielisiä.
- ☐ Aina englanninkielisiä.

14. Kuunteletko englannin- vai suomenkielisiä ääni- ja/tai CD-levyjä?

- ☐ Aina suomenkielisiä.
- ☐ Suomenkielisiä useammin kuin englanninkielisiä.
- ☐ Yhtä paljon molempia.
- ☐ Englanninkielisiä useammin kuin suomenkielisiä.
- ☐ Aina englanninkielisiä.

15. Katsotko englanninkielisiä ohjelmia televisiosta?

☐ En ☐ Kyllä

16. Jos ohjelma on tekstitetty,

- luetko aina koko tekstin ☐
- luetko suurimman osan tekstistä ☐
- luetko tavallisesti noin puolet tekstistä ☐
- luetko pätkän sieltä, toisen täältä ☐
- luetko tekstin lukematta lähes kokonaan ☐

17. Käytkö elokuvateatterissa katsomassa englanninkielisiä elokuvia?

☐ En ☐ Kyllä

18. Jos elokuva on tekstitetty,

- luetko aina koko tekstin ☐
- luetko suurimman osan tekstistä ☐
- luetko tavallisesti noin puolet tekstistä ☐
- luetko vain pätkän sieltä, toisen täältä ☐
- luetko tekstin lukematta lähes kokonaan ☐

19. Kirjoitatko ystävillesi tai kirjeenvaihtotovereillesi englanniksi?

☐ En ☐ Kyllä

20. Kuinka monta kirjettä kirjoitat englanniksi joka vuosi?

☐ 1-2 kirjettä vuodessa
☐ 3-4 kirjettä vuodessa
☐ 5-8 kirjettä vuodessa
☐ Enemmän kuin 8 kirjettä vuodessa

21. Mistä sinulla on päästötodistus/loppututkinto?

☐ Peruskoulu
☐ Lukio (ylioppilastutkinto)
☐ Opisto
☐ Korkeakoulu 1 (alempi tutkinto)
☐ Korkeakoulu 2 (ylempi tutkinto)

Työpaikka:

JOKO 22. Missä olet töissä tällä hetkellä ja millainen on työnkuvasi?

_____
_____
_____

JA/TAI 23. Jos olet opiskelija,

23a. missä tiedekunnassa opiskelet? _____

23b. missä laitoksessa/osastossa opiskelet? _____

23c. minä vuonna aloitit opiskelusi? ____

24. Käytätkö englannin kielen taitoasi töissä / kotona ?

24a. Töissä	**24b. Kotona**
☐ En koskaan	☐
☐ Harvoin (eli kerran tai pari vuodessa)	☐
☐ Silloin tällöin (eli noin kerran kuukaudessa)	☐
☐ Aika usein (eli noin kerran viikossa)	☐
☐ Säännöllisesti (eli joka päivä tai melkein joka päivä)	☐
☐ Koko ajan	☐

25. Miten muut ihmiset arvioivat mielestäsi tulotasoasi?

- ☐ Erittäin hyvätuloinen
- ☐ Hyvätuloinen
- ☐ Keskituloinen
- ☐ Matalapalkkainen
- ☐ Ilman tuloa tai työtön

26. Mitä työtä isäsi teki, kun kävit lukiota? _____

27. Mitä työtä äitisi teki, kun kävit lukiota? _____

28. Mikä seuraavista kuvaa parhaiten vanhempiesi tulotasoa silloin, kun kävit lukiota?

28a. ISÄN tulotaso

- ☐ Erittäin hyvätuloinen
- ☐ Hyvätuloinen
- ☐ Keskituloinen
- ☐ Matalapalkkainen
- ☐ Ilman tuloa tai työtön

28b. ÄIDIN tulotaso

- ☐ Erittäin hyvätuloinen
- ☐ Hyvätuloinen
- ☐ Keskituloinen
- ☐ Matalapalkkainen
- ☐ Ilman tuloa tai työtön

29. Mihin koulutusryhmään äitisi ja isäsi kuuluivat, kun kävit lukiota?

29a. ISÄN koulutus

- ☐ Ei ammatillista koulutusta
- ☐ Ammattikoulutus tai vastaava
- ☐ Opistotasoinen koulutus tai vastaava
- ☐ Akateeminen koulutus

29b. ÄIDIN koulutus

- ☐ Ei ammatillista koulutusta
- ☐ Ammattikoulutus tai vastaava
- ☐ Opistotasoinen koulutus tai vastaava
- ☐ Akateeminen koulutus

Swedish version

Frågeformulär

1. Kön ☐ Kvinna ☐ Man

2. Nationalitet ☐ Finsk ☐ Svensk ☐ Annan . . . Vilken _____

3. Modersmål ☐ Finska ☐ Svenska ☐ Same ☐ Annat . . .
Vilket _____

4. Födelseår ____

5. Skolgång

5a. På vilken ort gick du i skola?

_____

5b. I vilket län (i vilka län)? _____

6. Hur många år läste du engelska i skolan? ☐ .

7. Hur många år har du studerat engelska i högskolan? ☐ .

8. Hur pass bra kan du engelska jämfört med dina kompisar och andra personer som har studerat engelska lika mycket som du? Är dina kunskaper i engelska . . .

utmärkta? ☐
goda? ☐
medelmåttiga? ☐
hjälpliga? ☐
dåliga? ☐

9. Har du någon gång besökt ett land där engelskan talas som modersmål (t.ex. Storbritannien, USA)?

☐ Aldrig ☐ Ja

10. Om svaret är ja:

10a. - Hur ofta har du besökt ett land där engelskan talas som modersmål?

En gång ☐ Två gånger ☐ Flera gånger ☐

10b. - Hur många veckor har du bott i ett land där engelskan talas som modersmål?

1-4 veckor ☐ 5-8 veckor ☐ 9 veckor eller flera ☐

11. Vilka språk (förutom engelska) har du läst i skolan och/eller i högskolan?

	11a. I skolan	11b. I högskolan
Finska	☐	☐
Tyska	☐	☐
Ryska	☐	☐
Franska	☐	☐
Annat	☐	☐

12. Hur pass goda kunskaper har du i ovannämnda språk jämfört med andra personer?

12a. Finska

☐ utmärkta
☐ goda
☐ medelmåttiga
☐ hjälpliga
☐ dåliga

12b. Tyska

☐ utmärkta
☐ goda
☐ medelmåttiga
☐ hjälpliga
☐ dåliga

12c. Ryska

☐ utmärkta
☐ goda
☐ medelmåttiga
☐ hjälpliga
☐ dåliga

12d. Franska

☐ utmärkta
☐ goda
☐ medelmåttiga
☐ hjälpliga
☐ dåliga

12e. Annat, vilket? .....

☐ utmärkta
☐ goda
☐ medelmåttiga
☐ hjälpliga
☐ dåliga

13. Läser du engelsk- eller svenskspråkiga tidningar, tidskrifter, facktidningar, böcker?

- ☐ Alltid svenskspråkiga.
- ☐ Svenskspråkiga oftare än engelskspråkiga.
- ☐ Båda lika mycket.
- ☐ Engelskspråkiga oftare än svenskspråkiga.
- ☐ Alltid engelskspråkiga.

14. Lyssnar du på engelsk- eller svenskspråkiga LP- och/eller CD-skivor?

- ☐ Alltid svenskspråkiga.
- ☐ Svenskspråkiga oftare än engelskspråkiga.
- ☐ Båda lika mycket.
- ☐ Engelskspråkiga oftare än svenskspråkiga.
- ☐ Alltid engelskspråkiga.

15. Tittar du på engelskspråkiga TV-program?

☐ Nej ☐ Ja

16. Om programmet har en text på ditt modersmål,

läser du alltid hela texten? ☐
läser du största delen av texten? ☐
läser du vanligen ungefär hälften av texten? ☐
läser du bara en bit här, en bit där? ☐
läser du inte alls text medan du tittar? ☐

17. Gå du på bio när det visas engelskspråkiga filmer?

☐ Nej ☐ Ja

18. Om filmen har en text,

läser du alltid hela texten? ☐
läser du största delen av texten? ☐
läser du vanligen ungefär hälften av texten? ☐
läser du bara en bit här, en bit där? ☐
läser du inte alls text medan du tittar? ☐

19. Skriver du till dina vänner på engelska?

☐ Nej ☐ Ja

20. Hur många brev skriver du på engelska varje år?

☐ 1-2 brev per år
☐ 3-4 brev per år
☐ 5-8 brev per år
☐ flera än 8 brev per år

21. Från vilket läroverk har du avgångsbetyg/examensbetyg?

☐ Grundskolan
☐ Gymnasiet (studentexamen)
☐ Institut
☐ Högskolan (lägre examen)
☐ Högskolan (högre examen)

Arbetsplats :

Antingen 22. Var arbetar du för till fället och hurdana arbetsuppgifter har du?

_____
_____
_____

Eller 23. Om du är studerande, . . .

23a. vid vilken fakultet studerar du? _____

23b. på vilken institution/avdelning studerar du?

_____

23c. vilket år började du dina studier? ____

24. Använder du engelska på jobbet / hemma?

24b. På jobbet		24a. Hemma
☐	Aldrig	☐
☐	Sällan (ett par gånger per år)	☐
☐	Då och då (ungefär en gång i månaden	☐
☐	Ganska ofta (ungefär en gång i veckan)	☐
☐	Regelbundet (dagligen eller nästan dagligen)	☐
☐	Hela tiden	☐

25. Vad anser du själv, hur värderar andra människor din inkomstnivå?

☐ Mycket hög
☐ Hög
☐ Medelhög
☐ Låg
☐ Inga inkomster eller arbetslös

26. Vilket arbete hade din pappa när du gick i gymnasiet?

_____

27. Vilket arbete hade din mamma när du gick i gymnasiet?

_____

28. Vilket uttryck av de följande beskriver bäst dina föräldrars inkomstnivå under tiden då du gick i gymnasiet?

28a. PAPPAS inkomstnivå	28b. MAMMAS inkomstnivå
☐ Mycket hög	☐ Mycket hög
☐ Hög	☐ Hög
☐ Medelhög	☐ Medelhög
☐ Låg	☐ Låg
☐ Inga inkomster eller arbetlös	☐ Inga inkomster eller arbetslös

29. Till vilken utbildningsgrupp hörde dina föräldrar då när du gick i gymnasiet?

29a. PAPPAS utbildning	29b. MAMMAS utbildning
☐ Ingen yrkesutbildning	☐ Ingen yrkesutbildning
☐ Yrkesutbildning eller motsvarande	☐ Yrkesutbildning eller motsvarande
☐ Institutnivå eller motsvarande	☐ Institutnivå eller motsvarande
☐ Akademisk/Högskoleutbildning	☐ Akademisk/Högskoleutbildning

REFERENCES

Abercrombie D. 1968. 'Paralanguage.' *British Journal of Disorders of Communication* 3:55–9.

