Exploring the forensic audiofit

Non-linguist perceptions, conceptions, descriptions and evaluations of unfamiliar voices in a forensic context

Mark Griffiths

Centre for Language and Communication Research
Cardiff University

Thesis submitted in part
fulfilment of the degree of
DOCTOR OF PHILOSOPHY
April 2009

This research was fully funded by:

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Award no. PTA-030-2002-01485

UMI Number: U585256

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



UMI U585256
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Research in the field of forensic linguistics has in recent years advanced our understanding of the processes and practice of speaker identification, yet the area of forensic speaker *description* remains unexplored. Drawing on research and methodologies from forensic linguistics, language attitudes and social psychology, this study investigates how non-linguists perceive and conceive of unknown speakers' accents and voices, and how language ideologies and social stereotypes interact with these perceptions. The study presents two groups of non-linguist respondents with recordings of five speakers responding to interview questions in pseudo-legal and non-legal control contexts. The analysis considers three major aspects of the non-linguists' responses: firstly, descriptions of speaker accent, accuracy of respondent estimations and possible mediating factors; secondly, non-linguists' voice descriptions from the first cohort of respondents, and their presentation to the second cohort in the form of an audiofit framework; thirdly, the study compares speaker evaluations of voices against the findings of previous studies, uniting research on general speaker evaluations of accents with studies on the attributions of guilt. Interacting mechanisms that mediate these evaluations are discussed.

Through quantitative statistical analysis, the study finds that non-linguists from the south of the UK perform well at locating accents geographically and socially when provided with information through the vocal channel only, and that accuracy of estimations is further enhanced by providing an assisting framework. For description of voice characteristics, the study finds that many of the features that are salient to and articulated by non-linguists form a useful basis for the an audiofit descriptor of voice characteristics, and are enhanced by respondents' consideration of possible *soundalikes* as a shorthand for conveying the salient characteristics. Finally, the study finds that the relationship between accent and attributions of guilt is more indirect and complex than previous literature may have indicated, noting that the common favouring or disfavouring of accents may be bound with unconscious evaluations of intelligence and aggression, and also with the voice characteristics of the individual being evaluated.

Acknowledgements

I would first of all like to thank the five people who have supervised this work and from whom I have learnt so much, Dr Peter Garrett, Dr Frances Rock, Dr Janet Cotterill, Professor Alison Wray and Dr Rebecca Playle. I thank them for their inspirational guidance and support, and for their invaluable patience. I cannot conceive of having completed this work without them.

I would also like to extend my thanks to the Centre for Language and Communication Research at Cardiff University to all the academic staff who have offered their informal guidance along the way, and especially to the CLCR Administrators, Julia Bullough, Carol Rees and Dawn Harrington for their assistance and support, and for retaining their sense of humour at all times.

I should also like to acknowledge and thank the ESRC for believing that the aims and outcomes of this research were deserving of institutional support and taxpayers' money.

My gratitude is also extended to the two police officers, anonymised here, from South Wales Police (Roath Station) for their time and insight into the handling of voice description evidence in the British context.

An enormous thank you is also extended to the mother of all proof readers.

And finally, to Anthony; thank you for your unflinching support and patience. I simply could not have done this without you there.

For Mum, Dad and Anthony

Look! A book!

TABLE OF CONTENTS

Abstract ii
Acknowledgements iii

1 - THE CONTEXT OF FORENSIC LINGUISTICS 1

1.1 Introduction 1
1.2.1 The rise of forensic linguistics and obstacles faced 3
1.2.2 Technology, language and witnesses 6
1.2.3 The police and forensic speaker description evidence 8
1.2.4 Linguists 11
1.2.5 Non-linguists 12
1.3 Voice evidence, practitioners and practice 14
1.3.1 Voice evidence 14
1.3.2 Brief historical background to voice evidence 14
1.3.3 Inter and intra-speaker variation 17
1.3.4 Current practice 18
1.3.5 Voice evidence: major issues and next steps 20

2 - NON-LINGUIST PERCEPTIONS, CONCEPTIONS AND EVALUATIONS 25

2.1 Introduction 25
2.2 The naïve non-linguist 28
2.3 Voice evidence 32
2.3.1 Voice evidence: accents 32
2.3.2 Voice evidence: physical speech features 46
2.4 Attitudes 54
2.4.1 Why do attitudes form? 54
2.4.2 What is an attitude? 55
2.4.3 Some difficulties with attitudes 58
2.5 Attitudes and language 60
2.5.1 Language attitudes in the public domain 61
2.5.2 'Good English', 'Correct English' and RP 65
2.6 Planning research on forensic speaker description: methodological choices 67
2.7 System variables: The selection of language attitudes objects 69
2.8 System variables: Indirect and direct measures 74
2.8.1 Indirect measures 74
2.8.2 Direct measures: benefits in relation to MGT 77
2.8.3 Direct measures: methodological considerations - biases 78
2.9 System variables: The questionnaire 83
2.10 System variables: Voice samples 87
2.10.1 Voice samples: matched-guise 88
2.10.2 Voice samples: accents 89
2.10.3 Voice samples: male versus female voice donors 90
2.10.4 Voice samples: ethnicity 91
2.10.5 Voice samples: the use of actors 91
2.10.6 Recording voice samples 91
2.10.7 Selection of voice samples 94
2.10.8 Voice samples: texts and the Picture Memory Interview 96
2.10.9 Voice samples: length of recording 98
2.11 System variables: The task 99
2.11.1 Outlet for respondents' attitudes – maps 100
2.11.1.1 Limitations of perceptual mapping 103
2.11.2 Guided perceptual mapping 105
2.11.3 Playing the samples 106

2.12 Estimator variables 106
2.12.1 Estimator variables: salience and speech sensitivity 106
2.12.2 Estimator variables: respondent memory 107
2.12.3 Estimator variables: respondent sex 108
2.12.4 Estimator variables: respondent age and age estimations 108
2.12.5 Estimator variables: respondent provenance 109
2.12.6 Estimator variables: in and outgrouping 113
2.12.7 Estimator variables: the role of social stereotypes 116
2.12.7.1 Conscious and unconscious processes 117
2.12.7.2 Acquisition of stereotypes as part of socialisation 118
2.12.7.3 Accessing social stereotypes 119
2.13 Estimator variables: language evaluations and the attribution of guilt 120
2.14 Research questions 123
2.14.1 The known and unknown 123
2.14.2 Research questions 125

3 - METHOD 126

3.1 Introduction 126
3.2 Participants: Stages 1 and 2 126
3.2.1 Respondents' linguistic background: Stages 1 and 2 127
3.2.2 Respondents' ages: Stages 1 and 2 127
3.2.3 Respondents' sex: Stages 1 and 2 128
3.2.4 Respondents' educational background: Stages 1 and 2 129
3.2.5 Respondents' regional background: Stages 1 and 2 129
3.2.5.1 Stage 1 129
3.2.5.2 Stage 2 138
3.3 Materials 141
3.3.1 Suggested context of recordings: Stage 1 141
3.3.2 Suggested contexts of recordings: Stage 2 142
3.3.3 Selection of voice samples: Stages 1 and 2 142
3.3.4 The selection of accents: Stages 1 and 2 144
3.3.5 Voice samples: Stages 1+2 - Picture Memory Interview 147
3.3.6 The questionnaires: Stage 1 153
3.3.6.1 The question sequence: Stage 1 156
3.3.7 The questionnaires: Stage 2 157
3.3.7.1 Perceptual maps 157
3.3.7.2 Stage 2, Crime context 158
3.3.7.3 Rating the speakers' accents 158
3.3.7.4 Describing and rating the speakers' voices 159
3.3.7.5 Question order 161
3.4 Procedure: Stages 1 and 2 161

4 - RESULTS: Stage 1 163

4 .1 Introduction 163
4.2 Where did respondents believe speakers are from? 164
4.3 Estimations of speaker origin: are they associated with respondent origin? 168
4.4 Which accents were most salient to each region? 174
4.5 Estimations of speaker age 179
4.6 Qualitative descriptions of the speakers' voices 182
4.7 Estimations of speaker intelligence 186
4.8 Estimations of speaker honesty 192
4.9 Estimations of speaker aggression 196
4.10 Attributions of guilt 199
4.10.1 Crime type 199
4.10.2 Question order 206

4.11 Principal component analysis and correlations **207**
4.11.1 Principal component analysis & correlations: Speaker 1 208
4.11.2 Principal component analysis & correlations: Speaker 2 210
4.11.3 Principal component analysis & correlations: Speaker 3 211
4.11.4 Principal component analysis & correlations: Speaker 4 213
4.11.5 Principal component analysis & correlations: Speaker 5 214
4.12 Summary **215**

5 - RESULTS: Stage 2 ..... **216**

5.1 Introduction **216**
5.2 Where did respondents believe speakers are from? **217**
5.3 Estimations of speaker origin: are they associated with respondent origin? **224**
5.4 Which accents were most salient to each region? **231**
5.5 Accent 'strength': regional/no accent/well-spoken **233**
5.6 Voice features 1 – Bipolar semantic scales **238**
5.6.1 Voice features: comparison of means features by feature 239
5.6.1.1 Speaker speed 239
5.6.1.2 Speaker confidence/hesitancy 240
5.6.1.3 Speaker pitch 241
5.6.1.4 Speaker monotone/melodious 241
5.6.1.5 Speaker loudness 242
5.6.1.6 Speaker steadiness 243
5.6.2 Voice features: comparison speaker by speaker 244
5.6.2.1 Speaker 1 – Wales 245
5.6.2.2 Speaker 2 – South West 248
5.6.2.3 Speaker 3 – South East 251
5.6.2.4 Speaker 4 – South East (RP-like) 254
5.6.2.5 Speaker 5 – Midlands 256
5.7 Voice features 2 – Non-polar features **260**
5.7.1 Small data sets 260
5.7.2 Analysis 261
5.8 Soundalikes **263**
5.8.1 Speaker 1 – Wales 264
5.8.2 Speaker 2 – South West 265
5.8.3 Speaker 3 – South East 266
5.8.4 Speaker 4 – South East (RP-like) 268
5.8.5 Speaker 5 – Midlands 269
5.9 Attributions of Guilt **270**
5.10 'Got the Girl' **274**
5.11 Summary **277**

6 - DISCUSSION **279**

6.1 Introduction **279**
6.2 Estimations of speaker age **280**
6.3 Estimations of speaker origin (Stages 1 and 2) **282**
6.3.1 Estimations of the origins of traditional regional accents 282
6.3.2 Estimations of the origin of the RP-like accent 293
6.3.3 Estimations of the origin of Speaker 5 297
6.4 Audiofit voice descriptions **301**
6.4.1 Bipolar semantic differential scales 301
6.4.2 Correlations and factors for bi-polar features 307
6.4.3 Non-polar features 309
6.4.4 Soundalikes 311
6.4.5 Forensic voice description: 3 descriptive channels 314

6.5 The role of language attitudes 321
6.5.1 Intelligence 321
6.5.2 Honesty 324
6.5.3 Aggression 326
6.5.4 Attributions of speaker guilt 328
6.5.5 'Got the girl' 337
6.6 Summary 338

7 - CONCLUSIONS 340

7.1 Introduction 340
7.2 Methodology 340
7.3 Non-linguist perceptions and conceptions 341
7.4 Non linguist descriptions of unknown speakers 345
7.5 Non-linguist evaluations of unknown speakers 350
7.6 Limitations and other future work 354
7.7 A prototype audiofit 357
7.8 Closing remarks 363

REFERENCES 364

APPENDICES 374

LIST OF TABLES AND FIGURES

Table 1.1	Comparison of visual and linguistic evidence sources.	21
Figure 2.4.2.1	Tripartite structure of attitudes	56
Figure 2.4.2.2	Illustration of tripartite structure of attitudes: attitude to French	57
Figure 2.5	Examples of language attitudes in the public domain	62
Figure 2.9.1	Comparison of Zahn & Hopper's and Perloff's dimensions	86
Figure 2.9.2	Comparison of dimensions	86
Figure 3.2.2.1	Respondents ages, Stage 1	128
Figure 3.2.2.2	Respondents ages, Stage 2	128
Figure 3.2.3.1	Respondent sex, Stage 1	129
Figure 3.2.3.2	Respondent sex, Stage 2	129
Figure 3.2.5.1	BBC Map of British Regions	132
Figure 3.2.5.2	Map of ITV Broadcasting Regions	132
Figure 3.2.5.3	Perceptual map of Great Britain	133
Figure 3.2.5.4	Regional division of Cardiff University catchment area	135
Table 3.2.5.1	Respondents' identified places of origin	136
Figure 3.2.5.5	Stage 1 - Respondents by region	138
Figure 3.2.5.6	Stage 2, idealised map of UK regions and cities	139
Figure 3.2.5.7	Stage 2 Geographical respondent/speaker locator	140
Figure 3.2.5.8	Stage 2 - Respondents by region	141
Figure 3.3.4.1	Geographical provenance of voice donors	148
Figure 4.2.1	Areal representation of accurate identifications of Speaker 1's provenance	165
Figure 4.2.2	Areal representation of accurate identifications of Speaker 2's provenance	165
Figure 4.2.3	Areal representation of accurate identifications of Speaker 3's provenance	166
Figure 4.2.4	Areal representation of 'accurate' identifications of Speaker 4 provenance	166
Figure 4.2.5	Areal representation of accurate identifications of Speaker 5 provenance	167
Table 4.3.1	Speaker 1, identified as from 'Wales' by respondents from:	170
Table 4.3.2	Speaker 2, identified as 'South West' by respondents from:	170
Table 4.3.3	Speaker 3, identified as 'South East' by respondents from:	170
Table 4.3.4	Speaker 3, identified as 'South East/London' by respondents from:	171
Table 4.3.5	Speaker 4, identified as 'South/South East/London' by respondents from:	171
Table 4.3.6	Speaker 4, identified as 'South East/London' by respondents from:	171
Table 4.3.7	Speaker 5, identified as 'Midlands' by respondents from:	171
Table 4.3.8	Speaker 5, identified as 'South' by respondents from:	172
Table 4.4.1	Salience and availability rates for Welsh respondents	175
Figure 4.4.1	Map of salience rates for Welsh respondents	175

Table 4.4.2	Salience rates for West/South West respondents	176
Figure 4.4.2	Map of salience rates for West/South West respondents	176
Table 4.4.3	Salience rates for South East respondents	177
Figure 4.4.3	Map of salience rates for South East respondents	177
Table 4.4.4	Salience rates for Midlands respondents	178
Figure 4.4.4	Map of salience rates for Midlands respondents	178
Figure 4.5.1	Estimations of Speaker 1's age	179
Figure 4.5.2	Estimations of Speaker 2's age	180
Figure 4.5.3	Estimations of Speaker 3's age	180
Figure 4.5.4	Estimations of Speaker 4's age - graphic representation	180
Figure 4.5.5	Estimations of Speaker 5's age - graphic representation	181
Figure 4.6.1	Professional linguist vs. non-linguist comments on voice qualities	184
Figure 4.7.1	Intelligence ratings for Speaker 1 - Wales	188
Figure 4.7.2	Intelligence ratings for Speaker 2 - South West	188
Figure 4.7.3	Intelligence ratings for Speaker 3 - South East	189
Figure 4.7.4	Intelligence ratings for Speaker 4 - South East (RP-like)	189
Figure 4.7.5	Intelligence ratings for Speaker 5 - Midlands	189
Figure 4.7.6	Speaker intelligence: table of means for each speaker	190
Figure 4.7.7	Repeated measures ANOVA for Speaker Intelligence ($\rho < 0.05$)	190
Figure 4.8.1	Honesty ratings for Speaker 1 - Wales	192
Figure 4.8.2	Honesty ratings for Speaker 2 - South West	192
Figure 4.8.3	Honesty ratings for Speaker 3 - South East	193
Figure 4.8.4	Honesty ratings for Speaker 4 - South East, 'RP-like'	193
Figure 4.8.5	Honesty ratings for Speaker 5 - Midlands 'no accent'	193
Figure 4.8.6	Speaker honesty: comparison of means for each speaker	194
Figure 4.8.7	Repeated measures ANOVA for speaker honesty ($\rho < 0.05$)	194
Figure 4.9.1	Aggression ratings for Speaker 1 - Wales	196
Figure 4.9.2	Aggression ratings for Speaker 2 - South West	196
Figure 4.9.3	Aggression ratings for Speaker 3 - South East	197
Figure 4.9.4	Aggression ratings for Speaker 4 - South East, RP-like	197
Figure 4.9.5	Aggression ratings for Speaker 5 - Midlands, 'no accent'	197
Figure 4.9.6	Speaker aggression: comparison of means for each speaker	198
Figure 4.9.7	Repeated measures ANOVA for speaker aggression ($\rho < 0.05$)	198
Figure 4.10.1.1	Attributions of speaker guilt: means for sexual assault	200
Figure 4.10.1.2	Repeated measures ANOVA for guilt - sexual assault ($\rho < 0.05$)	200
Figure 4.10.1.3	Attributions of speaker guilt: means for assault and robbery	201

Figure 4.10.1.4	Repeated measures ANOVA for guilt - assault and robbery ($\rho < 0.05$)	201
Figure 4.10.1.5	Attributions of guilt: sexual assault versus assault and robbery	202
Figure 4.10.1.6	Combined guilt ratings for Speaker 1 - Wales	203
Figure 4.10.1.7	Combined guilt ratings for Speaker 2 - South West	203
Figure 4.10.1.8	Combined guilt ratings for Speaker 3 - South East	204
Figure 4.10.1.9	Combined guilt ratings for Speaker 4 – Eastbourne - RP-like	204
Figure 4.10.1.10	Combined guilt ratings for Speaker 5 - Birmingham	204
Figure 4.10.1.11	Speaker guilt: comparison of means for each speaker in combined crimes	205
Figure 4.10.1.12	Repeated measures ANOVA for speaker guilt ($\rho < 0.05$)	205
Figure 4.10.2.1	Attributions of guilt: crime first versus region first	206
Figure 4.10.2.2	Attributions of guilt - crime or region 1st: Tests of within-subjects effects	207
Table 4.11.1.1	Principal component analysis for Speaker 1	208
Table 4.11.1.2	Component matrix for Speaker 1	208
Figure 4.11.1.1	Pearson r correlation coefficient (2 tailed) for Speaker 1	209
Table 4.11.2.1	Principal component analysis for Speaker 2	210
Table 4.11.2.2	Component matrix for Speaker 2	210
Figure 4.11.2.1	Pearson r correlation coefficient (2 tailed) for Speaker 2	211
Table 4.11.3.1	Principal component analysis for Speaker 3	211
Table 4.11.3.2	Component matrix for Speaker 3	212
Figure 4.11.3.1	Pearson r correlation coefficient (2 tailed) for Speaker 3	212
Table 4.11.4.1	Principal component analysis for Speaker 4	213
Table 4.11.4.2	Component matrix analysis for Speaker 4	213
Figure 4.11.4.1	Pearson r correlation coefficient (2 tailed) for Speaker 4	213
Table 4.11.5.1	Principal component analysis for Speaker 5	214
Table 4.11.5.2	Component matrix for Speaker 5	214
Figure 4.11.5.1	Pearson r correlation coefficient (2 tailed) for Speaker 5	215
Figure 5.2.1	Areal representation of respondents' identifications of Speaker 1	218
Figure 5.2.2	Areal representation of respondents' identifications of Speaker 2	219
Figure 5.2.3	Areal representation of respondents' identifications of Speaker 3	219
Figure 5.2.4	Areal representation of respondents' identifications of Speaker 4	220
Figure 5.2.5	Areal representation of respondents' identifications of Speaker 5	220
Figure 5.2.6	Speaker 3 - % Identifications of London/South East, Study 1 vs. Study 2	222
Figure 5.2.7	Speaker 4 - % Identifications for Oxbridge, Study 1 vs. Study2	223
Table 5.3.1	Speaker 1, identified as from 'Wales' by respondents from:	226
Table 5.3.2	Speaker 1, identified as from 'South Wales' by respondents from:	227
Table 5.3.3	Speaker 1, identified as from 'Cardiff' by respondents from:	227
Table 5.3.4	Speaker 2, identified as 'South West' by respondents from:	227

Table 5.3.5	Speaker 2, identified as from 'Bristol' by respondents from:	227
Table 5.3.6	Speaker 3, identified as from 'South East/London by respondents from:	228
Table 5.3.7	Speaker 3, identified as from 'London' by respondents from:	228
Table 5.3.8	Speaker 4, identified as from 'South East/London' by respondents from:	228
Table 5.3.9	Speaker 4, identified as from 'South/S.East/London by respondents from:	228
Table 5.3.10	Speaker 4, identified as from 'Oxbridge' by respondents from:	229
Table 5.3.11	Speaker 5, identified as from 'Midlands' by respondents from:	229
Table 5.3.12	Speaker 5, identified as from 'South/South East' by respondents from:	229
Table 5.3.13	Speaker 5, identified as from 'Oxbridge' by respondents from:	229
Table 5.4.1	Salience rates for Welsh respondents	231
Table 5.4.2	Salience rates for South West England respondents	232
Table 5.4.3	Salience rates for South East England respondents	232
Table 5.4.4	Salience rates for Midlands respondents	232
Figure 5.5.1	Attributions for Speaker 1 'accent strength'	234
Figure 5.5.2	Attributions for Speaker 2 'accent strength'	235
Figure 5.5.3	Attributions for Speaker 3 'accent strength'	235
Figure 5.5.4	Attributions for Speaker 4 'accent strength'	235
Figure 5.5.5	Attributions for Speaker 5 'accent strength'	236
Figure 5.5.6	Table of means for 'accent strength'	236
Table 5.6.1	Repeated measures ANOVA for 'accent strength' ($\rho < 0.05$)	236
Figure 5.6.1.1	Mean scores for 'speaker speed'	239
Table 5.6.1.1	Repeated measures ANOVA for 'speaker speed' ($\rho < 0.05$)	240
Figure 5.6.1.2	Mean scores for 'speaker confidence/hesitancy'	240
Table 5.6.1.2	Repeated measures ANOVA for 'speaker confidence/hesitancy' ($\rho < 0.05$)	240
Figure 5.6.1.3	Mean scores for 'speaker pitch'	241
Table 5.6.1.3	Repeated measures ANOVA for 'speaker pitch' ($\rho < 0.05$)	241
Figure 5.6.1.4	Mean scores for 'speaker monotone/melodious'	241
Table 5.6.1.4	Repeated measures ANOVA for 'speaker monotone/melodious ($\rho < 0.05$)	242
Figure 5.6.1.5	Mean scores for 'speaker loudness'	242
Table 5.6.1.5	Repeated measures ANOVA for 'speaker loudness' ($\rho < 0.05$)	242
Figure 5.6.1.6	Mean scores for 'speaker steadiness'	243
Table 5.6.1.6	Repeated measures ANOVA for 'speaker steadiness'	243
Table 5.6.1.7	Ranked order for 'speaker speed'	244
Table 5.6.2.1.1	Means and medians for Speaker 1 voice characteristics	245
Figure 5.6.2.1.1	Boxplots for Speaker 1 voice characteristics	246
Table 5.6.2.1.2	Principal component analysis for Speaker 1 voice characteristics	247

Table 5.6.2.1.3	Rotated Component Matrix for Speaker 1 voice characteristics	248
Table 5.6.2.2.1	Means and medians for Speaker 2 voice characteristics	248
Figure 5.6.2.2.1	Boxplots for Speaker 2 voice characteristics	249
Table 5.6.2.2.2	Principal component analysis for Speaker 2 voice characteristics	250
Table 5.6.2.2.3	Rotated Component Matrix for Speaker 2 voice characteristics	250
Table 5.6.2.3.1	Means and medians for Speaker 3 voice characteristics	251
Figure 5.6.2.3.1	Boxplots for Speaker 3 voice characteristics	252
Table 5.6.2.3.2	Principal component analysis for Speaker 3 voice characteristics	253
Table 5.6.3.3.3	Rotated Component Matrix for Speaker 3 voice characteristics	253
Table 5.6.2.4.1	Means and medians for Speaker 4 voice characteristics	254
Figure 5.6.2.4.1	Boxplots for Speaker 4 voice characteristics	254
Table 5.6.2.4.2	Principal component analysis for Speaker 4 voice characteristics	256
Table 5.6.2.4.3	Rotated Component Matrix for Speaker 4 voice characteristics	256
Table 5.6.2.5.1	Means and medians for Speaker 5 voice characteristics	257
Figure 5.6.2.5.1	Boxplots for Speaker 5 voice characteristics	257
Table 5.6.2.5.2	Principal component analysis for Speaker 5 voice characteristics	259
Table 5.6.2.5.3	Rotated Component Matrix for Speaker 5 voice characteristics	259
Table 5.7.2	Non-polar voice features for all speakers	261
Table 5.8.1	Soundalikes for Speaker 1 – Wales	264
Table 5.8.2	Soundalikes for Speaker 2 – South West	265
Table 5.8.3	Soundalikes for Speaker 3 – South East	266
Table 5.8.4	Soundalikes for Speaker 4 – RP-like	268
Table 5.8.5	Soundalikes for Speaker 5 – Midlands	269
Figure 5.9.1.1	Distribution of attributions of guilt – Speaker 1	270
Figure 5.9.1.2	Distribution of attributions of guilt – Speaker 2	271
Figure 5.9.1.3	Distribution of attributions of guilt – Speaker 3	271
Figure 5.9.1.4	Distribution of attributions of guilt – Speaker 4	271
Figure 5.9.1.5	Distribution of attributions of guilt – Speaker 5	272
Table 5.9.1	Stage 2, attributions of guilt - means	272
Table 5.9.2	Stage 2, repeated measures ANOVA for speaker guilt	272
Table 5.9.10	Comparison of mean ratings for all three crime scenarios	273
Figure 5.10.1	Distribution of evaluations for 'Got the girl' – Speaker 1	275
Figure 5.10.2	Distribution of evaluations for 'Got the girl' – Speaker 2	275
Figure 5.10.3	Distribution of evaluations for 'Got the girl' – Speaker 3	275
Figure 5.10.4	Distribution of evaluations for 'Got the girl' – Speaker 4	276
Figure 5.10.5	Distribution of evaluations for 'Got the girl' – Speaker 5	276
Figure 5.10.6	Comparison of means for 'Got the Girl' – all speakers	276

Table 5.10.1 Repeated measures ANOVA for 'Got the girl' 277
Figure 6.2.3 Interpretation of respondents' evaluations of Speaker 5 'strength of accent'.... 300
Figure 6.5.4 Comparison of correlations: 'intelligence', 'honesty', 'aggression' 'guilt' 333
Figure 7.7 Diagrammatic representation of an audiofit prototype 362

1 - THE CONTEXT OF FORENSIC LINGUISTICS

1.1 Introduction

'A Portsmouth armed robber was convicted by the sound of his voice alone in the first case of its kind.' – Portsmouth News *(Bull and Clifford 1999:1)*

The late 20th century witnessed a great expansion in the nature and scope of what could serve as evidence in law courts around the world. The development of such technologies as CCTV cameras recording footage on streets, trains, buses, and DNA sampling of crime scenes and suspects allowed for a shift away from the historically predominant reliance on anecdotal evidence and corroboration to more sophisticated, objective and reliable means of gathering and interrogating evidence. Academic and technological advances in the field of forensic sciences have increasingly enabled researchers to turn once apparently meaningless artefacts into meaningful evidence, delivering highly complex data, and yielding compelling results. The progress made was and continues to be enormous, such that those legal professionals involved in criminal investigations and prosecutions at the beginning of the 20th could never have envisaged how, in technical terms, the century would end.