Abbot P. & Sapsford R. 1987. *Women and Social Class*. London: Tavistock Publications.

Aboud F.E., Clément R. & Taylor D.M. 1974. 'Evaluational reactions to discrepancies between social class and language.' *Sociometry* 37,2:239–50.

Addington D.W. 1968. 'The relationship of selected vocal characteristics to personality perception.' *Speech Monographs* 35:492–503.

Agheyisi R. & Fishman J. 1970. 'Language attitude studies: a brief survey of methodological approaches.' *Anthropological Linguistics* 12,5:137–57.

Ajzen I. 1988. *Attitudes, Personality and Behavior*. Milton Keynes: Open University Press

Allport G.W. 1935. 'Attitudes' in C. Murchison (ed.), *A Handbook of Social Psychology*. Clark University Press, Worcester, Massachusetts. Pp. 798–884.

Allport G.W. & Cantril H. 1934. 'Judging personality from voice.' *Journal of Social Psychology* 5:37–55.

Alreck P.L. & Settle R.B. 1995. *The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey*. Chicago: Irwin.

Amabile Y.M. & Glazebrook. 1981. 'A negativity bias in interpersonal evaluation.' *Journal of Experimental Social Psychology* 18:1–22.

Ambady N. & Rosenthal R. 1995. 'Experimenter effects' in A.S.R. Manstead & M. Hewstone (eds), *The Blackwell Encyclopaedia of Social Psychology*. Oxford: Blackwell. Pp. 230–35.

Anderson-Hsieh J., Johnson R. & Koehler K. 1992. 'The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure.' *Language Learning* 42,4:529–55.

Anderson-Hsieh J. & Koehler K. 1988. 'The effect of foreign accent and speaking rate on native speaker comprehension.' *Language Learning* 38,4:561–613.

Anisfeld E. & Lambert W.E. 1964. 'Evaluational reactions of bilingual and monolingual children to spoken languages.' *Journal of Abnormal and Social Psychology*, 69,1:89–97.

Anisfeld M., Bogo N. & Lambert W.E. 1962. 'Evaluational reactions to accented English speech.' *Journal of Abnormal and Social Psychology*, 65,4:223–31.

Apple W., Streeter L.A. & Krauss R.M. 1979. 'Effects of pitch and speech rate on personal attributions.' *Journal of Personality and Social Psychology* 37,5: 715–727.

Argyle M., Salter V., Nicholson H., Williams M. & Burgess P. 1970. 'The communication of inferior and superior attitudes by verbal and non-verbal signals.' *British Journal of Social and Clinical Psychology* 9:222–31.

Arthur B., Farrar D. & Bradford B. 1974. 'Evaluation reactions of college students to dialect differences in the English of Mexican-Americans.' *Language and Speech* 17:255–270.

Asch S.E. 1946. 'Forming impressions of personality.' *Journal of Abnormal and Social Psychology* 41:258–90.

Austin W.M. 1965. 'Some social aspects of paralanguage.' *Canadian Journal of Linguistics* 11:31–9.

Baker C. 1992. *Attitudes and Language*. Clevedon: Multilingual Matters.

Ball P., Byrne J., Giles H., Berechree P., Griffiths J., MacDonald H. & McKendrick I. 1982. 'The retrospective speech halo effect: some Australian Data.' *Language and Communication* 2,3:277–84.

Barber B. 1957. *Social Stratification: a Comparative Analysis of Structure and Process*. New York: Harcourt Brace & World.

Bass B.M. 1955. 'Authoritarianism or acquiescence?' *Journal of Abnormal and Social Psychology* 51:616–23.

Beebe L. & Giles H. 1984. 'Speech accommodation theories: a discussion in terms of second-language acquisition.' *International Journal of the Sociology of Language* 46:5–32.

Bell B.D. & Stanfield G.G. 1973. 'Chronological age in relation to attitudinal judgments: an experimental analysis.' *Journal of Gerontology* 28,4:491–6.

Bem D.J. 1967. 'Self-perception: an alternative interpretation of cognitive dissonance phenomena.' *Psychological Review* 74,3:183–200. [Also reprinted in Warren & Jahoda (eds) 1973. Pp. 74–99.]

Bem D.J. 1968. 'Attitudes as self-descriptions: another look at the attitude–behavior link' in A.G. Greenwald, T.C. Brock & T.M. Ostrom (eds), *Psychological Foundations of Attitudes*. New York: Academic Press. Pp. 197–215.

Bem D.J. 1970. *Beliefs, Attitudes, and Human Affairs*. Belmont, California: Brooks/Cole.

Bernstein B. 1962a. 'Linguistic codes, hesitation phenomena and intelligence.' *Language and Speech* 5:31–46.

Bernstein B. 1962b. 'Social class, linguistic codes and grammatical elements.' *Language and Speech* 5:221–40.

Berry D.S. 1992. 'Vocal types and stereotypes: joint effects of vocal attractiveness and vocal maturity on person perception.' *Journal of Nonverbal Behavior* 16,1:41–54.

Bierhof H.-W. 1989. *Person Perception and Attribution*. Berlin: Springer-Verlag.

Bohrnstedt G.W. & Knoke D. 1988. *Statistics for Social Data Analysis*. 2nd ed. Itasca, Illinois: Peacock Publishers.

Bond Z.S. & Fokes J. 1985. 'The timing of English words by non-native speakers.' *Journal of the Acoustical Society of America* Suppl. 1, 77:S53.

Boswell J. 1967. *Life of Johnson*. 2 vols. London: Everyman's Library, Dent. (First published 1791)

Bourhis R. & Giles H. 1976. 'The language of co-operation in Wales: a field study.' *Language Sciences* 42:13–16.

Bourhis R.Y., Giles H. & Lambert W.E. 1975. 'Social consequences of accommodating one's style of speech: a cross-national investigation.' *International Journal of the Sociology of Language* 6:55–71.

Bradac J.J. & Giles H. 1991. 'Social and educational consequences of language attitudes.' *Moderna Språk* 85,1:1–11.

Breckler S.J. 1984. 'Empirical validation of affect, behavior and cognition as distinct components of attitude.' *Journal of Personality and Social Psychology* 47,6:1191–205. [Also reprinted in M. Hewstone et al. (eds) 1997. Pp. 221–45.]

Brennan E.M. & Brennan J.S. 1981a. 'Accent scaling and language attitudes: reactions to Mexican American English speech.' *Language and Speech* 24,3: 207–21.

Brennan E.M. & Brennan J.S. 1981b. 'Measurements of accent and attitude toward Mexican-American speech.' *Journal of Psycholinguistic Research* 10,5: 487–501.

Brennan E.M., Ryan E.B. & Dawson W.E. 1975. 'Scaling of apparent accentedness by magnitude estimation and sensory modality matching.' *Journal of Psycholinguistic Research* 4,1:27–36.

Brown B.L. 1980. 'Effects of speech rate on personality attributions and competency evaluations' in H. Giles, W.P. Robinson & P.M. Smith (eds), *Language: Social Psychological Perspectives*. Oxford: Pergamon Press. Pp. 293–300.

Brown B.L., Giles H. & Thakerar J.N. 1985. 'Speaker evaluations as a function of speech rate, accent and context.' *Language & Communication* 5,3:207–20.

Brown B.L., Strong W.J. & Rencher A.C. 1973. 'Perceptions of personality from speech: effects of manipulations of acoustical parameters.' *Journal of the Acoustical Society of America* 54:29–35.

Brown B.L., Strong W.J. & Rencher A.C. 1974. 'Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency and variance of fundamental frequency on ratings of personality from speech.' *Journal of the Acoustical Society of America* 55,2:313–18.

Brown J.D. 1988. *Understanding Research in Second Language Learning: a Teacher's Guide to Statistics and Research Design*. Cambridge: Cambridge University Press.

Bryman A. & Cramer D. 1997. *Quantitative Data Analysis with SPSS for Windows: a Guide for Social Scientists*. London: Routledge.

Burgess A. 1982. *Enderby*. Harmondsworth: Penguin.

Busch M. 1993. 'Using Likert scales in L2 research.' *TESOL Quarterly* 24,4: 733–36.

Cairns E. & Duriez B. 1976. 'The influence of speaker's accent on recall by Catholic and Protestant school children in Northern Ireland.' *British Journal of Social and Clinical Psychology* 15:441–2.

Calvert D.R. 1980. *Descriptive Phonetics*. 2nd ed. New York: Georg Thieme.

Cameron D. 1985. *Feminism and Linguistic Theory*. London: Macmillan.

Cameron D. & Coates J. 1988. 'Some problems in the sociolinguistic explanation of sex differences' in J. Coates & D. Cameron (eds), *Women in their Speech Communities*. London: Longman. Pp. 13–26.

Cardno J.A. 1955. 'The notion of "attitude": an historical note.' *Psychological Reports* 1:345–52.

Cargile A.C. & Giles H. 1997. 'Understanding language attitudes: exploring listener affect and identity.' *Language & Communication* 17,3:195–217.

Cargile A.C., Giles H., Ryan E.B. & Bradac J.J. 1994. 'Language attitudes as a social process: a conceptual model and new directions.' *Language & Communication* 14,3:211–36.

Carranza M.A. & Ryan E.B. 1975. 'Evaluative reactions of bilingual Anglo and Mexican American adolescents toward speakers of English and Spanish.' *International Journal of the Sociology of Language* 6:83–104.