One of the advances that has come to be used both in every day cases and high profile events is the photofit, a technique whereby the descriptions of eyewitness with firsthand knowledge about a crime are used by the police to build up an accurate image of a wanted criminal. In the United Kingdom, it is technique with which most members of the public are familiar, be it through personal experience or having seen it as an ever present feature on the BBC's *Crimewatch* television program. It was even actively used in a case where one

of the Crimewatch presenters themselves became a victim of a stalker. After her murder, a photofit was broadcast on the programme in order to solicit evidence from eyewitnesses.

Turning to language, much progress has been made in the use of language as evidence, with analysis of written and spoken texts becoming more frequently employed by the police and law courts as part of the investigative procedure and prosecutions. The specific concern to the research in this thesis is the linguistic parallel to the photofit, which I will refer to as an *audiofit*.

This thesis addresses a question of major significance for forensic linguistic research, and for the wider legal community, namely, whether there is the potential for audiofits to be developed for use in police investigations. The proposal here is to investigate the potential for an *earwitness* with firsthand knowledge of a crime to provide a pen sketch description of the voice and/or accent of a wanted criminal. In doing so, this research proposes to investigate *non-linguist* perceptions of voices and accents, and also explore the descriptions provided by non-linguists. The word 'propose' is used in this context, for the simple reason that unlike the photofit, the area of audiofit has not yet, to the best of the author's knowledge, received any dedicated research attention, and rather than being a standard part of police practice around the world, it is currently only an unexplored theoretical possibility.

As will become evident in the course of this chapter and Chapter 2, the research gap for possible audiofits is so large that no single thesis can fully research the potential for audiofit designs all the way through to trialling audiofit reliability in police investigations. However, the present study aims to offer some modest observations and proposals that may contribute to the answer.

The remainder of this first chapter provides a broad brush overview of the context for the research. Chapter 2 develops the issues identified in Chapter 1 in relation to designing an audiofit, culminating in a set of research questions which form the basis of the rest of the present study. In Chapter 3, the design of the study is described, including appropriate justification with reference to the

literature discussed in Chapter 2, for the key methodological decisions taken. The analysis Chapters 4 and 5 present the results of two studies, in their respective order. The results of the studies are discussed in Chapter 6 within the context of the research questions. The account thus naturally leads to the final section of the thesis, the Conclusions (Chapter 7), where the review questions and issues arising from them are considered in their context, including problems with the design of the study and opportunities for future research, which are identified and discussed. The findings of this study are located in the broader agenda of developing a future of forensic audiofits. In the Conclusions chapter it is possible to address, in the light of all that has been covered, the main question as to the feasibility of an audiofit being developed for use in the police context.

1.2.1 The rise of forensic linguistics and obstacles faced

In recent years, interest in language in the forensic context has increased rapidly. Practitioners of *forensic linguistics*, the "scientific study of language as applied to forensic purposes and contexts" (McMenamin 2002:67), have become ever-more frequent visitors to the courts in the role of expert witnesses leading to a growth in the community of forensic linguists and the establishment of organisations such as the International Association of Forensic Linguists (IAFL), and the International Association for Forensic Phonetics and Acoustics (IAFPA), whose members are regularly consulted by the police and other legal professionals on matters of linguistic evidence. For example, Professor Malcolm Coulthard, founder of the IAFL, cites his involvement in over 180 civil and criminal cases over the last twenty years. (http://www.aston.ac.uk/lss/staff/coulthardm.jsp, accessed 12/3/07). And Professor Peter French, with Malcolm Coulthard a founding editor of *The International Journal of Speech, Language and the Law*, cites over 200 appearances in the witness box and around 3000 consultations as an expert witness (http://www.jpfrench.com/docs/frenchcv.pdf, accessed12/3/07).

Analytical advances in discourse analysis have informed the juries, judges and appeal courts (see for example, Svartvik (1968), Coulthard (1994)), preventing

and reversing miscarriages of justice, whilst technical advances have enabled forensic phoneticians to provide acoustic analyses as evidence (French and Harrison, in press).

One such high profile case for forensic linguistics, detailed by Coulthard (2000) is that of Derek Bentley, the 19 year old hanged in 1952 for his part in the murder of PC Sidney Miles; the fatal shot had been fired by Bentley's sixteen-year-old friend, Christopher Craig, when Bentley was already in police custody. Bentley had been arrested at the scene but was tried for murder after the police alleged that his instruction to his gun-wielding friend to 'let him have it', was an incitement for his friend to shoot dead the police officer. Coulthard's linguistic analysis demonstrated through a combination of features he had identified that the Bentley 'confession' was not simply a word-for-word transcript of Bentley's spoken statement as the police had claimed, but was in part co-authored, and specifically the written work of the police officers. On the basis of Coulthard's linguistic analysis, in 1998 Bentley's murder conviction was posthumously overturned.

Nevertheless, the appearance of headline-grabbing stories such as the Portsmouth armed robber at the beginning of this section and the augmenting recognition amongst legal practitioners of the evidential role that language can play, have not altered the fact that currently, many legal professionals are still unfamiliar with what forensic linguistic analysis may have to offer the courts, and remain uncertain as to the reliability and validity of many of the specific applications of forensic linguistics. To illustrate this point, we can contrast the acceptance of forensic linguistic analysis in the case above with a further case involving language as a central part of the evidence. Both cases employed the same expert witness, Professor Malcolm Coulthard, but in his description of the rejection of forensic linguist analysis in a failed appeal, Coulthard (1997) describes how British courts disregarded the merits of linguistic analysis and the opinions of professional linguists.

A defendant accused of four armed robberies was arrested and interviewed, during which time he was driven to the places where he claimed to have been

when the robberies were taking place. Whilst driving, contemporaneous notes were taken by a police officer writing on a clipboard on his lap. After the journey, the defendant read and signed the notes as an accurate record, but later claimed that incriminating additions had been made after he had signed them, specifically on alternate lines of the text that he had witnessed.

Coulthard's expert opinion on the record of the car journey was used in an Appeal against Conviction and highlighted inconsistencies in graphology, register, concord, anaphora and other grammatical features. The expert witness analysis argued that as much as half of the text presented to the court may have been added after the defendant had signed it.

The response from the Lord Justices included a dismissal of Professor Coulthard's evidence for pointing out that two grammars were present in the text - a note-taking and a full grammar - on the basis that there is 'only correct and incorrect grammar' (Coulthard 1997:300). Moreover, in Judgement, they did not feel that Coulthard's evidence was 'sufficiently cogent' (Coulthard 1997:301). Thus, the appeal failed, and an analysis that Coulthard felt compelling, failed to convince judges in a real-world context.

This limited comprehension and appreciation of linguistic argument is a phenomenon familiar to linguists when dealing with members of the public (see, for example, Bauer and Trudgill's (1998) 'Language Myths'). However, for linguists dealing with law, the situation is quite unpalatable. Rodman (2002) when discussing a case study of a Haitian-born American convicted on the basis of him being able to 'drop' his accent and speak a dialect of American Black English reports the linguistic naiveties that lead to the conviction and concludes:

> *If a person were convicted of a crime because of a geocentric solar system, or of the phlogiston theory of combustion, or of the four humours, there would be outrage, and the conviction would (one prays) be overturned. Yet the chain of linguistic non-sequiturs that allowed the Haitian to be convicted is no less absurd, no less naïve in today's society, than these once respected ancient theories.*
>
> *(Rodman 2002:103)*

The courts' dismissal of linguistic evidence, analysis and expert opinion may in some ways also be explained by the paucity of research in the area of language and the law. This is evident not only in the linguistic analysis of written texts presented as in Coulthard's examples, but also in research centred on oral texts. For example, in perhaps the nearest research area to that presented in this thesis, *forensic speaker identification*, Künzel (1994), Bull and Clifford (1999) and Yarmey (2001) have all cited problems caused, at least in part, by the relative paucity and controversy of the research evidence. Not only do we have uncertainty amongst the courts, law enforcement agencies and legal practitioners, but also amongst phoneticians and linguists, where there remains much debate surrounding what voice evidence involves, what constitutes appropriate methodology, what it can achieve and what its limitations are (Rose 2002:2).

Given the current state of forensic linguistic research and its standing in the courts, we can summarise the above issues into two questions: Firstly, are criminal justice systems around the world making full use of all available tools at their disposal? Secondly, can the academic community make a contribution to this area? The answer to the former question must be, on the basis of current trial examples, 'no', as in the rejected appeal case described by Coulthard (1997). In answer to the second question, it is the contention of this thesis that the answer is 'yes', and it is this belief that will be driving the research. On this basis, the thesis will identify and focus on a specific area of forensic linguistics where the academic community can move things forward, an apparently as yet researched area that I have entitled '*forensic speaker description*'. The following sections set out a context within which this research area has evolved.

1.2.2 Technology, language and witnesses

I mentioned in the Introduction that some of the most striking technological advances in evidential data are the ubiquitous CCTV and the use of DNA profiling. Moreover, where there are witnesses to the crime taking place, the modern technology that police forces have at their disposal can assist the *eyewitness* in providing a description of the suspect's face, by reconstructing

and fine-tuning computer-generated images to match what they saw. Such technological advances have doubtless had a major impact on crime detection; however, we must remember that it is not only the visual channel that witnesses employ when encoding and recalling the details of a crime. As Nolan and Grabe (1996) state:

> *Criminals, contrary to the ideal for children in Victorian England, are often heard but not seen. Victims may not have seen attackers, but may have heard them speak; witnesses may have seen little of masked armed robbers, but have overheard their interactions; and recipients of obscene telephone calls have been exposed only to the voice of the perpetrator.*
>
> *(Nolan and Grabe 1996:74)*

It is on the idea of the voice as evidence that this thesis will be focusing. Voice evidence has a unique role to play in criminal investigations and the technologies and techniques for handling recorded voice data have developed significantly in recent times. For example, real progress has been made in the field of forensic speaker identification in terms of the technology for performing acoustic analysis in the way that voices can be isolated from background noise and analysed and in the last decade since Coulthard's (1997) example mentioned above, the analysis of recorded speech evidence appears to have gained credibility in the eyes of the court. Recent high-profile examples of recorded speech as evidence in the UK include the alleged cheating during the television show, 'Who Wants to Be a Millionaire' (French and Harrison 2004) when an audience member was accused of having mouthed words and coughed to indicate correct answers, and the arrest and successful prosecution of the Yorkshire Ripper Hoaxer (French et al. 2006) - more details of which are given below in Section 1.3.2 (p.16).

However, as I described above, there remains insufficient research in the area of language evidence. This thesis therefore seeks to focus on language evidence, and specifically on forensic speaker description, an undeveloped research area which has yet to benefit from any substantial technological advances or increased prominence in the investigative process. The paucity of research and technological development in this area can be illustrated if one takes examples of the advances made in the analysis of recorded speech for speaker identification, the use of CCTV across the urban landscape, the

pinpoint accuracy of DNA and the interactive nature of the modern eyewitness videofit, and one compares these innovations with the manner in which the evidential descriptions of voices provided by *earwitnesses* is handled; there are few technological parallels. What this means is that no system, protocol or precedent has been established regarding how to deal with descriptions rather than recordings of voices. That is, in dealing with any police force around the world, an earwitness, having heard some example of a suspect's speech, does not have any technology at their disposal with which to recall, describe and replicate what they heard. Moreover, the academic community has made very few inroads into understanding what it is that non-linguists perceive in a voice or how they describe what they perceive.

It is unclear why this gap in our understanding of earwitness description has remained unaddressed. It may be that until recently, the relationship between non-linguists, language and crime has been seen as too complex to address; alternatively, it may be that compared with linguistic evidence, there have been other, more prominent, significant or reliable lines of enquiry to pursue; or perhaps more cynically, the gap in our understanding may be the result of neglect on the part of governments and the legal systems that are made up of non-linguists who consider themselves to be expert language users themselves with little appreciation of the power and utility of additional linguistic analysis, and therefore do not solicit such research from academics. Whatever the reason, it is the contention of this thesis that the field of forensic speaker description merits greater attention, and that developing our capacity for eliciting reliable earwitness descriptions will yield positive outcomes for the criminal justice system.

1.2.3 The police and forensic speaker description evidence

As a launch pad for this research, I set out to gain a better insight into how the police deal with issues of forensic speaker description by speaking directly to those at the front line of criminal investigation and arranging informal meetings with a Police Constable and Detective Constable with South Wales Police. The

meetings took place independently of each other and were short and informal in nature.

In the meetings I raised the following issues:
1/. How often do you feel that a description of a suspect's voice may be useful in their apprehension? Is it a routine line of enquiry to ask for a speaker description?

2/. How confident do you feel that members of the public can give you useful speaker descriptions?

3/. What types of questions do you ask regarding a speaker description?

The responses
1/. Both the PC and DC responded that they did not feel that a speaker's voice was a feature that they would frequently ask about. The PC felt that she would be more likely to concentrate on what a suspect said, i.e. the content rather than the mode of delivery.

The DC did add, however, that he felt that speaker description could be a potentially important piece of evidence when relevant, although he felt, that such evidence would never be the single most important piece, and that it would be followed up in conjunction with information elicited from witnesses and the SOCO's (Scene of Crime Officers). This echoes Kredens and Góralewska-Łach's (1998) description of the prevailing view amongst forensic phoneticians, that evidence based on speech analysis is largely corroborative. It would appear that both the police and forensic speech scientists are in broad agreement on this issue.

2/. The responses to the second question were entirely consistent with the first. Neither the PC nor the DC felt confident in the public's general ability to provide speaker descriptions. Both said they would ask 'if he had an accent' but the PC felt that there were so many different accents in Cardiff (referring to the considerable ethnic diversity of Cardiff where the interviews took place) that it

was hard to describe them. She did, however, distinguish between 'foreign' and 'English'.

This foreign/English distinction and the remark regarding whether the speaker 'had an accent' are interesting in themselves. Firstly, the interviews took place in Wales, with Welsh Officers working for South Wales Police, yet 'English' was seen as the opposite of foreign - perhaps a reference to the language rather than the nation - but this remark may nevertheless give an insight into South Walian linguistic ideology and what is seen as the linguistic standard, an issue that merits further research. Secondly, in terms of 'having an accent', it was unclear how somebody would sound if this was *not* the case. If they had no accent, would that mean that they share the community accent? That is, does 'no accent' mean that the speaker is community-neutral and therefore does not standout as an outsider? Or does such a remark refer to an English-centred standard accent such as southern English/RP? The idea of what constitutes 'no accent' is one which this thesis will investigate, incorporating the issue into the methodology.

3/. This question elicited different responses from each of the Officers. Both said that they would ask about accent, however, the PC explained that she would not ask any further questions regarding voice quality, etc, as such questions were 'leading'. She then took the conversation into describing the 'cognitive interview' and away from asking about the speaker's voice.

In contrast, the DC said that he would also ask if there was anything distinctive about the speaker, such as a speech impediment, and did not seem to be concerned that such a question could be construed as leading.

These informal interviews with two Police Officers, although short, were nevertheless illuminating. It appears firstly that speaker description is rarely solicited when talking to witnesses, and that when it is, only an undefined concept of accent and perhaps speech impediments would be focused on, with no attention paid to other voice features, such as pitch, tone and force. Secondly, the interviews confirmed that the Officers were non-linguists and had

received no linguistic training, with only a little focus being given to issues around accent and voice, apparently at the officers' discretion. It would appear that certainly in South Wales Police (but most likely across the whole of the UK as Officers are trained centrally), no specialist training is given in this area and that the Police themselves may be as untrained or lay as the public they deal with. I will return specifically to the non-linguist status of the police below in Section 1.2.5. (p.12). Finally, a lack of training in the area may be directly impacting on the confidence of the Officers to engage with linguistic issues, further exacerbating the absence of language-related enquiry.

The interviews with South Wales Police raise an important issue which needs addressing before we consider voice evidence in detail. In criminal investigations we have a situation in which both the police and the public are discussing potentially complex linguistic issues; but what training do they have to perform this discussion, description or analysis? The next two sections set out a distinction to be used throughout the thesis regarding what makes a linguist and what makes a non-linguist.

1.2.4 Linguists

The term 'linguist' is easier to apply than to define. 'Linguist' is popularly understood to mean 'someone who studies foreign languages or speaks them very well' and in the public domain, rarely understood to mean 'one who practises linguistics.'

However, throughout this research, reference will be made to applied and forensic linguists. Assuming forensic linguistics to be one of many possible applications of linguistics, practitioners of both may be considered to be 'professional linguists', people who engage in a remunerated, academic capacity with the world of linguistics. The terms 'professional linguist' and 'forensic linguist' may, in the context of this research, be used interchangeably, as it often is in the eyes of the courts.

In contrast, there are those who do not practise professionally or have any formal accreditation as linguists, yet are not bereft of linguistic insight. Their linguistic insight may come from their own personal endeavours or more formally from completing undergraduate courses, an entire degree, or higher post-graduate degree in linguistics; they may alternatively have received specialised vocational training – the spectrum of linguistic activities is wide; they may not be practising themselves as professional linguists, but they are not naïve to the technical issues that linguists regularly engage with. Any individual sufficiently experienced and competent to engage with the issues in the field of what is academically known as linguistics is then, for the purposes of this research, referred to as a 'linguist'. The definition is broad but does enable us to draw a contrast with the target group of respondents who are truly naïve to the field – non-linguists.

1.2.5 Non-linguists

Non-linguists and their linguistic naivety/awareness will be at the centre of this research. In the UK, despite the huge increase in the number of students attending higher education courses over the last fifteen years, and despite the pre-university English Language 'A' Level exams being the fastest growing exams of this standard (http://news.bbc.co.uk/1/hi/magazine/6499797.stm; accessed 30/03/07), courses in language-related subjects such as Linguistics, English Language Studies and Communication still represent a very small proportion of the total number of students in UK higher education at any one time. Swell this number with those who never attend a higher education course and the proportion of the adult general public in the UK who have never undergone any language awareness training is very large indeed. The lack of awareness or experience of linguistic issues is of course not uniformly low. There will inevitably be unevenly distributed peaks and troughs of language awareness throughout the population, with many people sharing a general interest in language at least at an anecdotal level. However, it is clear that very few members of the general public in the UK could be called linguists and the general public could interchangeably be known as 'non-linguists'.

It is probably helpful at this point to explore the term 'general public'. In the police context, the 'general public' is usually used to contrast those who work in law enforcement with those who do not. From the linguist's perspective, the contrast here is between those with formal linguistic training and those without, regardless of a person's relationship to the criminal justice system. This includes the police, whose role, despite being tasked with eliciting evidence and solving crimes, offers no linguistic advantage. For example, and as mentioned previously, although officers (in the UK at least) are provided with training in general questioning techniques, such as the cognitive interview, few police officers have any specific linguistic training, and receive no training courses and no guidelines on the elicitation or provision of voice descriptions, excepting the general guidelines for taking and providing statements. The police are therefore from the linguist's point of view as much members of the non-linguist general public as the suspects, victims and witness they deal with. This assumption can also be extended to those staffing the courts who also receive no specific linguistic training.

The resultant situation is that suspects, victims, witnesses, arresting officers and legal professionals are therefore almost exclusively non-linguists, accessing an identical linguistic resource for the provision, elicitation and interpretation of voice descriptions. Moreover, there may be little difference in terms of technical linguistic sophistication between a description of a voice in a pub, a police station or a law court, being that all the venues are populated almost exclusively by non-linguists. We can easily surmise that every hour of every day, up and down the country and around the world, non-linguist members of the general public are appointed to elicit the best possible linguistic evidence, from other non-linguist members of the general public, which other non-linguists then represent in law courts. It would be remiss of a linguist to allow this situation to persist without remark and the discussion of how non-linguists and their non-linguist evidence can be handled will be developed in Chapter 2.

1.3 Voice evidence, practitioners and practice

1.3.1 Voice evidence

When considering language evidence, as forensic linguistics does, we are potentially engaging with a remarkably wide range of linguistic data. It is useful at this point to distinguish between language evidence and voice evidence. On the one hand, as Gibbons (2003) notes, language evidence may focus on illegal speech acts – offering or accepting bribes, suborning a person to a crime, soliciting an illegal act (e.g. hiring a hitman) - and research analysis into this field of language evidence is well-established and ongoing. However, the linguistic evidence that is the focus of this thesis is not the illocutionary act or content of the message, but rather the voice that carries it, the vocal medium itself which can serve as a strong marker of identity.

> *The voice is the very emblem of the speaker, indelibly woven into the fabric of speech. In this sense, each of our utterances of spoken language carries not only its own message, but through accent [...] it is at the same time an audible declaration of our membership of particular social or regional groups, of our individual physical and psychological identity.*
> *(Laver 1994:2)*

Investigating voice evidence is therefore the process of dealing with a medium with manifold possibilities for conveying the patterns of a language, whilst at the same time conveying all manner of physiological and socio-economic information. The rest of this chapter will deal specifically with voice evidence and the area of language crime will not be returned to.

1.3.2 Brief historical background to voice evidence

Despite the lack of research into forensic speaker description, voice evidence has a long history dating back to at least biblical times, with the story of 'shibboleth', recounting how some 42,000 Ephraimites are said to have been killed for not being able to convincingly use the dialect of the tribe they were disguising themselves as. Whilst voice evidence has been referred to throughout history, what is perhaps surprising is that such crude linguistic practices were still being used in law courts as recently as the C20th.

Bruno Richard Hauptmann, for example, was executed in 1936 for the kidnap and murder of the son of the aviation pioneer Charles Lindbergh. Hauptmann was convicted when an earwitness positively identified him as the killer largely on the basis of him having to repeat a six-word sentence in court 29 months after the murder (Solan and Tiersma 2003:373). At the time, there were no expert witnesses retained by the courts to advise on matters of linguistics, no procedures for arranging a voice line-up, and no research to establish the reliability of earwitness testimony as a function of memory.

Moving the clock forward, a technical 'sophistication' was brought to an emerging field in the 1960's with the development in the US of the 'voice print' or spectrogram, a visual representation of a spectral analysis of speech, showing how the balance of energy at different pitches or frequencies changes through time. At first, this comparison of broadband spectrograms was seen, especially by the courts, as a great leap forward as it gave the analyst, for the first time, a means of visually comparing speech samples.

However, speaker description and identification have evolved greatly since the 1960's. And as seductive as it may be to analogise voice prints with finger prints, suggesting just as a finger print will reveal characteristics unique to the individual, so might a voice print, such an allusion has received widespread opposition from the academic community (Nolan 1991:483). For example, Braun and Künzel highlight:

> *It is probably not unfair to say that the visual interpretation of spectrograms does not solve the problem of identification, but simply shifts it from the auditory to the visual domain.*
>
> *(Braun and Künzel 1998:12)*

In other words, it is a non-sequitur to imply that being able to 'see' the voice enables us to consistently distinguish individuals. Finally, Nolan (1991:483) observes that the exaggeration of the reliability of voice print technology, by analogising 'voice print' with finger prints, coupled with courtroom evidence offered by people professing to be 'experts', most of whom had no scientific

education in speech science or related areas - 'charlatans' (Hollien 2002:64) - led the technique particularly, in exclusivity, into disrepute.

In contrast to an exclusively technology-reliant approach, Braun and Künzel (1998) describe a long tradition of *aural*-perceptual analysis of evidence speaker identification in the UK, that is, performing detailed impressionistic assessments of regional accents combined with an evaluation of the voice pitch, quality and other prosodic features. These auditory assessments were only to be carried out by skilled phoneticians. (As long ago as 1980, the Colloquium of the British Association of Academic Phoneticians passed a motion that, 'phoneticians should not consider themselves expert in speaker identification until they have demonstrated themselves to be so' (Nolan 1991:483)). This aural-perceptual approach has most commonly been used for *speaker profiling*, gleaning whatever information possible from a recorded voice sample, in an attempt to build a picture of the speaker's voice, accent and possible socio-economic background.

An example of the potentials of speaker profiling via an aural-perceptual approach undertaken by a skilled phonetician was documented by Ellis (1994) in his description of the hunt for a suspect in the Yorkshire Ripper case. During the investigation, letters and a tape recording were received by the police sent by someone who appeared to have an encyclopaedic knowledge of the Ripper's movements. Brought in to trace the caller, (who turned out to be a hoaxer), Ellis concluded that the hoaxer on the 257 word recording was from the Castletown area of Sunderland, north-east England. Some twenty-eight years after the investigation, in 2006, John Humble was arrested and imprisoned for perverting the course of justice by making hoax calls, having finally succumbed to DNA analysis provided when licking the envelope he had sent back in the 70's. Humble was from the Castletown area of Sunderland, north-east England. (For further details see French et al. 2006.)