Cartier F. 1987. 'Whassa mispernunciation?' *TESOL Quarterly* 21,1:157–8.

Catford J.C. 1987. 'Phonetics and the teaching of pronunciation: a systemic description of English phonology' in J. Morley (ed.), *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, D.C.: TESOL. Pp. 83–100.

Celce-Murcia M., Brinton D.M. & Goodwin J.M. 1996. *Teaching Pronunciation: a Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.

Cercignani F. 1981. *Shakespeare's Works and Elizabethan Pronunciation*. Oxford: Oxford University Press.

Cheyne W.M. 1970. 'Stereotyped reactions to speakers with Scottish and English regional accents.' *British Journal of Social and Clinical Psychology* 9:77–9.

Clark H.H. 1973. 'The language-as-fixed-effect fallacy: a critique of language statistics in psychological research.' *Journal of Verbal Learning and Verbal Behavior* 12,4:335–59.

Cohen A. 1974. 'Mexican-American evaluation judgements about language varieties.' *International Journal of the Sociology of Language* 3:33–51.

Cohen J. 1992. 'A power primer.' *Psychological Bulletin* 112,1:155–9.

Cook S.W. & Selltiz C. 1964. 'A multiple indicator approach to attitude measurement.' *Psychological Bulletin* 62:36–55. [Reprinted in N. Warren & M. Jahoda (eds) 1973. Pp. 354–94.]

Coombs C.H. 1967. 'Thurstone's measurement of social values revisited 40 years later.' *Journal of Personality and Social Psychology* 6:85–90.

Cooper R.L. 1975. 'Introduction.' *International Journal of the Sociology of Language* 6:5–9.

Corder S.P. 1967. 'The significance of learner's (sic) errors.' *IRAL: International Review of Applied Linguistics* 5,4:161–70.

Corder S.P. 1973. *Introducing Applied Linguistics*. Harmondsworth: Penguin.

Côté P. & Clément R. 1994. 'Language attitudes: an interactive situated approach.' *Language & Communication* 14,3:237–51.

Coupland N. 1985. ' "Hark, hark, the lark": social motivations for phonological style-shifting.' *Language & Communication* 5,3:153–71.

Crawford W.W. 1987. 'The pronunciation monitor: L2 acquisition considerations and pedagogical priorities' in J. Morley (ed.), *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, D.C.: TESOL. Pp. 101–21.

Creber C. & Giles H. 1983. 'Social context and language attitudes: the role of formality–informality of the setting.' *Language Sciences* 5,2:155–61.

Cronbach L.J. 1946. 'Response sets and test validity.' *Educational and Psychological Measurement* 6:475–94.

Cronbach L.J. 1950. 'Further evidence of response sets and test design.' *Educational and Psychological Measurement* 10,1:3–31.

Cruttenden A. 1994. *Gimson's Pronunciation of English*. 5th ed. London: Edward Arnold.

Cunningham-Andersson U. 1997. 'Native speaker reactions to non-native speech' in A. James & J. Leather (eds), *Second-Language Speech: Structures and Process*. Berlin: Mouton de Gruyter. Pp. 133–44.

Cunningham-Andersson U. & Engstrand O. 1989. 'Perceived strength and identity of foreign accent in Swedish.' *Phonetica* 46:138–54.

d'Anglejan A. & Tucker G.R. 1973. 'Sociolinguistic correlates of speech style in Quebec' in R. Shuy & R. Fasold (eds), *Language Attitudes: Current Trends and Prospects*. Washington, D.C.: Georgetown University Press. Pp. 1–27.

Davies E.E. 1983. 'Error evaluation: the importance of viewpoint.' *English Language Teaching Journal* 37,4:304–11.

Davis L.M. 1986. 'Sampling and statistical inference in dialectology.' *Journal of English Linguistics* 19,1:42–8.

Davitz J.R. & Davitz L.J. 1959a. 'Correlates of accuracy in the communication of feelings.' *Journal of Communication* 9:110–17.

Davitz J.R. & Davitz L.J. 1959b. 'The communication of feelings by content-free speech.' *Journal of Communication* 6–13.

Davitz J.R. & Davitz L.J. 1961. 'Nonverbal vocal communication of feeling.' *Journal of Communication* 11:81–6.

Dawes R.M. 1972. *Fundamentals of Attitude Measurement*. New York: John Wiley.

DeFleur M.L. & Westie F.R. 1963. 'Attitude as a scientific concept.' *Social Forces* 42:17–31.

Delia J.G. 1972. 'Dialects and the effects of stereotypes on interpersonal attraction and cognitive processes in impression formation.' *Quarterly Journal of Speech* 58,3:285–97.

Derwing T.M. & Munro M.J. 1997. 'Accent, intelligibility and comprehensibility: evidence from four L1s.' *Studies in Second Language Acquisition* 19,1: 1–16.

Diab L.N. 1967. 'Measurement of social attitudes: problems and prospects' in C.W. Sherif & M. Sherif (eds), *Attitude, Ego-Involvement, and Change*. New York: John Wiley. Pp. 140–58.

Diehl C.F., White R. & Burk K.W. 1959. 'Voice quality and anxiety.' *Journal of Speech and Hearing Research* 2,3:282–5.

Dimitrijevic N.R. & Djordjevic D. 1971. 'The reliability of the subjective assessment of the pupils' pronunciation of English as a foreign language.' *IRAL: International Review of Applied Linguistics* 9:245–65.

Dretzke B. 1987. 'Beurteilung von Aussprachefehlern im Englischen.' *Die Neueren Sprachen* 86,6:507–17.

Dretzke B. & Martin J.N. 1975. 'Assessment of a student's phonological command of English—an evaluative chart of phonology.' *Die Neueren Sprache* 4:326–33.

Duncan M.H. 1945. 'An experimental study of some of the relationships between voice and personality among students of speech.' *Speech Monographs* 12: 47–60.

Edwards A.L. 1953. 'The relationship between the judged desirability of a trait and the probability that the trait will be endorsed.' *Journal of Applied Psychology* 37:90–93.

Edwards A.L. 1957a. *Techniques of Attitude Scale Construction*. New York: Appleton-Century-Crofts.

Edwards A.L. 1957b. *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden.

Edwards A.L. & Kilpatrick F.P. 1948. 'A technique for the construction of attitude scales.' *Journal of Applied Psychology* 32:374–84.

Edwards J.R. 1982. 'Language attitudes and their implications among English speakers' in E.B. Ryan & H. Giles (eds), *Attitudes towards Language Variation: Social and Applied Contexts.* London: Edward Arnold. Pp. 20–33.

Eisenstein M. 1983. 'Native reactions to non-native speech: a review of empirical research.' *Studies in Second Language Acquisition* 5,2:160–76.

Eiser J.R. 1980. *Cognitive Social Psychology: a Guidebook to Theory and Research.* London: McGraw-Hill.

Eiser J.R. & van der Pligt J. 1984. 'Accentuation theory, polarization, and the judgment of attitude statements' in J.R. Eiser (ed.), *Attitudinal Judgment.* New York: Springer-Verlag. Pp. 43–63.

El-Dash L. & Tucker G. 1975. 'Subjective reactions to various speech styles in Egypt.' *International Journal of the Sociology of Language* 6:33–54.

Elms A.C. 1976. *Attitudes.* Milton Keynes: The Open University Press.

Elyan O., Smith P., Giles H. & Bourhis R. 1978. 'RP-accented female speech: the voice of perceived androgyny?' in P. Trudgill (ed.), *Sociolinguistic Patterns in British English.* London: Edward Arnold. Pp. 122–31.

Esling J.H. 1978. 'The identification of features of voice quality in social groups.' *Journal of the International Phonetic Association* 8,1/2:18–23.

Esling J.H. & Wong R.F. 1983. 'Voice quality settings and the teaching of pronunciation.' *TESOL Quarterly* 17,1:89–95.

Esser H. 1993. 'Response set: habit, frame or rational choice?' in D. Krebs & P. Schmidt (eds), *New Directions in Attitude Measurement.* Berlin: Walter de Gruyter. Pp. 293–314.

Fabrigar L.R. & Krosnick J.A. 1996. 'Attitude measurement and questionnaire design' in A. Manstead & M. Hewstone (eds), *The Blackwell Encyclopedia of Social Psychology.* Oxford: Blackwell. Pp. 42–7.

Fairbanks G. & Hoaglin L.W. 1941. 'An experimental study of the durational characteristics of the voice during the expression of emotion.' *Speech Monographs* 85–90.

Fasold R. 1984. *The Sociolinguistics of Society.* Oxford: Blackwell.

Fay P.J. & Middleton W.C. 1939a. 'Judgement of occupation from the voice as transmitted over a public address system and over a radio.' *Journal of Applied Psychology* 23:586–601.

Fay P.J. & Middleton W.C. 1939b. 'Judgement of Spranger personality types from the voice as transmitted over a public address system.' *Character Personality* 8:144–45.

Fay P.J. & Middleton W.C. 1940a. 'Judgement of intelligence from the voice as transmitted over a public address system.' *Sociometry* 3:186–91.

Fay P.J. & Middleton W.C. 1940b. 'Judgement of Kretschmerian body types from the voice as transmitted over a public address system.' *Journal of Social Psychology* 12:151–62.

Fay P.J. & Middleton W.C. 1940c. 'The ability to judge the rested or tired condition of a speaker from his voice as transmitted over a public address system.' *Journal of Applied Psychology* 24:645–50.

Fay P.J. & Middleton W.C. 1941a. 'Judgment of emotional balance from the transmitted voice.' *Character and Personality* 10:109–13.

Fay P.J. & Middleton W.C. 1941b. 'The ability to judge sociability from the voice as transmitted over a public address system.' *Journal of Social Psychology* 13:302–09.

Fay P.J. & Middleton W.C. 1941c. 'The ability to judge truth-telling, or lying, from the voice as transmitted over a public address system.' *Journal of General Psychology* 24:211–15.

Fay P.J. & Middleton W.C. 1942a. 'Judgement of introversion from the transcribed voice.' *Quarterly Journal of Speech* 28:226–28.

Fay P.J. & Middleton W.C. 1942b. 'Measurement of the persuasiveness of the transcribed voice.' *Journal of Psychology* 14:259–67.