It is clear from examples such as the identification of the hoax Ripper, that auditory analysis of voice evidence has had some high-profile successes, having been either the basis of or a contributory factor to the identification of

suspects and their being brought to justice. Such involvement has not been limited to individual scene-of-the-crime investigations. Tanner and Tanner (2004:48) note that since September 11th 2001 identifying a person's country of origin has taken on greater urgency in many countries around the world, with speaker profiling in particular being seen as one of the many tools that can be used to prevent future terrorist attacks.

1.3.3 Inter and intra-speaker variation

> *For various theoretical reasons, I cannot foresee a day when phoneticians will be able to identify a speaker with the degree of certainty associated with the matching of finger-prints or DNA profiles.*
>
> *(Baldwin and French 1990:62)*

Despite some successful applications of speaker profiling and voice evidence, at the heart of forensic speaker identification is the unromantic acceptance of its limitations. In seeking to describe speakers' voices for evidential purposes, we are not searching for some kind of linguistic 'dna', some congenital traits that predispose each individual to speak uniquely. The notion of each speaker on the planet having a unique idiolect is largely frowned upon by the linguistic community. Hollien (2002:7) points out that it is simply not known whether every one of the 5-6 billion speakers in the world produces utterances which are unique to them and different from all others. Moreover, Nolan (1991:489) states that in the absence of large-scale demonstration, he remains sceptical about the extent of idiosyncrasy permissible in a homogenous speech community. Scepticism regarding inter-speaker variability is matched by serious misgivings regarding intra-speaker variability.

In challenging the suggestion of a 'voice print' technique capturing an image unique to the individual, Nolan (1991:486, 1993:2) also highlights the intra-speaker variation or 'plasticity' of a given speaker, referring not only to the flexibility of the speech organs but also to the language they are delivering at any given moment. Braun and Künzel (1998) elaborate:

> *The most fundamental issue is that it has never been shown that the inter-individual variation of spectral and/or temporal features of a spectrogram is greater than intra-individual variation.*
>
> *(Braun and Künzel 1998:11)*

In other words, a single individual's speech may vary to such a degree that they could be seen, if relying on spectrographic analysis alone, to be a different speaker, and vice versa. Thomas (2002:138) notes that intra-speaker variation is fundamental to sociolinguistics but that these differences are often regarded lightly by more experimental branches of linguistics. It is the recognition of the role of both inter- and intra-speaker variation within forensic linguistics that has focused efforts on codes of practice and innovation in the handling of voice evidence.

1.3.4 Current practice

Regardless of the issues around intra- and inter-speaker variation, voice data are still being processed and utilised in legal contexts and there is a rationale for why this practice should continue. We must remember firstly the number of potential candidate speakers or 'authors' (Coulthard 2004) for voice identification. For example, when carrying out any form of speaker description or analysis, we are unlikely ever to be dealing with a database of candidate suspects from everywhere in the world:

> *For one thing, you do not have to compare your target speaker (i.e. the unknown) to everyone else in the world. If it is a man, you immediately eliminate women and children. If he speaks a particular dialect and/or language, you eliminate speakers of other languages and dialects, and so on [...] thus you end up with a reasonably small number of 'potentials'.*
>
> *(Hollien 2002:2)*

What this means for the handling of voice evidence is that while the jury may be out in terms of potential uniqueness of an individual human voice, linguists are unanimous with regard to the potentials for linguistic analysis and the corroborative role it may play in an investigation.

Moreover, the last 20 years have seen a development in professional practice for forensic linguists, not only in the formation of professional organisations such as the IAFL and IAFPA (p.3) and the establishment of codes of practice (p.16) but also in the 21st century, the exclusive application by a skilled phonetician of auditory-only analysis, can no longer be seen as sufficient for speaker profiling or speaker identification due to its subjective component

(Braun and Künzel 1998). There are now very few practitioners left who do not employ a combination of both auditory and acoustic analysis, a view asserted by Dr Peter French, in 2002 in the Appeal Trial of Anthony O'Doherty, a man who, five years earlier, had been convicted for aggravated burglary and causing grievous bodily harm based on an auditory-only analysis of a mobile phone call. Dr French stated that he knew of only one remaining expert in the UK who relied on auditory analysis alone (O'Doherty 2002).

Most practitioners agree then, that the combination of both auditory and acoustic analysis is legally the safest and analytically the most productive way forward, with each analysis providing a cross-check against the other. This 'self-imposed' best practice has been adopted world-wide by most experts in the field and has more recently been recognised by the courts. When both Dr French for the prosecution and Dr Francis Nolan for the defence articulated their view during the O'Doherty Appeal Trial that neither analysis should be used without the other, such was the strength of their expert witness evidence that the Appeal Court Judge Nicholson concluded:

> *In the present state of scientific knowledge, no prosecution should be brought in Northern Ireland in which one of the planks is voice identification given by an expert which is solely confined to auditory analysis. There should also be expert evidence of acoustic analysis such as is used by Dr Nolan, Dr French and all but a small percentage of experts in the United Kingdom and by all experts in the rest of Europe, which includes formant analysis.*
> *(O'Doherty 2002:17)*

Moreover, the O'Doherty ruling is seen as so significant that it has in most cases been extended across the whole of the United Kingdom to apply to courts countrywide. In other words, the academically rigorous trajectory of modern forensic linguistics has endowed the field with increasing and substantial legal credibility in the eyes of the courts and academia, with only those professionals of a high calibre practising in the field, and the courts broadly concurring with their analytical approach to the evidence.

Forensic speaker identification nowadays combines both auditory and acoustic methods to analyse recorded speech samples and produce speaker profiles, advising law enforcement authorities of the socio-economic and/or regional

background of a speaker. Practitioners nowadays would be expected to consider both approaches when identifying pronunciation idiosyncrasies unique to one speaker or the content of poor quality recordings, or undertaking tape authentication and even constructing an earwitness line-up (parallel to the eyewitness line-ups more familiar to the general public through the popular media).

What we can confidently infer from the increasingly higher profile of cases involving voice evidence, also from the professionalisation of the field in the form of the IAFL and IAFPA, and from the increased research activity in academic institutions around the world, is that the modern approach to handling voice evidence is enjoying greater credibility with legal professionals than at any time before.

1.3.5 Voice evidence: major issues and next steps

Handling voice evidence in the context of forensic speaker description is always likely to present problems not encountered in other lines of investigation. Consider for example an assault that takes place in a park after nightfall. There may be CCTV footage or DNA samples available, which could be of great assistance to the police. However, dependent on the positioning of cameras, available light and precautions taken by the assailant, there may not be any recorded evidence that can independently be analysed under scientific conditions. Moreover, if there is any recorded and/or analysable material, CCTV and DNA samples reveal little with respect to the social or regional background of the suspect in comparison with voice evidence (Rose 2002). The overall effect of a lack of voice recording is inevitably to diminish the forensic resources available to the law enforcement agencies and is detrimental to the investigative process as a whole. Of course, with no recorded voice evidence, the police are unlikely to be able to call upon the expertise of a forensic linguist; neither will they be able to use sophisticated technical equipment (formant analysis, etc) to assist their investigation.

A comparison with the visual medium further illustrates the difficulties that voice evidence or lack of it may pose. In the UK in particular, there is often likely to be some CCTV footage at least of a perpetrator approaching or leaving the crime scene, due to the widespread use of CCTV in British towns and cities. However, where this is not available, but a member of the public or *eyewitness* has seen the perpetrator, they may provide a description of the person. The critical difference here is that the eyewitness will be assisted in providing a description at the police station by computer video technology with which the eyewitness can reproduce an image of the suspect. The *earwitness,* by contrast, is not provided with any technological resources to assist their description of a suspect's voice, simply because the technology to reproduce what was referred to in the Introduction (p.2) as an 'audiofit', a synthesised reconstruction of the suspect's speech, voice qualities and accent, does not currently exist. What evidence there is, is subjectively encoded, existing solely in the mind of the earwitness. It should at this point be stated that it is important to distinguish between the encoding, memorisation and recoding of visual data compared with aural data, and the current research does not assume that the two processes are cognitively identical. It is, however, seen as valuable to compare the *aims* of videofit and audiofit processes, and how the evidence may contribute to the investigative process.

A simple table comparison of forensic speaker description (FSD) with visual identification via CCTV (CCTV), with eyewitness description (EYW) and with forensic speaker identification (FSI) illustrates some of the major issues facing forensic speaker description work:

Table 1.1 Comparison of visual and linguistic evidence sources.

	CCTV	EYW	FSI	FSD
A recorded sample of evidence is available	✓	×	✓	×
The evidence has a technical source	✓	×	✓	×
The evidence can be technically assisted/enhanced	✓	✓	✓	×

As we can see in the table above, where there is no recorded voice evidence available, the likely scenario is that the only distinctive evidential feature of the perpetrator – the voice – may be manifest solely in the memory of the earwitness with no established practice for how to access it. Forensic speaker description is what we might call the most impoverished source of evidence.

The problem of accessing descriptions from earwitnesses is further exacerbated by the fact that few of the people who are ever likely to be earwitnesses will ever have had any linguistic training. There is a weak diffusion of linguistic expertise (referring to the analysis of language rather than the ability to speak other languages) throughout a given population and professional linguists are, relatively speaking, few and far between. Barring a particularly targeted vendetta, linguists are never likely to form more than a fraction of those who are a victim of or witness to a crime. What this means, is that few earwitnesses will be sufficiently skilled or knowledgeable to perceive and describe in detail the regional or social characteristics of the perpetrator's voice or any idiosyncrasies in the delivery. Forensic linguists agree: Yarmey (2001) talking from his own experience in the field claims that even when prompted, witnesses lack the analytic skills or available terms or concepts to describe relevant voice characteristics and when they do they are very general in nature. Shuy concurs:

> *"The non-linguist does not have the meta-language to describe his/her own language."*
> *(Shuy 1993:14)*

This issue is made graver by the fact that non-linguist status extends to both sides of the earwitness/police interaction. These interactions will almost certainly involve exclusively *non-linguist* participants, handling *linguistic* information, as the police too have almost invariably been left untouched by any specific linguistic training. Philippon et al. (2007:111) in their investigation of non-linguists' and police officers' knowledge regarding earwitness identification, note that police officers would be no more efficient at dealing with, organising and collecting this kind of evidence than ordinary members of the public. They conclude that job experience (i.e. police service) appears to make no obvious difference in performance. It is perhaps not surprising then, that forensic

speaker description has not been the most productive or best-pursued line of enquiry.

In summary, the professionalisation of the fields of forensic linguistics, phonetics and identification has highlighted the main problems for the field of forensic speaker description: on the one hand, it has thrown the lack of linguistic training on the part of the general public into sharp relief and cast doubt within the academic community over the abilities and reliability of non-linguist earwitnesses to produce an anecdotal speaker profile of the voices encountered; on the other hand, it has highlighted the tension between the academic community's lack of understanding of non-linguist description and the fact that so long as there are earwitnesses to crimes that have no aural recordings, eliciting forensic speaker descriptions will continue to be a necessary part of the legal investigative process.

It is the contention of this thesis, firstly, that what non-linguists notice, how they describe what they notice and what factors may influence their perceptions and reports, such as the extent to which an earwitness's own social stereotypes may play a role when encoding and describing the voice and accent encountered, is an area that is little understood rather than of little academic research merit. Secondly, it is incumbent upon academic researchers to consider the contribution they may make in exploring the potentials and limits of non-linguist audio-perceptual descriptions.

It is at this point that the thesis turns its first corner. Existing research and procedures for handling voice evidence give us methodological starting points and boundaries, such as the combination of the established acoustic and audio-perceptual approaches mentioned in the O'Doherty ruling above (Criminal Cases Review Commission 2002). Yet they nevertheless pertain to *recorded* voice evidence and do not illuminate the field of forensic speaker description sufficiently to enable the construction of an audiofit. Moreover, even the perception literature is sparse - Thomas (2002:113) points out that perception research has been the neglected stepsister of production research. What is needed, therefore, is to look beyond the forensic literature and pursue a wider

socioperceptual line of enquiry, examining the literature in more diverse linguistic fields in order to expand our understanding of the potentials for an audiofit and establish possible research paths. The following chapter will examine the wider research conducted by phoneticians, forensic linguists and sociolinguists, social psychologists and will represent a drawing together of findings from the loosely-titled *socioperceptual field*.

2 - NON-LINGUIST PERCEPTIONS, CONCEPTIONS AND EVALUATIONS

2.1 Introduction

Chapter 1 offered an introduction to forensic linguistics and speaker identification, highlighting the role of the professional forensic linguist, developments in technology and practice, and issues surrounding voice evidence produced by non-linguist earwitnesses. The case was made that despite the difficulties inherent in working with non-linguists and their voice evidence, such evidence may nevertheless play a significant role in criminal investigations. Yet the academic community's dim view of the integrity of forensic speaker description has resulted in non-linguist perceptions of voice evidence being largely overlooked. It is the task of countering this neglect that will be driving the thesis.

When focusing on speaker perception we can, from the outset, identify three major themes that *forensic speaker description* can address: (1) What do non-linguists perceive in a voice and what words do they use to articulate these perceptions? (2) What influences the impressions they have of the speaker and their voice and how does this influence their description and evaluation of the speaker? (It should be understood that throughout the remainder of this thesis, the term *forensic speaker description* assumes both descriptive and evaluative aspects of non-linguists' articulations.) (3) What are the practical considerations in designing research around non-linguists' perceptions of speech? The following chapter addresses these questions by engaging with the wider literature, bringing together the relevant findings and methodologies that have been utilised in research with a socioperceptual angle.

Chapter 2 begins with sections oriented towards the first of the major thematic questions identified, above, i.e. what non-linguists perceive in a voice and what words they use to articulate their perceptions. Section 2.2 discusses the role of non-linguists, their naivety and their potential as informative and insightful respondents to questions on linguistic perceptions. Section 2.3 then considers voice evidence, addressing accent and physical speech features. This section will provide a brief gloss of the modern history of accent investigation, focusing in particular on the role that accents can play in speech perception, their associated social characteristics in the criminal context and the ways in which non-linguists describe them. The chapter will then move on to how phoneticians and linguists have sought to categorise and characterise physical speech features and the terms that researchers have employed in order to present basic voice descriptor frameworks to non-linguists.

Both Sections 2.2 and 2.3 will include the view from what one might call traditional sociolinguistics and phonetics but will also look beyond to include discussion of evidence found in the emerging field of folk linguistics - a small but methodologically highly relevant literature on what non-linguists believe about language, the social information that can be offered by members of a speech community, the cognitive underpinnings of those beliefs, and how one might go about eliciting and interpreting non-linguists' views. Folk linguistics explicitly questions non-linguists regarding their perceptions and beliefs regarding language and language variation, examples of which will be given throughout the chapter. It would appear that some of the commonly-held perceptions regarding non-linguist capabilities, held by linguists and the police, are yet to be fully substantiated and the folk linguistics literature explains why further research on non-linguist beliefs is necessary.

The second of the major thematic questions above is addressed in Sections 2.4 and 2.5. Section 2.4 begins with an exploration and discussion of the nature of attitudes: What is an attitude - in general and as a cognitive activity? How do they form? When is an attitude not an attitude? What purpose do they have? Section 2.5 focuses specifically on language attitudes, with examples from the public domain and common non-linguist beliefs regarding language and English

specifically. This section discusses the social meanings of language varieties including social stereotypes, and the processes by which they and their speakers are labelled by speech communities. Language attitudes research is seen as a highly relevant resource for this thesis due to the substantial body of literature that has focused on how non-linguists' evaluations of voices, opinions, decisions and actions, even attributions of guilt may be mediated by factors such as the context of the language encountered, by the earwitnesses' own linguistic background and by their attitudinal beliefs and social stereotypes regarding language varieties and their speakers. Both Section 2.4 and 2.5 serve as a backdrop to the decision to include language attitudes methodology in the current research design (Chapter 3).

The third of the major themes regarding the practicalities of designing research into speaker perception is the focus of the remaining sections, 2.6 to 2.14. These sections are more practical in content, setting out a summary list of issues a researcher should take into consideration when planning research on forensic speaker description. The methodological decisions are briefly introduced in Section 2.6. Sections 2.7 to 2.11 focus on *system variables* by examining in 2.7 the selection of attitudes objects, followed by a discussion in Section 2.8 of direct and indirect approaches to speaker perception research. Section 2.9 considers issues relevant to questionnaire design. Section 2.10 explores considerations for the selection and recording of voice samples. Section 2.11 addresses the types of tasks that one might include in speaker perception research.

Continuing the exploration of the major theme of research design, Sections 2.12 and 2.13 focus on *estimator variables*, exploring the influence of factual details, age, sex, respondent region of origin, etc. whilst also more psychological aspects such as the role of social stereotypes. Section 2.13 focuses particularly on estimator variables in relation to the interaction between social stereotypes, evaluations and of attributions of guilt.

The aims of this chapter then, are to review the socioperceptual literature dedicated to dealing with, interviewing and surveying non-linguists' responses

to language behaviour and to explore the ways in which linguists can assist non-linguists in providing maximally detailed, accurate and meaningful descriptions of encountered language behaviour. Based on the considerations set out from 2.2 to 2.13, this chapter closes with a set of central research questions that this study has been designed to answer, with the aim of developing our understanding of forensic speaker description.

2.2 The naïve non-linguist

The first step on this chapter's journey through socioperceptual literature is to consider in a little more detail those people who are the focus of this study, the non-linguists. As discussed in Chapter 1 (p.12), the term 'non-linguist' refers to participants, on both sides of the police-witness interaction. And as a matter of respect to participants in social surveys, this study follows the view espoused by Niedzielski and Preston (1999), employing the term 'non-linguist' to indicate "those not trained in the area under investigation," eschewing any value-laden connotations such as "rustic, ignorant, uneducated, backward, primitive, minority, isolated, marginalized or lower status individuals" (Niedzielski and Preston 1999:vii). It is important to raise this issue at the outset, as our understanding of what distinguishes non-linguists from linguists will guide our expectations of what they may provide and how we might best elicit their views and descriptions.

Having decided the non-pejorative referent of the term 'non-linguist', one is bound to observe that this is a term that would nevertheless be meaningless amongst those to whom it is applied. Despite the fact that non-linguists form the overwhelming majority of language users, and despite the scale and breadth of language use amongst non-linguists, using and responding to language every day, playing with it, modifying its words, syntax and sounds, teaching it to their offspring, informing, negotiating and making decisions based upon the ideational, indexical and pragmatic features of it, it is apparent that the ability to use language does not align with a technical understanding of how language works.

If we analogise non-linguists' use of language with the use of a motor car, one can say that non-linguists are mostly drivers who can navigate their intended courses and adapt to innovations and surprises along the daily route. They do not, however, possess the design knowledge of the draughtsman, the technical knowledge of the engineer or the diagnostic knowledge of the mechanic. Non-linguists have not been trained to understand how language is constructed, how it works or how it evolves; they are rarely more than users of a vehicle for a variety of personal and professional purposes with mere occasional glimpses of technical insight picked up along the way.

Moreover, not only do non-linguists lack a technical understanding of language, but they are also unlikely to reflect on their lack of technical linguistic understanding as a potential problem. For example, in questioning non-linguists on their views on technical issues such as attaining proficiency in a language, the response to Niedzielski and Preston was that such issues do not need to be relegated to 'experts' (1999:201). Davies (2007:137) remarks that one of the challenges facing professional linguists today is that the status of the professional linguist is not as clear cut in the public perception as that of other fields; that is, if they can successfully navigate the social world every day with language as their rudder, what value is there in studying what they can already do 'competently'? Many speakers consider themselves equally qualified to comment on linguistic issues, and seem to think that linguists complicate things unnecessarily. One could argue that the public perception of linguists serves as a barometer of public perceptions of language, and there will be no change in one without a change in the other.

We can surmise that Niedzielski and Preston's respondents are largely unaware of their own linguistic naïvety and to the author's best knowledge, there is no research indicating that non-linguists view themselves as linguistically naïve or unable make linguistic analyses, and there are no extant studies indicating large-scale non-linguist insecurity regarding their ability to evaluate their own linguistic accuracy.

On a wider scale, the fact that non-linguists are unaware of their non-linguist status does not (on the surface, at least) appear to trouble societies in general, and it certainly does not inhibit the non-linguists in police stations from eliciting and offering linguistic descriptions. One may even infer, on an institutional level, that this is a situation that the police and the courts are in fact satisfied with. However, those trained in linguistic science will be all too aware of both the lack of technical linguistic understanding shared by the general public and the general public's lack of concern for what they do not know. Moreover, professional linguists have experienced first-hand the potentially serious and far-reaching consequences for the criminal justice system that may arise from an institutional disregard for those who specialise in linguistic science, as was documented earlier in Chapter 1 (p.5) with the example cases of Coulthard's (1997) dismissed High Court appeal and Rodman's (2002) summary of the 'black accent' trial.

What we are presented with here is a tension: forensic linguistics concerns itself with language evidence, and language evidence requires, or would be best dealt with, by those with a technical acquaintance with linguistics; however, most language evidence is both provided by and handled by non-linguists. From the linguists' point of view, a resolution to this tension could be to dismiss a large amount of language evidence that non-linguists could potentially provide on the basis of it being technically ill-informed and lacking rigour, objectivity and consistency. However, there is a number of arguments that should deter such dismissal: firstly, an anthropological argument points to the need to understand the nature of the linguistics of the group providing the data. Preston (2005) and Niedzielski and Preston (1999) offer warnings:

> *"How could one imagine doing applied linguistics without knowing the folk linguistics of the group with whom the work is to be done? To do so is to invite at worst disaster and at best unexpressed or even sullen disdain for one's attempts."*
>
> *(Preston 2005)*

> *"It is simply dangerous not to know what people believe about language and how they respond to it: it is dangerous to general linguistics, dangerous to applied linguistics, and even debilitating to the desire for a complete account of language and its users."*
>
> *(Niedzielski & Preston 1999: xiv)*

A second reason for not dismissing non-linguists' language evidence is practical, in that we must remember that non-linguists' evidence is still evidence and it must all be dealt with. Most witnesses to a crime are non-linguists, and most linguistic evidence will be provided by non-linguists. Non-linguist language evidence is therefore impossible to avoid; better, surely, that it is dealt with by someone with linguistic training than not. Thirdly, to ignore linguistic evidence on the basis that it is being provided by and handled by non-linguists would be detrimental not only to forensic speaker description and forensic linguistics, but to criminal justice also.

Finally, there is encouraging evidence to suggest that non-linguists can in fact provide insightful and beneficial data with regard to linguistic enquiry. Preston's and Niedzielski's (1999) work has helped to highlight that non-linguists do indeed have perceptions regarding language, across the range of linguistic variables. But of particular interest to this research is that Preston has obtained much data regarding non-linguist perceptions of speech articulation, and non-linguist perceptions of regional variation, by providing, for example, hand-drawn maps with which non-linguists may articulate their perceptions. Preston (1996) in his study of non-linguists placing accents along the transect from Indiana to Alabama, noted the high degree of accuracy with which listeners correctly identified speakers on the north-south scale.

To conclude, there is a pressing demand to engage with the group most affected by the issues under investigation - non-linguists. As linguists, we must, regardless of the lack of technical insight that characterises non-linguists, find ways of engaging them on subjects of linguistic detail; we must hear the world through the non-linguists' ears; Preston (1996) made a plea for the systematic study of non-linguists' opinions of language varieties to complement professional linguists' insights about 'scientifically discovered aspects of language structure and use' (Preston 1996:72). We must draw comparisons between the observations and approaches of the professional linguist and the non-linguist, and identify what non-linguists notice; we must observe what they hear, and find common themes that enable us to engage non-linguists on their own terms; we must consider what we can do to facilitate non-linguist

description; we must observe attitudinal behaviour and develop sound predictions on how language beliefs and language evidence may interact; finally, we must experiment with identifying fruitful methodologies for eliciting these aims. The next two sections embark upon addressing these issues, reviewing the literature on how linguists and non-linguists perceive and describe both the accent and physical speech features of voice data.

2.3 Voice evidence

2.3.1 Voice evidence: accents

Although the term 'accent' is a part of every day speech, this section considers accent from a technical perspective, noting some of the landmark issues that have formed the focus of accent study in the recent history of linguistics. The section will then progress to discussing research that has been conducted on how non-linguists describe accents.

Abercrombie (1967:3-7) explained that the pattern-forming potentials of a language are much greater than a language requires us to use. The medium of speech, not being completely taken up by the patterns that convey language, is able to accommodate a quite complex system of non-linguistic signs, which Abercrombie terms *indexical* features (as distinct from *linguistic* features related to conveying the patterns of the language). When we discuss accents, it is these indexical features that we are referring to. Such indexical features may relate to a number of social meanings across a speech community, frequently acting as a cue to signal a speaker's group membership(s).

The first of these social cues to be considered relates to regional variation. Regional variation in a language has been the focus of a long tradition of sociolinguistic research dating back to the early twentieth century linguistic atlases through to more substantial dialectological undertakings such as the Orton's Survey of English Dialects (Orton et al. 1962-1971), which culminated in the Linguistic Atlas of England (Orton et al. 1978). Early dialectological work concerned itself with capturing the disappearing traditional dialects of the

country and consequently was consciously biased towards rural communities with static populations, the respondents being mostly what Chambers and Trudgill (1980:33) refer to as NORMs: 'nonmobile, older, rural males' (males being seen as more conservative in their speech forms than women). The questions asked of respondents focused on every-day aspects of their lives, such as farming, the house, animals and the weather. The data, once collected, enabled dialectologists to plot isoglosses on to a map to represent where traditional variation takes place across regions, utilising what was to be considered the classical model of dialect variation research; a two-dimensional spatial distribution of indexical features and variations, which served an important function as a record of archaic speech, in reference book form. However, the gaps that this research did not fill are large: it did not take account of either gender, the urban environment, or the range of ages and socio-economic backgrounds and that inhabit real, every-day communities.

Many of the limitations of traditional dialectology were to be addressed by later research. Studies such as Labov's (1966) 'Social Stratification in New York' and Trudgill's (1974) work on the 'Social Differentiation of English in Norwich', both moved the field forward by shifting the focus to include the urban environment where, in England, some 90% of the population live, thereby increasing the representativeness of the findings. A further innovation was provided by Kerswill (1996), who extended this new branch of dialect investigation by investigating the birth of dialects in new towns where speakers from far afield were coming into prolonged and regular contact with each other, resulting in dialect levelling.