Fay P.J. & Middleton W.C. 1943. 'Judgment of leadership from the transmitted voice.' *Journal of Social Psychology* 17:99–102.

Fay P.J. & Middleton W.C. 1944. 'Judgment of confidence from voice.' *Journal of General Psychology* 30:93–5.

Fayer J.M. & Krasinski E. 1987. 'Native and nonnative judgments of intelligibility and irritation.' *Language Learning* 37,3:313–26.

Festinger L. 1957. *A Theory of Cognitive Dissonance*. London: Tavistock Publications.

Fielding G. & Evered C. 1980. 'The influence of patients' speech upon doctors: the diagnostic interview' in R.N. St Clair & H. Giles (eds), *The Social and Psychological Contexts of Language*. Hillsdale, N.J.: Erlbaum. Pp. 51–72.

Firth J.R. 1950. 'Personality and language in society.' *Sociological Review* 42: 37–52.

Fishbein M. (ed.) 1967. *Readings in Attitude Theory and Measurement*. New York: Wiley.

Fishbein M. & Ajzen I. 1975. *Belief, Attitude, Intention and Behavior: an Introduction to Theory and Research*. London: Addison-Wesley.

Fitz-Gibbon C.T. & Morris L.L. 1987. *How to Analyze Data*. Newbury Park: Sage.

Flege J.E. 1984. 'The detection of French accent by American listeners.' *Journal of the Acoustical Society of America* 76,3:692–707.

Flege J.E. 1988. 'Factors affecting degree of perceived foreign accent in English sentences.' *Journal of the Acoustical Society of America* 84,1:70–9.

Flege J.E. & Fletcher K.L. 1992. 'Talker and listener effects on degree of perceived foreign accent.' *Journal of the Acoustical Society of America* 91,1: 370–89.

Flores N. & Hopper R. 1975. 'Mexican Americans' evaluations of spoken Spanish and English.' *Speech Monographs* 42,2:91–8.

Frandsen K.D. & Clement D.A. 1984. 'The functions of human communication in informing: communication and processing information' in C.C. Arnold & J.W. Bowers (eds), *Handbook of Rhetorical and Communication Theory*. Boston: Allyn & Bacon.

Fransella F. & Bannister D. 1977. *A Manual for Repertory Grid Technique*. London: Academic Press.

Frender R., Brown B. & Lambert W.E. 1970. 'The role of speech characteristics in scholastic success.' *Canadian Journal of Behavioural Science* 2,4:299–306.

Gallois C. & Callan V.J. 1981. 'Personality impressions elicited by accented English speech.' *Journal of Cross-Cultural Psychology* 12,3:347–59.

Gardner M. (ed.) 1970. *The Annotated Alice (Lewis Carroll)*. Harmondsworth: Penguin.

Gardner R.C. & Taylor D.M. 1968. 'Ethnic stereotypes: their effects on person perception.' *Canadian Journal of Psychology* 22,4:267–76.

Gaudio R.P. 1994. 'Sounding gay: pitch properties in the speech of gay and straight men.' *American Speech* 69,1:30–57.

Giles H. 1970. 'Evaluative reactions to accents.' *Educational Review* 22: 211–227.

Giles H. 1971a. *A Study of Speech Patterns in Social Interaction: accent evaluation and accent change*. Unpub. Ph.D. Thesis. University of Bristol.

Giles H. 1971b. 'Ethnocentrism and the evaluation of accented speech.' *British Journal of Social and Clinical Psychology* 10:187–8.

Giles H. 1971c. 'Patterns of evaluation to R.P., South Welsh and Somerset accented speech.' *British Journal of Social and Clinical Psychology* 10:280–1.

Giles H. 1971d. 'Teachers' attitudes towards accent usage and change.' *Educational Review* 24:11–25.

Giles H. 1972a. 'Evaluation of personality content from accented speech as a function of listeners' social attitudes.' *Perceptual and Motor Skills* 34:168–70.

Giles H. 1972b. 'The effect of stimulus mildness–broadness in the evaluation of accents.' *Language and Speech* 15,3:262–9.

Giles H. 1973a. 'Accent mobility: a model and some data.' *Anthropological Linguistics* 15:87–105.

Giles H. 1973b. 'Communicative effectiveness as a function of accented speech.' *Speech Monographs* 40:330–31.

Giles H. 1980. 'Accommodation theory: some new directions' in S. de Silva (ed.), *Aspects of Linguistic Behaviour*. York: York University Press. Pp. 105–36.

Giles H., Baker S. & Fielding G. 1975. 'Communication length as a behavioral index of accent prejudice.' *International Journal of the Sociology of Language* 6:73–81.

Giles H. & Bourhis R.Y. 1973. 'Dialect perception revisited.' *The Quarterly Journal of Speech* 59,3:337–42.

Giles H. & Bourhis R.Y. 1976. 'Methodological issues in dialect perception: some social psychological perspectives.' *Anthropological Linguistics* 18,7: 294–304.

Giles H., Bourhis R. & Davies A. 1979. 'Prestige speech styles: the imposed norm and inherent value hypotheses' in W.C. McCormack & S.A. Wurm (eds), *Language and Society: Anthropological Issues*. The Hague: Mouton. Pp. 589–96. (First published in 1975 in *Language in Anthropology IV: Language in Many Ways*. Same editors, same publisher.)

Giles H., Bourhis R.Y. & Taylor D.M. 1977. 'Towards a theory of language in ethnic group relations' in H. Giles (ed.), *Language, Ethnicity and Intergroup Relations*. London: Academic Press. Pp. 307–48.

Giles H., Bourhis R., Trudgill P. & Lewis A. 1974. 'The imposed norm hypothesis: a validation.' *The Quarterly Journal of Speech* 60,4:405–10.

Giles H. & Byrne J.L. 1982. 'An intergroup approach to second language acquisition.' *Journal of Multilingual and Multicultural Development* 3,1:17–40.

Giles H. & Coupland N. 1991. *Language: Contexts and Consequences*. Milton Keynes: The Open University.

Giles H. & Farrar K. 1979. 'Some behavioural consequences of speech and dress styles.' *British Journal of Social and Clinical Psychology* 18:209–10.

Giles H., Harrison C., Creber C., Smith P.M. & Freeman N.H. 1983. 'Developmental and contextual aspects of children's language attitudes.' *Language & Communication* 3,2:141–6.

Giles H. & Marsh P. 1979. 'Perceived masculinity, androgyny and accented speech.' *Language Sciences* 1,2:301–15.

Giles H. & Powesland P. 1975. *Speech Style and Social Evaluation*. London: Academic Press.

Giles H. & Sassoon C. 1983. 'The effect of speaker's accent, social class background and message style on British listeners' social judgements.' *Language & Communication* 3,3:305–13.

Giles H. & Smith P.M. 1979. 'Accommodation theory: optimal levels of convergence' in H. Giles & R. St Clair (eds), *Language and Social Psychology*. Oxford: Blackwell. Pp. 45–65.

Giles H., Smith P., Browne C., Whiteman S. & Williams J. 1980. 'Women's speech: the voice of feminism' in S. McConnell-Ginet, R. Borker & N. Furman (eds), *Women and Language in Literature and Society*. New York: Praeger. Pp. 150–6.

Giles H., Smith P.M., Ford B., Condor S. & Thakerar J.N. 1980. 'Speech style and the fluctuating salience of sex.' *Language Sciences* 2,2:260–82.

Giles H., Taylor D.M. & Bourhis R. 1973. 'Towards a theory of interpersonal accommodation through language: some Canadian data.' *Language in Society* 2:177–92.

Giles H., Williams A., Mackie D.M. & Rosselli F. 1995. 'Reactions to Anglo- and Hispanic-American accented speakers: affect, identity, persuasion, and the English-only controversy.' *Language & Communication* 15,2:107–20.

Giles H., Wilson P. & Conway A. 1981. 'Accent and lexical diversity as determinants of impression formation and perceived employment suitability.' *Language Sciences* 3,1:91–103.

Goldman-Eisler F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.

Gordon E. & Abell M. 1990. ' "This objectionable colonial dialect": historical and contemporary attitudes to New Zealand speech' in A. Bell & J. Holmes (eds), *New Zealand Ways of Speaking English*. Clevedon: Multilingual Matter. Pp. 21–48.

Graddol D. & Swann J. 1989. *Gender Voices*. Oxford: Blackwell.

Greenwald A.G. 1968. 'On defining attitude and attitude theory' in A.G. Greenwald et al., *Psychological Foundations of Attitudes*. New York: Academic Press. Pp. 361–388.

Greenwald A.G., Brock T.C. & Ostrom T.M. (eds). 1968. *Psychological Foundations of Attitudes*. New York: Academic Press.

Grosjean F. 1980. 'Comparative studies of temporal variables in spoken and sign languages: a short review' in H.W. Dechert & M. Raupach (eds), *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton. Pp. 307–12.

Guttman L. 1944. 'A basis for scaling qualitative data.' *American Sociological Review* 9:139–150. [Reprinted in G.F. Summers 1977. Pp. 174–86.]

Guttman L. 1947. 'The Cornell technique for scale and intensity analysis.' *Educational and Psychological Measurement* 7:247–80. [Excerpted in G.F. Summers 1977.]

Guttman L. 1954. 'The principal components of scalable attitudes' in P.F. Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences*. Glencoe, Illinois: Free Press.

Gynan S.N. 1985. 'Comprehension, irritation and error hierarchies.' *Hispania* 68:160–5.

Hargreaves W.A. & Starkweather J.A. 1964. 'Voice quality changes in depression.' *Language and Speech* 7:84–8.

Hargreaves W.A., Starkweather J.A. & Blacker K.H. 1965. 'Voice quality in depression.' *Journal of Abnormal Psychology* 70,3:218–20.

Hatch E. & Farhady H. 1982. *Research Design and Statistics for Applied Linguistics*. Rowley, Massachusetts: Newbury House.

Hatch E. & Lazaraton A. 1991. *The Research Manual: Design and Statistics for Applied Linguistics*. Boston: Heinle & Heinle.

Heider F. 1946. 'Attitudes and cognitive organization.' *Journal of Psychology* 21:107–12.

Heise D.R. 1977. 'The semantic differential and attitude research' in G.F. Summers (ed.), *Attitude Measurement*. London: Kershaw. Pp. 235–53.

Helfrich H. 1979. 'Age markers in speech' in K.R. Scherer & H. Giles (eds), *Social Markers in Speech*. Cambridge: Cambridge University Press. Pp. 63–107.

Helson H. 1947. 'Adaptation-level as frame of reference for prediction of psychophysical data.' *American Journal of Psychology* 60,1:1–29.

Helson H. 1959. 'Adaptation level theory' in S. Koch (ed.), *Psychology: a Study of a Science, Vol. 1*. New York: McGraw-Hill. Pp. 565–621.

Henton C.G. 1989. 'Fact and fiction in the description of female and male pitch.' *Language & Communication* 9,4:299–311.

Herzog H. 1933. 'Stimme und Persönlichkeit.' *Zeitschrift für Psychologie* Bd 130, Heft 3 bis 5 300–69.

Hewstone M., Manstead A.S.R. & Stroebe W. (eds). 1997. *The Blackwell Reader in Social Psychology*. Oxford: Blackwell.

Higgins J. 1996. 'Minimal Pairs.' *John Higgins's Home Page*. http://www.stir.ac.uk/epd/celt/staff/higdox/wordlist/phonfreq.htm

Honey J. 1989. *Does Accent Matter? – The Pygmalion Factor*. London, Boston: Faber & Faber.

Hopper R. & Williams F. 1973. 'Speech characteristics and employability.' *Speech Monographs* 40:296–302.

Hovland C.I. & Sherif M. 1952. 'Judgmental phenomena and scales of attitude measurement: item displacement in Thurstone scales.' *Journal of Abnormal and Social Psychology* 47:822–32.

Hultfors P. 1986. *Reactions to Non-Native English: Native English-Speakers' Assessments of Errors in the Use of English Made by Non-Native Users of the Language. Part 1*. Stockholm: Almqvist & Wiksell.

Hultfors P. 1987. *Reactions to Non-Native English: Native English-Speakers' Assessments of Errors in the Use of English Made by Non-Native Users of the Language. Part 2*. Stockholm: Almqvist & Wiksell.

Hyrkstedt I. & Kalaja P. 1998. 'Attitudes toward English and its functions in Finland: a discourse-analytic study.' *World Englishes* 17,3:345–57.

Hänninen R. 1979. 'The voiceless–voiced opposition of English consonants: difficulties of pronunciation and perception in communication between native and Finnish speakers' in J. Lehtonen & K. Sajavaara (eds), *Papers in Contrastive Phonetics (Jyväskylä Cross-Language Studies, No. 7)*. Jyväskylä. Pp. 93–164.

I.P.A. *The Principles of the International Phonetic Association*. London: I.P.A., University College. [Revised versions of the International Phonetic Alphabet are to be found in A. Cruttenden, 1994, P. 32, and in J. Laver, 1994, P. 593.]

James C. 1972. 'The diagnosis and analysis of error: some insights from linguistics.' *Audio-visual Language Journal* 10:75–9.

James C. 1977. 'Judgements of error gravities.' *English Language Teaching Journal* 31,2:116–24.

Jaspars J.M.F. 1978. 'The nature and measurement of attitudes' in H. Tajfel & C. Fraser (eds), *Introducing Social Psychology*. Harmondsworth: Penguin. Pp. 256–76.

Johansson S. 1973. 'The identification and evaluation of errors in foreign languages: a functional approach' in J. Svartvik (ed.) *ERRATA: Papers in Error Analysis*. Lund. Pp. 102–14.

Johansson S. 1978. *Studies of Error Gravity: Native Reactions to Errors Produced by Swedish Learners of English*. Göteborg: Acta Universitatis Gothoburgensis.

Jones R.H. 1997. 'Beyond "Listen and repeat": pronunciation teaching materials and theories of second language acquisition.' *System* 25,1:103–12.

Jones D., Roach P. & Hartman J. (eds). 1997. *English Pronouncing Dictionary*. Cambridge: Cambridge University Press.

Kalin R. & Rayko D.S. 1978. 'Discrimination in evaluative judgments against foreign-accented job candidates.' *Psychological Reports* 43:1203–9.

Kalin R., Rayko D.S. & Love N. 1980. 'The perception and evaluation of job candidates with four different ethnic accents' in H. Giles, W.P. Robinson & P.M. Smith, *Language: Social Psychological Perspectives*. Oxford: Pergamon. Pp. 197–202.

Kaplan K.J. 1972. 'On the ambivalence-indifference problem in attitude theory and measurement: a suggested modification of the semantic differential technique.' *Psychological Bulletin* 77,5:361–72.

Kelley H. 1973. 'The processes of causal attribution.' *American Psychologist* 28: 107–28.

Kelley H.H. & Michela J.L. 1980. 'Attribution theory and research.' *Annual Review of Psychology* 31:457–501.

Kelly G.A. 1955. *The Psychology of Personal Constructs*. New York: Norton.

Kerlinger F.N. 1973. *Foundations of Behavioral Research*. 2nd ed. New York: Holt, Rinehart & Winston.

Kidder L.H. & Campbell D.T. 1977. 'The indirect testing of social attitudes' in G.F. Summers (ed.), *Attitude Measurement*. London: Kershaw. Pp. 333–85.

Kiesler C.A., Collins B.E. & Miller N. 1969. *Attitude Change: a Critical Analysis of Theoretical Approaches*. New York: John Wiley.

Kline P. 1993. *The Handbook of Psychological Testing*. London: Routledge.

Klopfer F.J. & Madden T.M. 1980. 'The middlemost choice on attitude items: ambivalence, neutrality, or uncertainty?' *Personality and Social Psychology Bulletin* 6,1:97–101.

Knowles G. 1987. *Patterns of Spoken English: an Introduction to English Phonetics*. London: Longman.

Kramer E. 1963. 'Judgment of personal characteristics and emotions from nonverbal properties of speech.' *Psychological Bulletin* 60,4:408–20.

Kramer E. 1964. 'Personality stereotypes in voice: a reconsideration of the data.' *Journal of Social Psychology* 62:247–51.

Kramer C. 1977. 'Perceptions of female and male speech.' *Language and Speech* 20,2:151–61.

Kristiansen T. & Giles H. 1992. 'Compliance-gaining as a function of accent: public requests in varieties of Danish.' *International Journal of Applied Linguistics* 2,1:17–35.

Laaksonen K. & Lieko A. 1988. *Suomen kielen äänne- ja muoto-oppi*. Helsinki: Finn Lectura.

Labov W. 1965. 'Stages in the acquisition of standard English' in R. Shuy (ed.), *Social Dialects and Language Learning*. Champaign, Illinois: National Council of Teachers of English. Pp. 77–103.

Labov W. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Ladefoged P. 1975. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Lambert W.E. 1967. 'A social psychology of bilingualism.' *Journal of Social Issues* 23,2:91–109.

Lambert W.E., Anisfeld M. & Yeni-Komshian G. 1965. 'Evaluational reactions of Jewish and Arab adolescents to dialect and language variations.' *Journal of Personality and Social Psychology* 2,1:84–90.

Lambert W.E., Frankel H. & Tucker G.R. 1966. 'Judging personality through speech: a French-Canadian example.' *Journal of Communication* 16:305–21.

Lambert W.E., Giles H. & Picard O. 1975. 'Language attitudes in a French-American community.' *International Journal of the Sociology of Language* 4: 127–52.

Lambert W.E., Hodgson R.C., Gardner R.C. & Fillenbaum S. 1960. 'Evaluational reactions to spoken languages.' *Journal of Abnormal and Social Psychology* 60,1:44–51.

Lamminmäki R. 1979. 'The discrimination and identification of English vowels, consonants, junctures and sentence stress by Finnish comprehensive school pupils' in J. Lehtonen & K. Sajavaara (eds), *Papers in Contrastive Phonetics (Jyväskylä Cross-Language Studies, 7)*. Jyväskylä. Pp. 165–231.

Laver J. 1968. 'Voice quality and indexical information.' *British Journal of Disorders of Communication* 3:43–54.

Laver J. 1974. 'Labels for voices.' *Journal of the International Phonetic Association* 4:62–75.

Laver J. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Laver J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Laver J. & Trudgill P. 1979. 'Phonetic and linguistic markers in speech' in K.R. Scherer & H. Giles (eds), *Social Markers in Speech.* Cambridge: Cambridge University Press. Pp. 1–32.

Lee R.R. 1971. 'Dialect perception: a critical review and re-evaluation.' *Quarterly Journal of Speech* 57,4:410–17.

Lee H.O. & Boster F.J. 1992. 'Collectivism–individualism in perceptions of speech rate: a cross-cultural comparison.' *Journal of Cross-Cultural Psychology* 23,3:377–88.

Lehiste I. 1970. *Suprasegmentals*. Cambridge MA: Massachusetts Institute of Technology Press.

Lehtonen J. & Koponen M. 1977. 'Signalling of morphophonological boundaries by Finnish speakers of English: preliminary findings' in K. Sajavaara & J. Lehtonen (eds), *Contrastive Papers: Jyväskylä Studies, 4*. Jyväskylä. Pp. 75–87.

Lehtonen J. & Sajavaara K. 1985. 'The silent Finn' in D. Tannen & M. Saville-Troike (eds), *Perspectives on Silence*. Norwood, N.J.: Ablex.

Lehtonen J., Sajavaara K. & May A. 1977. *Spoken English: the Perception and Production of English on a Finnish–English Contrastive Basis*. Jyväskylä: Gummerus.

Lemon N. 1973. *Attitudes and their Measurement*. London: Batsford.

Likert R.A. 1932. 'A technique for the measurement of attitudes.' *Archives of Psychology* 140:40–53. [Excerpted in G.F. Summers 1977. Pp. 149–58.]

Lind E.A. & O'Barr W.M. 1979. 'The social significance of speech in the courtroom' in H. Giles and R. St. Clair (eds), *Language and Social Psychology*. Oxford: Blackwell. Pp. 66–87.