What work such as Labov's and Trudgill's early works also achieved was to shift the research focus to social variation and *sociolects*, examining how accent varies socially (as opposed to regionally) within communities. Both made explicit how a speaker's use of prestige or standard linguistic features in a community significantly depends on a speaker's social status. The variationist approach to language analysis used in these and subsequent studies documented how, when language varies from one user to another, these variations tend to appear not randomly, but in socially meaningful patterns, enabling us to identify and distinguish co-existing regional *and* social speech

communities. As a result, sociolinguists have come to observe that the act of producing indexical features, or examples of a certain speech variety, is likely to serve as a predictor for membership not only of a particular regional group, but also a social group, endowing the listener with further information regarding the geographical and social background of the speaker. Reciprocally, membership of a social or regional group often (but not always) acts as a predictor of the indexical features or speech variety to be produced.

Shifting the focus to the social complexity of accent variation, other research has considered how non-linguists index the accent variation they encounter. A large body of language attitudes research has developed investigating how non-linguists index social and regional patterns that exist in a speech community. Lambert et al. (1960) noted that hearing a language variety is likely to arouse in the non-linguist mind generalised, stereotyped characteristics of a group. In their research on the evaluations of French and English Canadian speakers in the bilingual context of Montreal, Lambert et al. found that the social meaning of accents resulted in French and English speakers being evaluated differently, with correlations being noted between the first language of the respondents and their judgements of the speakers. For example, the English-speaking respondents evaluated English speakers more favourably than French speakers; the French speakers not only evaluated English speakers more favourably but that their evaluations of French speakers were reliably less favourable than those of English speakers (Lambert et al. 1960:50).

This early language attitude work was expanded upon by Giles (1970), working in the UK context, where he asked British school students to rate a number of accents, including Birmingham, Cockney, Liverpool, Northern English, RP Scottish, South Welsh and the non-UK accents of French, German, Irish, Italian and North American. Giles found that not only were the speakers/accents evaluated differently, but that there was a distinct hierarchical structure to the evaluations: RP was rated highest in terms of prestige, and accents related to industrial areas and cities such as Liverpool, London (specifically the 'Cockney' accent) and Birmingham received low evaluations. Foreign accents such as

French and North American occupied positions of relatively high prestige, being afforded higher status than any regional British accents.

What Lambert et al., Giles (1970) and subsequent research (Williams et al., 1996; Garrett, et al., 2004; Bishop and Coupland, 2005) have shown, is that whether questioned directly or observed indirectly, non-linguists meaningfully index the differing accents heard, not only along lines of regional and social distribution, but also along other evaluative dimensions, such as social prestige and social attractiveness. This evaluation need not only be prompted by physically hearing a recording of the accent; Giles (1970), Garrett et al. (2003) and Garrett et al. (2005) have all noted that the simple mention of the social or regional group - *conceptual labelling* - can provoke the same evaluative responses from non-linguists.

Furthermore, research in language attitudes has revealed how the evaluative dimensions of speaker accent by non-linguists exhibit a tri-partite hierarchical structure: Zahn and Hopper (1985) pooled scale adjectives from a large number of speech evaluation research programmes and used them in a single study involving nearly 600 judges. They found, through a factor analysis, that each of the scales loaded on to one of three differentiated factors which they labelled 'superiority', 'attractiveness' and 'dynamism'. It is now well established in language attitudes research that it is these three dimensions that judges operate with when evaluating speaker accents (Garrett et al. 2003:63), and that standard accents are rated more positively, especially on traits associated with competence or status (e.g. intelligence or literacy). We can observe, for example, how accents may systematically (dis)advantage people in such institutional contexts as job interviews for management positions, etc (Dixon et al. 2002:162). The implication for forensic research is that if accent evaluation depends on the sociolinguistic context (Giles and Coupland 1992) then research regarding earwitness descriptions of unknown speakers should investigate the extent to which accent evaluation may be mediated by criminal inferences. The work of language attitudes researchers will be discussed at greater length in Section 2.5.

The link between accent evaluation and forensic context has to some extent been explored by a few studies in the previous thirty years. Lind and O'Barr (1979:67) described research involving mock juries in the courtroom and argue that credibility judgements are a major psychological process involved in the reception of court speech. They claim that considerable social evaluation occurs in the courtroom, with judges or jurors including evaluation of speaker accent in their attempts to assess witnesses on dimensions such as competence, trustworthiness, attractiveness and social dynamism. In turn, a limited but revealing number of studies has been conducted specifically on the interaction between speaker accent and guilt.

In the first example, Seggie (1983) divided respondents into three groups, with each group being presented with the details of a crime. The crime differed for each group.[1] The first group read details of a crime of violence against a middle-aged man in a city park leaving the victim unconscious, with serious injuries; the second group read details of a young middle-class married couple with two children, whose home was vandalised whilst they were away for the night visiting grandparents; the third group read a success story of a young man who had set up his own thriving business. The auditors then discovered that an employee had embezzled a large sum of money. All respondents were informed that the police had interviewed a large number of suspects and that three had had their voices recorded. The respondents were invited to listen to the three in turn and asked to rate the probability of each speaker having committed the particular crime. The voices were recorded using the matched-guise technique: one speaker recorded the accounts in three accents - broad Australian, Malay Chinese and English RP. The foil's three voices were played to the respondents under the guise of being three different people. The results indicated that pre-existing evaluative structures for the English and Australian accents are highly formed, with the Australian being rated as significantly more probable of being guilty of the crimes involving violence, whilst the English accent attracted a higher guilt attribution for the crime of embezzlement. The results point to the conclusion that accents and their crime type interacted to shape assessments of guilt, and that these assessments reflected social stereotypes regarding the actions of suspects with standard and non-standard accents.

In a second example, Dixon et al. (1994), asked white South African student respondents to assess the guilt of a Cape Afrikaans vernacular speaker (a vernacular which is 'linked to a working class ethos') being interviewed on suspicion of committing a crime. In the recordings, the speaker under interrogation either diverged away from the interlocutor into Cape Afrikaans or converged into English. Dixon et al. also examined the influence of crime type on the attribution of guilt, utilising both 'blue-collar' crime (physically and verbally assaulting a colleague) and 'white collar' (passing fraudulent cheques) as a variable. Two main findings arose from the research: firstly, the suspect who diverged into Cape Afrikaans was found guilty more often than the one who converged into English. Secondly, the suspect was deemed guilty more often when he was accused of committing a blue-collar crime than when he was accused of a white collar crime. Dixon et al. conclude that the raters responded to the suspect as an exemplar of his social category rather than as an individual and that the results were consistent with previous research regarding accent and social stereotypes and crime-related racial stereotypes.

In the UK context, Dixon et al. (2002) conducted a small study replicating Seggie's (1983) research (above), using a Birmingham accent, a 'third class' accent that has, since Giles' first research in the 1970's, been more negatively evaluated than either rural, regional or RP accents. They found firstly that there was a positive correlation between suspect's guilt ratings and the general perceptions of the suspect's competence and attractiveness. They also conjecture that non-standard speakers are generally perceived as guiltier than standard speakers because their testimony is deemed less assured and therefore more closely associated with shiftiness or related criminal stereotypes.

The most recent of the studies regarding the interaction between attributions of guilt and accent was conducted by Dixon and Mahoney (2004). Their research presented mock interviews with the police to a group of 199 British undergraduate students. They employed a matched-guise technique with the suspect foil recording interviews in an RP and a Birmingham accent; they then asked the respondents to rate the speakers along the lines of the hypothetical suspects' guilt, criminal typicality and likelihood of being accused of a future

offence. They found that the regionally-accented suspect was rated as more typically criminal and more likely to be re-accused of a crime than was a standard-accented speaker. Their evidence is consistent with the previous research, suggesting that linguistic characteristics of accents may influence judgements of criminality.

The research detailed above gives us a strong indication of the relevance of the indexical features described by Abercrombie (1967:3-7), their status as social cues within a speech community, how they interact with learned associations with reference groups and evoke stereotyped responses in non-linguists. Non-linguist responses are mediated by preferences and ideologies that dominate in listeners' own communities, and which influence other evaluative judgements they make (Garrett et al. 2003:209). In forensic contexts, indexical features such as speaker accent appear to function not only as salient markers of group identity, but may also have a particular evaluative significance. Dixon and Mahoney conclude that such indexical features can therefore be added to the list of extralegal factors that may affect forensic impressions, attributions and decisions (Dixon and Mahoney 2004:64). To understand the psychological processes behind how such social cues such as accent evoke attitudinal evaluations, one needs to understand more regarding the nature of attitudes and this will be considered in more detail in the following Attitudes section (2.4) and the Attitudes and Language section (2.5).

The final facet of accent to be considered in this section is one of description. We know that for over 100 years, phoneticians and dialectologists have been transcribing regional and social variation in speech communities around the world and there has developed a rich tradition of linguists describing speech production. We also know that in the forensic context, as the interviews with the PC and DC from South Wales Police revealed, there is also a need for non-linguists to describe speech and whether the speaker had 'an accent'. However, where there has been great progress in the development of highly detailed systems such as the International Phonetic Alphabet, which enable trained linguists to record and encode what they hear in speech produced by others, and which have formed the basis of sociolinguistic dialect maps, non-linguist

speech perception has been studied far less (Thomas 2002:113). There is, for example no research that directly investigates earwitnesses' perceptions and descriptions of accents; most research at the interface between non-linguists and the accents they encounter has tended to investigate non-linguist evaluations rather than descriptions of accents. Preston (1996) complained that whilst the same research asked people to evaluate accents, it had failed to determine where the non-linguists thought the regional voices were from. In other words, even those working most closely with speaker perceptions have neglected to investigate accent descriptions, even when to do so would presumably strengthen the validity of their research claims.

Despite the impoverishment of the field where speaker description is concerned, there is nevertheless a handful of studies that have dealt with some aspects of accent perception. They can be subdivided into two thematic groups: (1) recognition of accent; (2) description of accent.

(1) Recognition of accent: Recognition can be viewed as the cognitive mapping of audible speech-features on to individuals' records of the usage norms of particular communities (Garrett et al. 2003:208). This is not a transparent process, as what non-linguists record is filtered by their perceptual labelling abilities, practices and strategies (Kerswill and Wright, 1990). For example, non-linguists are unlikely to remark on the same depth of features that one would expect a professional linguist to identify, given that professional linguists identify greater detail in speech features and have academically researched records onto which to map linguistic behaviour. In order to account for what non-linguists record and replay, Preston sets out four modes of folk linguistic awareness:

Unavailable: the folk will not comment on certain linguistic topics
Available: the folk will comment on certain linguistic topics if engaged, but do not normally do so
Suggestible: seldom discussed in ordinary conversation, but they will discuss them if they arise, and without much assistance
Common: topics of usual folk linguistic discussion (Preston 1996:40)

If one accepts Preston's model as a possible taxonomy for non-linguist awareness, we must consider what determines which mode of folk awareness is invoked. Kerswill and Williams (2002) note that although there are many factors that may play a role in non-linguist awareness, this field has been the focus of relatively few studies. Factors of particular note are:

(i) The life experience of the judges: depending on location and culture that one has grown up in, some language varieties are better known than others. An example is Stephan (1997) who played recordings of twelve speakers of English dialects from around the world to university students in Germany. Stephan found that American English was identified correctly most frequently, despite English English being the nearest dominant variety. The data exemplified how in many parts of the world, especially the former allied-occupied West Germany, US media dominance through music, film, tv and information technology has resulted in US culture and voices becoming highly familiar to the masses, geographical distance becoming a secondary factor. Moving to native speaker respondents, Preston (1996) observed how the respondents' regional backgrounds influence the identification of varieties. In his study on a transect from Michigan to Alabama, two groups of respondents, one from south-eastern Michigan and one from southern Indiana were played two groups of speakers and were then asked to plot on a map the voices that they heard. The respondents from Michigan and Indiana differed in which speakers they were best able to discriminate, with a bias towards regional success.

(ii) The sociolinguistic maturities of the judges: Garrett et al. (2003) note that in their comparative study between Welsh teenagers and Welsh teachers, a difference was observed between Welsh teachers' ability to recognise Welsh dialects, and that of the Welsh teenagers, with teenagers being less successful at the task than their teachers. They point out that recognition can be accounted for by their predictably lower levels of dialect experience; the young adults have probably experienced lower geographical mobility and less access to dialect speakers, either face-to-face or in the broadcast media, resulting in a less accurate or detailed

cognitive template of English language variation in Wales (Garrett et al. 2003:208-209).

(iii) The subjectively-perceived social attractiveness of the speaker: Garrett et al. (2003) noted that an affective factor played a part in the recognition of dialect speakers. They recorded school students (15-16 years) in the six major dialect regions of Wales, telling stories in their local Welsh English dialects. They then played the recording to teenagers in the corresponding six dialect regions of Wales who noted their first impressions and then rated the speakers on scales of affiliation, status and Welshness. The Welsh teenage respondents' recognition of the dialects was often mediated by the likeability of a speaker, with a tendency for a very likeable speaker to be appropriated into the ingroup of respondents. Almost cyclically, Kerswill and Williams add the suggestion that perceived in-group membership in itself has the power to make a voice attractive (Kerswill and Williams 2002:174).