Loewenthal K.M. 1996. *An Introduction to Psychological Tests and Scales*. London: UCL Press.

Lord C.G., Ross L. & Lepper M.R: 1979. 'Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence.' *Journal of Personality and Social Psychology* 37:2098–109.

Ludwig J. 1982. 'Native-speaker judgments of second-language learners' efforts at communication: a review.' *Modern Language Journal* 66,3:274–83.

Luppescu S. & Day R.R. 1990. 'Examining attitude in teachers and students: the need to validate questionnaire data.' *Second Language Research* 6,2:125–34.

Macaulay R.K.S. 1976. 'Social class and language in Glasgow.' *Language in Society* 5:173–88.

Macaulay R.K.S. 1977. *Language, Social Class, and Education: a Glasgow Study*. Edinburgh: Edinburgh University Press.

McCretton E. & Rider N. 1993. 'Error gravity and error hierarchies.' *IRAL: International Review of Applied Linguistics* 31,3:177–88.

McGuire W.J. 1968. 'The nature of attitudes and attitude change' in G. Lindzey & E. Aronson (eds), *The Handbook of Social Psychology, Vol. 3*. Reading, Massachusetts: Addison-Wesley. Pp. 136–314.

McKennell A.C. 1970. 'Attitude measurement: use of alpha with cluster or factor analysis.' *Sociology* 4:227–45.

McNemar Q. 1946. 'Opinion-attitude methodology.' *Psychological Bulletin* 43: 289–374.

Mallory E.B. & Miller V.R. 1958. 'A possible basis for the association of voice characteristics and personality traits.' *Speech Monographs* 25:255–60.

Manstead A.S.R. & Hewstone M. (eds). 1996. *The Blackwell Encyclopedia of Social Psychology.* Oxford: Blackwell.

Marjomaa I. 1985a. 'It's not how long you make them—or is it? Englishmen and Finns producing one another's vowels' in *CDEF 84: Papers from the Conference of Departments of English in Finland.* Turku: Turun yliopiston englantilaisen filologian julkaisuja. Pp. 177–89.

Marjomaa I. 1985b. 'Mitta vai laatu: englannin ja suomen vokaalit hallittuina ja opittavina' in *Approaches to the Teaching of the English Language and Culture II.* Oulu: Publications of the Department of English, University of Oulu. Pp. 42–61.

Markel N.N. 1965. 'The reliability of coding paralanguage: pitch, loudness and tempo.' *Journal of Verbal Learning and Verbal Behavior* 4:306–308.

Markel N.N. 1969. 'Relationship between voice-quality profiles and MMPI profiles in psychiatric patients.' *Journal of Abnormal Psychology* 74,1:61–6.

Markel N.N., Eisler R.M. & Reese H.W. 1967. 'Judging personality from dialect.' *Journal of Verbal Learning and Verbal Behavior* 6:33–5.

Markel N.N., Meisels M. & Houck J.E. 1964. 'Judging personality from voice quality.' *Journal of Abnormal and Social Psychology* 69,4:458–63.

Markel N.N. & Roblin G.L. 1965. 'The effect of content and sex-of-judge on judgments of personality from voice.' *International Journal of Social Psychiatry* 11,4:295–300.

Masterson J., Mullins E. & Mulvihill A. 1983. 'Components of evaluative reactions to varieties of Irish accents.' *Language and Speech* 26,3:215–31.

Matthews W. 1938. *Cockney Past and Present: a Short History of the Dialect of London.* London: Routledge.

Mazzarella S. 1971. 'Mistakes commonly made by Finnish-speaking people in pronouncing English.' *Tempus* 1:19–21.

Meddis R. 1975. *Statistical Handbook for Non-Statisticians.* London/New York: McGraw Hill.

Mehrabian A. & Ferris S.R. 1967. 'Inference of attitudes from nonverbal communication in two channels.' *Journal of Consulting Psychology* 31,3: 248–52.

Mehrabian A. & Williams M. 1969. 'Nonverbal concomitants of perceived and intended persuasiveness.' *Journal of Personality and Social Psychology* 13,1: 37–58.

Meyerhoff M. 1998. 'Accommodating your data: the use and misuse of accommodation theory in linguistics.' *Language & Communication* 18:205–25.

Miller A.G., McHoskey J.W., Bane C.M. & Dowd T.G. 1993. 'The attitude polarization phenomenon: role of response measure, attitude extremity, and behavioral consequences of reported attitude change.' *Journal of Personality and Social Psychology* 64,4:51–74.

Miller N., Maruyama G., Beaber R.J. & Valone K. 1976. 'Speed of speech and persuasion.' *Journal of Personality and Social Psychology* 34,4:615–24.

Milroy L. 1987a. *Language and Social Networks.* 2nd ed. Oxford: Blackwell.

Milroy L. 1987b. *Observing and Analysing Natural Language.* Oxford: Blackwell.

Moisio R. & Valento E. 1976. 'Testing Finnish schoolchildren's learning of English consonants.' *Jyväskylä Studies 3.* Jyväskylä: University of Jyväskylä.

Morley J. (ed.). 1987. *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, D.C.: TESOL.

Morley J. 1994. 'A multidimensional curriculum design for speech-pronunciation instruction' in J. Morley (ed.), *Pronunciation Pedagogy and Theory: New Views, New Directions*. Alexandria, VA: TESOL.

Morris-Wilson I. 1992. *English Segmental Phonetics for Finns*. Loimaa: Finn Lectura.

Morris-Wilson I. 1997. 'Attitudes of British and Finnish listeners to specific segmental features of mispronunciation in Finnish-accented English' in J. Leather & A. James (eds), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*. Klagenfurt: University of Klagenfurt. Pp. 230–9.

Moser C.A. & Kalton G. 1971. *Survey Methods in Social Investigation.* 2nd ed. London: Heinemann.

Moses P.J. 1941a. 'Social Adjustment and the voice.' *Quarterly Journal of Speech* 27:532–6.

Moses P.J. 1941b. 'Theories regarding the relation of constitution and character through the voice.' *Psychological Bulletin* 38:746(A).

Moses P.J. 1942. 'The study of personality from records of the voice.' *Journal of Consulting Psychology* 6:257–61.

Mugglestone L. 1995. *Talking Proper: The Rise of Accent as Social Symbol.* Oxford: Oxford University Press.

Mulac A., Hanley T.D. & Prigge D.Y. 1974. 'Effects of phonological speech foreignness upon three dimensions of attitude of selected American listeners.' *Quarterly Journal of Speech* 60,4:411–420.

Munn P. & Drever E. 1990. *Using Questionnaires in Small-Scale Research: a Teacher's Guide*. Edinburgh: Scottish Council for Research in Education.

Munro M.J. & Derwing T.M. 1994. 'Evaluations of foreign accent in extemporaneous and read material.' *Language Testing* 11,3:253–66.

Munro M.J. & Derwing T.M. 1995. 'Foreign accent, comprehensibility and intelligibility in the speech of second language learners.' *Language Learning* 45:73–97.

Niemi J. 1979. 'Noun and noun phrase stress: a phonetic study of English supplemented with an error analysis using Finnish speaker–hearers, 1. production.' *Language and Speech* 22,3:221–35.

Nisbett R.E. & Wilson T.D. 1977. 'The halo effect: evidence for unconscious alteration of judgments.' *Journal of Personality and Social Psychology* 35,4: 250–6.

Norell P. 1989. 'Native English-speaking people's reaction to some Swedish pronunciation errors.' *Proceedings from the Fourth Nordic Conference for English Studies, 1989*. Copenhagen. Pp. 255–64.

Norušis M.J. & SPSS Inc. 1993. *SPSS for Windows Base System User's Guide Release 6.0.* Chicago, Illinois: SPSS Inc.

Nunnally J. 1961. *Popular Conceptions of Mental Health*. New York: Holt, Rinehart & Winston.

Oakley A. & Oakley R. 1979. 'Sexism in official statistics' in J. Irvine, I. Miles & J. Evans (eds), *Demystifying Social Statistics*. London: Pluto Press. Pp. 172–89.

OPCS: Office of Population Censuses and Surveys. 1980. *Classification of Occupations and Coding Index*. HMSO.

OPCS: Office of Population Censuses and Surveys. 1986. *1979–80, 1982–83: Occupational Mortality: Decennial Supplement*. HMSO.

Oppenheim A.N. 1992. *Questionnaire Design, Interviewing and Attitude Measurement*. London & New York: Pinter.

Osgood C.E. 1964. 'Semantic differential technique in the comparative study of cultures.' *American Anthropologist* 66(special issues):171–200.

Osgood C.E. 1965. 'Cross-cultural comparability in attitude measurement via multilingual semantic differentials' in I. Steiner & M. Fishbein (eds), *Current Studies in Social Psychology*. New York: Holt, Rinehart & Winston. [Reprinted in M. Fishbein (ed.), 1967, *Readings in Attitude Theory and Measurement*. New York: Wiley. Pp. 108–116.]

Osgood C.E., Suci G.J. & Tannenbaum P.H. 1957. *The Measurement of Meaning*. Urbana, Illinois: University of Illinois Press.

Oxford Advanced Learner's Dictionary of Current English. 1989. Hornby A.S. (ed.) [4th ed., A.P. Cowie (ed.)]. Oxford: Oxford University Press.

Paltridge J. & Giles H. 1984. 'Attitudes towards speakers of regional accents of French: effects of regionality, age and sex of listeners.' *Linguistische Berichte* 90:71–85.

Parish C. 1977. 'A practical philosophy of pronunciation.' *TESOL Quarterly* 11,3:311–17.

Pear T.H. 1931. *Voice and Personality*. London: Chapman & Hall.

Pennington M.C. 1994. 'Recent research in L2 phonology: implications for practice' in J. Morley (ed.), *Pronunciation Pedagogy and Theory: New Views, New Directions*. Alexandria, VA: TESOL. Pp. 92–108.

Pennington M.C. 1996. *Phonology in English Language Teaching: an International Approach*. London & New York: Longman.

Perrin E.H. 1954. 'The rating of defective speech by trained and untrained observers.' *Journal of Speech and Hearing Disorders* 19,1:48–51.

Phillis J.A. 1970. 'Children's judgments of personality on the basis of voice quality.' *Developmental Psychology* 3,3:411.

Piaget J. 1971. *The Language and Thought of the Child*. London: Routledge & Kegan Paul. (1st Published 1926)

Pihko M-K. 1997. *'His English Sounded Strange': the Intelligibility of Native and Non-Native English Pronunciation to Finnish Learners of English*. Jyväskylä: Centre for Applied Language Studies, University of Jyväskylä.

Pittam J. & Gallois C. 1986. 'Predicting impressions of speakers from voice quality: acoustic and perceptual measures.' *Journal of Language & Social Psychology* 5,4:233–247.

Pittenger R.E. 1957. 'Linguistic analysis of tone of voice in communication of affect.' *Psychiatric Research Reports* 8:41–54.

Potter J.P. & Wetherell M. 1987. *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. London: Sage Publications.

Prator C. 1971. 'Phonetics vs. phonemics in the ESL classroom: when is allophonic accuracy important?' *TESOL Quarterly* 5,1:61–72.

Ptacek P.H. & Sander E.K. 1966. 'Age recognition from voice.' *Journal of Speech and Hearing Research* 9,2:273–7.

Quirk R. & Svartvik J. 1966. *Investigating Linguistic Acceptability*. The Hague: Mouton.

Rahkonen M. & Juurakko T. 1998. 'Logit analysis in L2 research: measuring L1 and L2/Ln effects.' *International Journal of Applied Linguistics* 8,1:81–111.

Ramsaran S.M. 1978. *Phonetic and Phonological Correlates of Style in English: a Preliminary Investigation*. PhD dissertation, University of London.

Reid I. 1977. *Social Class Differences in Britain: a Sourcebook*. London: Open Books.

Rey A. 1977. 'Accent and employability: language attitudes.' *Language Sciences* 47:7–12.

Reynolds H.T. 1977. *Analysis of Nominal Data*. Beverly Hills: Sage.

Rifkin B. & Roberts F.D. 1995. 'Error gravity: a critical review of research design.' *Language Learning* 45,3:511–37.

Roach P. 1991. *English Phonetics and Phonology* (2nd ed). Cambridge: Cambridge University Press.

Robinson W.P. 1972. *Language and Social Behaviour*. Harmondsworth: Penguin.

Romaine S. 1980a. 'A critical overview of the methodology of urban British sociolinguistics.' *English Worldwide* 1:163–98.

Romaine S. 1980b. 'Stylistic variation and evaluative reactions to speech: Problems in the investigation of linguistic attitudes in Scotland.' *Language and Speech* 23,3:213–32.

Romaine S. 1982. *Socio-historical Linguistics: its Status and Methodology*. Cambridge: Cambridge University Press.

Rosenberg M.J. & Hovland C.I. 1960. 'Cognitive, affective, and behavioral components of attitudes' in M.J. Rosenberg, C.I. Hovland et al. (eds), *Attitude Organization and Change: an Analysis of Consistency among Attitude Components*. New Haven, Connecticut: Yale University Press. Pp. 1–14.

Rosenberg M.J., Hovland C.I., McGuire W.J., Abelson R.P. & Brehm J.W. (eds). 1960. *Attitude Organization and Change: an Analysis of Consistency among Attitude Components*. New Haven, Connecticut: Yale University Press.

Ruesch J. & Prestwood A.R. 1949. 'Anxiety: its initiation, communication and interpersonal management.' *Archives of Neurology and Psychiatry* 62,5: 527–50.

Ryan E.B. 1973. 'Subjective reactions toward accented speech' in R.W. Shuy & R.W. Fasold (eds), *Language Attitudes: Current Trends and Prospects*. Washington D.C.: Georgetown University Press. Pp. 60–73.

Ryan E.B. 1979. 'Why do low-prestige language varieties persist?' in H. Giles & R. St. Clair (eds), *Language and Social Psychology*. Oxford: Blackwell. Pp. 145–57.

Ryan E.B. 1983. 'Social psychological mechanisms underlying native speaker evaluations of non-native speech.' *Studies in Second Language Acquisition* 5,2:148–59.

Ryan E.B. & Bulik C.M. 1982. 'Evaluations of middle class and lower class speakers of standard American and German-accented English.' *Journal of Language and Social Psychology* 1,1:51–61.

Ryan E.B. & Capadano H.L. 1978. 'Age perceptions and evaluative reactions toward adult speakers.' *Journal of Gerontology* 33,1:98–102.

Ryan E.B. & Carranza M.A. 1975. 'Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English.' *Journal of Personality and Social Psychology* 31,5:855–63.

Ryan E.B. & Carranza M. 1977. 'Ingroup and outgroup reactions to Mexican American language varieties' in H. Giles (ed.), *Language, Ethnicity and Intergroup Relations*. London: Academic Press. Pp. 59–82.

Ryan E.B., Carranza M.A. & Moffie R.W. 1975. 'Mexican American reactions to accented English' in J.W. Berry & W.J. Lonner (eds), *Applied Cross-Cultural Psychology: Selected Papers from the Second International Conference of the International Association for Cross-Cultural Psychology*. Amsterdam: Swets & Zeitlinger. Pp. 174–8.

Ryan E.B., Carranza M.A. & Moffie R.W. 1977. 'Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals.' *Language and Speech* 20,3:267–73.

Ryan E.B. & Sebastian R.J. 1980. 'The effects of speech style and social class background on social judgements of speakers.' *British Journal of Social and Clinical Psychology* 19:229–33.

Samarin W.J. & Kalmar I. 1979. 'Evaluational reactions to foreign accent among immigrants in Toronto' in W.F. Mackey & J. Ornstein (eds), *Sociolinguistic Studies in Language Contact: Methods and Cases*. The Hague: Mouton. Pp. 181–96.

Sanford F.H. 1942a. 'Speech and personality.' *Psychological Bulletin* 39: 811–45.

Sanford F.H. 1942b. 'Speech and personality: a comparative case study.' *Character and Personality* 10:169–98.

Santos T. 1987. 'Markedness theory and error evaluation: an experimental study.' *Applied Linguistics* 8,3:207–18.

Sapir E. 1927. 'Speech as a personality trait.' *American Journal of Sociology* 32:892–905.

Saris W.E. 1993. 'Attitude measurement: is there still hope?' in D. Krebs & P. Schmidt (eds), *New Directions in Attitude Measurement*. Berlin: Walter de Gruyter. Pp. 187–205.

Saunders P. 1990. *Social Class and Stratification*. London: Routledge.

Scherer K.R. 1979a. 'Personality markers in speech' in K.R. Scherer & H. Giles (eds), *Social Markers in Speech*. Cambridge: Cambridge University Press. Pp. 147–209.

Scherer K.R. 1979b. 'Voice and speech correlates of perceived social influence in simulated juries' in H. Giles & R. St Clair (eds), *Language and Social Psychology*. Oxford: Blackwell. Pp. 88–120.

Schwarz N. & Wellens T.R. 1996. 'Survey Methods' in A.S.R. Manstead & M. Hewstone (eds), *The Blackwell Encyclopedia of Social Psychology*. Oxford: Blackwell. Pp. 643–7.

Sebastian R.J. & Ryan E.B. 1985. 'Speech cues and social evaluation: markers of ethnicity, social class, and age' in H. Giles & R.N. StClair (eds), *Recent Advances in Language, Communication and Social Psychology*. London: Erlbaum. Pp. 112–43.

Sebastian R.J., Ryan E.B., Keogh T.F. & Schmidt A.C. 1980. 'The effects of negative affect arousal on reactions to speakers' in H. Giles, W.P. Robinson & P.M. Smith (eds), *Language: Social Psychological Perspectives*. Oxford: Pergamon. Pp. 203–8.

Seggie I. 1983. 'Attribution of guilt as a function of ethnic accent and type of crime.' *Journal of Multilingual and Multicultural Development* 4,2/3:197–206.

Seligman C.R., Tucker G.R. & Lambert W.E. 1972. 'The effects of speech style and other attributes on teachers' attitudes towards pupils.' *Language in Society* 1:131–42.

Shaw M.E. & Wright J.M. 1967. *Scales for the Measurement of Attitudes*. New York: McGraw-Hill.

Sherif M. & Hovland C.I. 1961. *Social Judgement: Assimilation and Contrast Effects in Communication and Attitude Change*. New Haven, Connecticut: Yale University Press.

Sherif C.W. & Sherif M. (eds). 1967. *Attitude, Ego-Involvement, and Change*. New York: Wiley.

Sherif M. & Sherif C.W. 1967. 'Attitude as the individual's own categories: the social judgment-involvement approach to attitude and attitude change' in C.W. Sherif & M. Sherif (eds), *Attitude, Ego-Involvement, and Change*. New York: Wiley. Pp. 105–39.

Shosteck H. 1973. 'Factors influencing appeal of TV news personalities.' *Journal of Broadcasting* 18,1:63–71.

Shuy R.W., Wolfram W.A. & Riley W.K. 1968. *Field Techniques in an Urban Language Study*. Washington, D.C.: Center for Applied Linguistics.

Siegel S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Siegel S. & Castellan N.J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill.

Simard L.M., Taylor D.M. & Giles H. 1976. 'Attribution processes and interpersonal accommodation in a bilingual setting.' *Language and Speech* 19,4: 374–87.

Smith B.L., Brown B.L., Strong W.J. & Rencher A.G. 1975. 'Effects of speech rate on personality perception.' *Language and Speech* 18:145–52.

Smith L.E. & Bisazza J.A. 1982. 'The comprehensibility of three varieties of English for college students in seven countries.' *Language Learning* 32: 259–69.

Smith P.M. 1980. 'Judging masculine and feminine social identities from content-controlled speech' in H. Giles, W.P. Robinson & P.M. Smith (eds), *Language: Social Psychological Perspectives*. Oxford: Pergamon. Pp. 121–6.

Smith P.M. 1985. *Language, the Sexes and Society*. Oxford: Blackwell.

Smith P.M., Giles H. & Hewstone M. 1980. 'Sociolinguistics: a social psychological perspective' in R.N. St Clair & H. Giles (eds), *The Social and Psychological Contexts of Language*. Hillsdale, N.J.: Erlbaum.

Stagner R. 1936. 'Judgements of voice and personality.' *Journal of Educational Psychology* 27:272–77.

Starks D. & Paltridge B. 1996. 'A note on using sociolinguistic methods to study non-native attitudes towards English.' *World Englishes* 15,2:217–24.

Starkweather J.A. 1956a. 'Content-free speech as a source of information about the speaker.' *Journal of Abnormal and Social Psychology* 52:394–402.

Starkweather J.A. 1956b. 'The communication value of content-free speech.' *American Journal of Psychology* 69:121–3.

Starkweather J.A. 1961. 'Vocal communication of personality and human feelings.' *Journal of Communication* 11:63–72.

Starkweather J.A. 1964. 'Variations in vocal behaviour.' *Disorders of Communication* (Research Publications, Vol. XLII, Association for Research in Nervous and Mental Disorders, USA). Pp. 424–49.

Stevick E.W. 1978. 'Toward a practical philosophy of pronunciation: another view.' *TESOL Quarterly* 12,2:145–50.

Stewart M.A. & Ryan E.B. 1982. 'Attitudes toward younger and older adult speakers: effects of varying speech rates.' *Journal of Language and Social Psychology* 1,2:91–110.

Stone V.A. 1973. 'Attitudes towards television newswomen.' *Journal of Broadcasting* 18,1:49–62.

Street R.L. 1984. 'Speech convergence and speech evaluation in fact-finding interviews.' *Human Communication Research* 11:139–69.

Street R.L. 1985. 'Participant–observer differences in speech evaluation.' *Journal of Language and Social Psychology* 4,2:125–30.

Street R.L., Brady R.M. & Lee R. 1984. 'Evaluative responses to communicators: the effects of speech rate, sex and interaction content.' *The Western Journal of Speech Communication* 48:14–27.

Street R.L., Brady R.M. & Putnam W.B. 1983. 'The influence of speech rate stereotypes and rate similarity or (*sic*, but should be 'on') listeners' evaluations of speakers.' *Journal of Language and Social Psychology* 2,1:37–56.

Street R.L. & Giles H. 1982. 'Speech accommodation theory: a social cognitive approach to language and speech behaviour' in M.E. Roloff & C.R. Berger (eds), *Social Cognition and Communication*. Beverly Hills: Sage. Pp. 193–226.

Streeter L.A., Krauss R.M., Geller V., Olson C. & Apple W. 1977. 'Pitch changes during attempted deception.' *Journal of Personality and Social Psychology* 35,5:345–50.

Strongman K.T. & Woosley J. 1967. 'Stereotyped reactions to regional accents.' *British Journal of Social and Clinical Psychology* 6:164–67.

Summers G.F. (ed.). 1977. *Attitude Measurement*. London: Kershaw.

Suomi K. 1976. 'English voiceless and voiced stops as produced by native and Finnish speakers.' *Jyväskylä Studies 2*. Jyväskylä: University of Jyväskylä.

Suomi K. 1979. 'The English / ptk / – / bdg / distinction: data and discussion' in J. Lehtonen & K. Sajavaara (eds), *Papers in Contrastive Phonetics (Jyväskylä Cross-Language Studies No. 7)*. Jyväskylä. Pp. 53–91.

Suomi K. 1980. 'Voicing in English and Finnish stops: a typological comparison with an interlanguage study of the two languages in contact.' *Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja, nro 10*. Turku.

Tajfel H. 1962. 'Social perception' in G. Humphrey & M. Argyle (eds), *Social Psychology through Experiment*. London: Methuen. Pp. 20–54.

Tajfel H. 1969. 'Cognitive aspects of prejudice.' *Journal of Social Issues* 25,4: 79–97.

Tajfel H. 1972. 'Experiments in a vacuum' in J. Israel & H. Tajfel (eds), *The Context of Social Psychology: a Critical Assessment*. London: Academic Press. Pp. 69–119.

Tarone E. 1977. 'Conscious communication strategies in interlanguage: a progress report' in H.D. Brown, C.A. Yorio & R.H. Crymes (eds), *On TESOL '77*. Washington, D.C.: TESOL. Pp. 194–203.

Taylor J.B. 1961. 'What do attitude scales measure? The problem of social desirability.' *Journal of Abnormal and Social Psychology* 62,2:386–90.

Taylor O.L. 1973. 'Teachers' attitudes toward Black and non-standard English as measured by the language attitude scale' in R.W. Shuy & R.W. Fasold (eds), *Language Attitudes: Current Trends and Prospects*. Washington: Georgetown University Press. Pp. 173–201.

Thakerar J.N. & Giles H. 1981. 'They are - so they speak: non-content speech stereotypes.' *Language and Communication* 1,2/3:255–61.

Thakerar J.N., Giles H. & Cheshire J. 1982. 'Psychological and linguistic parameters of speech accommodation theory' in C. Fraser & K.R. Scherer (eds), *Advances in the Social Psychology of Language*. Cambridge: Cambridge University Press. Pp. 205–55.

Thurstone L.L. 1928. 'Attitudes can be measured.' *American Journal of Sociology* 333:529–54. [Reprinted in G.F. Summers 1977. Pp. 127–41]

Thurstone L.L. 1931. 'The measurement of social attitudes.' *Journal of Abnormal and Social Psychology* 26:249–69. [Reprinted in M. Fishbein (ed.) 1967. Pp. 14–25.]

Thurstone L.L. & Chave E.J. 1929. *The Measurement of Attitude*. Chicago: University of Chicago Press.

Tommola J. 1975. 'On the relationship between discrimination and production of English sounds by Finnish learners.' *Publications in English Studies, No. 5.* Turku: University of Turku.

Triandis H.C. 1971. *Attitude and Attitude Change*. New York: Wiley.

Trudgill P. 1972. 'Sex, covert prestige, and linguistic change in the urban British English of Norwich.' *Language in Society* 1:179–96.

Trudgill P. 1974a. *Sociolinguistics: an Introduction*. Harmondsworth: Penguin.

Trudgill P. 1974b. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Tucker G.R. & Lambert W.E. 1969. 'White and Negro listeners' reactions to various American-English dialects.' *Social Forces* 47,4:463–8. (University of North Carolina Press.)

Turner J. 1993. 'Using Likert scales in L2 research . . . Another researcher comments.' *TESOL Quarterly* 24,4:736–9.

Upshaw H.S. 1962. 'Own attitude as an anchor in equal-appearing intervals.' *Journal of Abnormal and Social Psychology* 64,2:85–96.

Valo M. 1994. *Käsitykset ja vaikutelmat äänestä: kuuntelijoiden arviointia radiopuheen äänellisistä ominaisuuksista*. Jyväskylä: Jyväskylän yliopisto.

Varonis E.M. & Gass S. 1982. 'The comprehensibility of nonnative speech.' *Studies in Second Acquisition* 4:114–36.

Waller P.J. 1987. 'Democracy and dialect, speech and class' in P.J. Waller (ed.), *Politics and Social Change in Modern Britain: Essays Presented to A.F. Thompson.* Brighton: Harvester Press. Pp. 1–33.

Wardaugh R. 1998. *An Introduction to Sociolinguistics*. 3rd ed. Malden, Massachusetts: Blackwell.

Warren N. & Jahoda M. (eds). 1973. *Attitudes: Selected Readings.* 2nd ed. Harmondsworth: Penguin.

Webb S.C. 1955. 'Scaling of attitudes by the method of equal-appearing intervals: a review.' *Journal of Social Psychology* 42:215–39.

Webster W.G. & Kramer E. 1968. 'Attitudes and evaluational reactions to accented English speech.' *Journal of Social Psychology* 75:231–40.

Weisberg H.F. & Bowen B.D. 1977. *An Introduction to Survey Research and Data Analysis*. San Francisco: Freeman.

Welcowitz J., Ewen R.B. & Cohen J. 1971. *Introductory Statistics for the Behavioral Sciences*. New York: Academic Press.

Wells J.C. 1982. *Accents of English*. 3 vols. Cambridge: Cambridge University Press.

Wells J.C. 1995. 'Age grading in English pronunciation preferences.' *Proceedings of the International Conference of Phonetic Sciences 1995*, Stockholm. Vol. 3:696–9.

Welsh I. 1996. *Trainspotting*. London: Minerva.

Wierzbicka A. 1986. 'Human emotions: universal or culture-specific?' *American Anthropologist* 88,3:584–94.

Wiik K. 1965. *Finnish and English vowels. A comparison with special reference to the learning problems met by native speakers of Finnish learning English.* Turku.

Williams F. 1970. 'Psychological correlates of speech characteristics: on sounding "disadvantaged" .' *Journal of Speech and Hearing Research* 13,3:472–88.

Williams F. 1974. 'The identification of linguistic attitudes.' *International Journal of the Sociology of Language* 3:21–32.

Williams F., Whitehead J.L. & Miller L.M. 1971. 'Ethnic stereotyping and judgments of children's speech.' *Speech Monographs* 30:166–70.

Wingstedt M. & Schulman R. 1984. 'Comprehension of foreign accents' in W.U. Dressler, H.C. Luschutzky, O.E. Pfeiffer & J.R. Ressnison (eds), *Phonologica 1984*. Cambridge: Cambridge University Press. Pp. 339–45.

Wolfram W.A. 1969. *A Sociolinguistic Description of Detroit Negro Speech.* Washington, D.C.: Center for Applied Linguistics.

Wyer R.S. & Srull T.K. 1989. *Memory and Cognition in its Social Context.* Hillsdale: Erlbaum.

Wölck W. 1973. 'Attitudes toward Spanish and Quechua in bilingual Peru' in R.W. Shuy & R.W. Fasold (eds), *Language Attitudes: Current Trends and Prospects*. Washington: Georgetown University Press. Pp. 129–47.

Youngman M.B. 1984. 'Designing questionnaires' in J. Bell, T. Bush, A. Fox, J. Goody & S. Goulding, *Conducting Small-scale Investigations in Educational Management*. London: Harper & Row/Paul Chapman.

Zajonc R.B. 1968. 'Attitudinal effects of mere exposure.' *Journal of Personality and Social Psychology - Monograph Supplement* 9,2–2:1–27.