(iv) Supplied information and social stereotypes: At the heart of indirect measures in language attitudes studies has often been the requirement to avoid formally bringing to the respondents' attention to the regional-social background of the speaker(s). A consequence of what happens if speakers are furnished with this information is illustrated by Niedzielski (1999). Niedzielski noted how the interaction between supplied information and social stereotypes influenced the identification of speaker accent. In a study of respondents from Detroit, half of the cohort was told that the speaker was from Detroit, half were told the speaker was from Canada. When asked to match resynthesized vowels with the original vowels, respondents in each group selected differently from each other along regional lines, indicating that their perception had been altered by their stereotypes of American and Canadian speech.

(v) The context in which the voices are encountered: Lee (1971) noted that one of the chief shortcomings of many perceptual studies is the failure to take the sociolinguistic setting into account. Preston (1996) exemplifies

why the sociolinguistic context may be important, citing communicative primacy as a governing function. Preston observes that if the primary purpose is to communicate, it can be to the exclusion of noticing or reporting detail. Non-linguists may not notice glaring grammatical or pronunciation oddities as they become submerged in the event by the communicative purpose (Preston 1996:46-47).

In a forensic context where, for example, respondents evaluate a suspect's guilt, Dixon and Mahoney (2004:71) note that a suspect's accent may represent only one meaningful criterion upon which guilt ratings are based. Kalven & Zeisel's (1966) liberation hypothesis elaborates on this point, suggesting that research on respondents' evaluations of guilt indicates that when evidence that strongly incriminates a suspect is available, participants tend to give the evidence informational priority, over-riding the effects of stereotypes. In linguistic terms, as Preston suggests, this may be an issue of communicative primacy. Researchers of dialect recognition need therefore to be mindful of the sociolinguistic context and the potential for non-linguists to observe no particular accent detail; in order to maintain ecological validity, researchers need to strike a balance between making their questions regarding accent more prominent in order to prompt recognition, diverting large amounts of the respondents' attention to the issue, or backgrounding all prompts and questions at the risk of allowing other factors to flood the respondents' attention. Moreover, in the laboratory, it is often the case that respondents require an (ethically-bound) pretext to elicit their agreement to participate. Ecological validity may therefore be further compromised if witnesses are forewarned of what is expected of them, as this may eventuate in a positive cognitive set encompassing maximally efficient processing strategies that may or may not reflect those strategies employed in real-life witnessing situations (Bull and Clifford 1984:119).

(2) Description of accent:

With regards to studies eliciting non-linguists' descriptions of accents, we encounter a sizeable gap in the research. The largest amount of work that has

been conducted in this area has been undertaken by Dennis Preston through his research in folk linguistics. Preston (1989) sees work with non-linguists as highly revealing, arguing that it is important to determine empirically how respondents themselves define linguistic varieties and their geolinguistic boundaries. He advocates research using perceptual dialectology to bridge the gap between variationist sociolinguistics and language attitudes research. As part of his work in perceptual dialectology, Preston (1996) has set out a taxonomy for approaching research on folklinguistic description, and the responses that may be elicited from non-linguists. The taxonomy consists of four continua, which span recognition and description:

AVAILABLE ⟷ UNAVAILABLE

DETAILED ⟷ GLOBAL

ACCURATE ⟷ INACCURATE

FULL CONTROL ⟷ NO CONTROL

(Preston 1996:41)

The *availability* continuum has been discussed in the previous section so will not be expanded upon here; the second continuum, *detail*, suggests that while some linguists may comment on such things as a specific phonological feature, others may make more reference to a speaker 'having an accent' (as was suggested by the interviews with the PC and DC in South Wales Police); the *accuracy* continuum reminds us that although non-linguists may perceive a distinct accent, we cannot expect that they will accurately identify that accent. An anecdotal personal experience may illustrate this point: I come from East Anglia and have in my accent features that are recognisable to those familiar with the local speech varieties. Yet I have throughout my life been mistakenly identified with a large degree of consistency by non-linguists from around the UK and Ireland as coming from Australia, illustrating non-linguists capacity to mis-locate certain perceived speech features; with regards to *control*, what we may expect is that whilst non-linguists may perceive a distinct accent, and

perhaps even be able to accurately place the accent in geographical terms, they may not be able to mimic or perform it. Relating this point to the forensic context, it is perhaps an awareness of this limitation that explains the absence of a request from the PC and DC in South Wales Police for the earwitness to perform the accent they heard.

Preston not only provides a taxonomy for approaching research on non-linguist description, but also demonstrates practical methodologies for the elicitation process. Preston (1994) called for content-oriented discourse and has shown in examples how such an approach may be productive. For example, his interviews with American non-linguists revealed much about their feelings regarding good and bad language, grammar, regional, social and ethnic dialects (Preston 1996; Niedzielski and Preston 1999; Preston 1999).

An alternative approach has, however, been demonstrated which affords a swifter way of capturing critical non-linguist descriptions. Perceptual mapping, pioneered also by Preston, invites respondents to demarcate boundaries (usually on a map) or label dialects on subjective measures. An illustration of this is the exercise discussed above (p.40), where Preston asked respondents to place dialect speakers on a north-south transect line from Indiana to Alabama; this he did by providing respondents with a map of the US and the line in question. Niedzielski and Preston (1999) give a full account of how this and other perceptual mapping exercises have been used in Preston's research in the US context. In the UK context, a perceptual mapping task was designed by Inoue (1999), who invited students at Essex University to draw lines on a map of the British Isles in order to divide the United Kingdom according to accent and dialect areas (for a reproduction of the composite map of responses, see for example Chapter 3, p.132). What is perhaps of most interest here is that despite it being wholly implausible to suggest that actual dialect variation could be adequately captured in a two-dimensional map (Garrett et al. 2002:116), the results produce a perceptual map of the UK which bears many similarities to traditional two-dimensional dialect maps of the UK (see Crystal 1995:325). In Inoue's task, respondents denoted a Scottish, English and Welsh division, which in terms of perceptual distinctions is no surprise, but within England,

respondents described a north, a midlands and a south, dividing the north to include Humberside, Lancashire and a 'Geordie' north east, and the south into a western province west of Somerset, and an East Anglian region. Such results reflect traditional views of English dialect variations and show promise in the elicitation of non-linguist regional conceptualisations.

However, keeping in mind the need for researchers in forensic speaker description to understand the perceptions of non-linguists at the local level, from within their own community, one of the most important points to note is where non-linguist dialect perceptions and the variationist sociolinguists' traditional dialect divisions differ. Williams et al. (1996:172) note that perceptual maps can contrast markedly with maps of actual dialect production. They report their use of perceptual maps in their study of teachers' perceptions of dialect communities in Wales, observing that community members have differentiated representations of English dialect variation in Wales. For example, South Walians were more likely to use the generalised term 'North Wales' whereas, North Walians themselves were less likely to use the term 'north', their geo-linguistic conceptualisation of North Wales being more multiply-centred, dividing the area into at least 'north west' or 'north east'; those in South Wales were less likely to use the term 'South Wales', preferring to label communities as 'Valleys', 'Cardiff', West Wales, etc; it was also clear that for non-linguists in Wales, certain Welsh towns and cities have much stronger geo-linguistic salience than others: Cardiff, Merthyr and Swansea in the south, Carmarthen and Cardigan in the west and Caernarfon and Bangor in the north. The teachers also distinguished accents as 'rural' or 'urban', but county boundaries had apparently no geo-linguistic salience with just one exception (Dyfed). Such linguistic divisions and geo-linguistic points of salience do not match the traditional variationist map of dialect areas in Wales, which demarcates six dialects on a regional basis (Garrett et al. 1999:325). Moreover, Williams et al. (1996:186) report that the respondents consolidated their perceptions with labelled folk-categorisations, encoding social evaluations and cultural stereotypes, 'Valleys boyo', 'Little England' (referring to a part of Pembrokeshire with particularly strong English links), or 'Swansea Jack' and 'Kaerdiff' (intended to reflect the

local pronunciation of Cardiff), denoting somewhat derogatory, class-laden designations.

Both the local self-labelling and the social evaluations illustrate the need to understand how local non-linguists conceptualise the geolinguistic variation around them. The approach of a linguist with a less intimate knowledge of the area may not succeed in eliciting or interpreting locally meaningful descriptions. Furthermore, for an audiofit to be developed for use by non-linguists (the police) for non-linguists (earwitnesses), the localised, subjective geo-linguistic perspective is an essential consideration. Techniques inviting non-linguists to use conceptual labelling and social evaluations can reveal meaningful descriptions with local currency. The findings above give considerable impetus to the argument that the starting point for any kind of audiofit is for researchers to devote greater attention to the investigation of the complex system of community-level evaluative tagging attached to language varieties by non-linguists, and how this tagging is used as street-level shorthand.

2.3.2 Voice evidence: physical speech features

When considering non-linguist descriptions of physical speech features, we must first acknowledge that the voice encountered by earwitnesses and described, (or *identified* in cases of forensic speaker identification) is only one token of the voice, encountered under particular circumstances. Any voice is likely to be subject to a certain degree of variation when produced under different circumstances, and in instances of criminal activity, the voice may be affected by factors such as stress and/or accent/voice disguise. Such variation will inevitably play a complicating role in the process of matching the description with the described. What we should be clear on is that in eliciting speaker descriptions, we are asking the earwitness to describe how the speaker *sounded* rather than how the speaker *sounds* - a report of a moment rather than a generalised statement. The fact that the speaker may sound to some degree different under other circumstances should not rule out the role of voice descriptions, in the same way that a criminal may dress differently, or change aspects of the appearance such as their hair or use a mask, yet this does not

instantly undermine the integrity or usefulness of the eyewitness's description. We should offer earwitnesses the opportunity to provide descriptions of voices in their criminal contexts and then evaluate the potential of this evidence in relation to other evidence available. The question arises: what does it mean to describe how a speaker sounded?

In section 1.2.3 above (p.8), I detailed the responses from a PC and a DC from South Wales Police with regard to what questions they might ask an earwitness in relation to any voice(s) encountered. Both officers suggested that in establishing how a speaker sounded, they would ask primarily about accent. This coincides with the auditory approach to forensic linguistic analysis of voice evidence discussed above, which has frequently taken the form of scrutinising accent and dialect variation in order to narrow the field of candidate speakers to identify where the speaker comes from (e.g. the Yorkshire Ripper trial and the Yorkshire Ripper hoax trail). However, those working in the field of forensic speaker identification will warn of the limitations of relying on accent information alone. Hollien (2002:54-55), for example, states that on the one hand, when comparing, say, an evidence tape and an exemplar tape, a finding that both speak a different dialect is significant as it suggests that the two speakers are different people. On the other hand, however, a finding that both speak a similar dialect is likely to be of modest significance, as since many thousands of people may exhibit that particular system, identification could not be made on this type of evidence alone.

Any linguist working in the field of forensic speaker description with the aim ultimately of contributing to the design of an audiofit (and in the absence of recorded material for acoustic analysis) should then consider looking beyond perceptions of accent alone, broadening the research to encompass non-linguists' perceptions of other voice features. As the DC in South Wales Police suggested, it may also be productive to ask further questions relating to voice features of the speaker, such as whether they had a high or low voice, or anything distinctive such as a speech impediment. And here we encounter the first and most fundamental problem with descriptions of physical speech features. Writing back in 1967, phonetician David Abercrombie noted that, at

the time, voice quality was one of the least investigated aspects of production (contrasted with, for example, segmental features) (Abercrombie 1967:91). And whilst progress has been made in the perceptual descriptions of voice quality over the last forty years, particularly by phoneticians, sociolinguists and forensic linguists, no substantive research has been conducted on how non-linguists perceive or describe the physical features of speech.

A starting point in the area of perceptual descriptions of physical speech features should be the view of the linguists and their technically informed depictions of voice quality. Laver (1991:187-188) makes an important distinction between the *phonetic quality* of voice i.e. the content, and the *voice quality*, i.e. the source. Rose (2002:277-280) elaborates on this distinction by noting that phonetic quality is related to segmental sounds like consonants and vowels, and also suprasegmental linguistic categories like intonation, tone, stress and rhythm. Voice quality, on the other hand is what we can hear when the phonetic quality is removed, as for example when someone can be heard speaking behind a door, but what they are actually saying is inaudible. According to Laver (1991), voice quality itself comprises two further aspects: an *organic* component, relating to the physiology of an individual speaker, and the *articulatory setting*, the quasi-permanent, habitual muscular settings that an individual adopts when they speak.

The controllable aspects of voice quality were described by Laver's (1980) phonetic descriptive framework. Designated by Nolan as 'the most comprehensive descriptive tool for the description of voice quality' (Nolan 2005:387), the framework provides some forty settings with which those trained in phonetics, may describe voice quality *settings*, such as 'falsetto', 'whisper', 'creak', 'harshness' and 'breathiness' and in combinations such as 'creaky whisper', or 'harsh breathy voice'. The framework enables us to distinguish settings and mobilise labels for their description and serves as a useful instrument in the phonetician's toolbox. However, familiarity with and proficiency in the application of this comprehensive framework lies within the realm of the phonetician, but is not, as Rose (2002:277-280) indicates, considered part of traditional auditory-phonetic training. Thus, the number of phoneticians

proficient in this approach is much smaller than those proficient in describing phonetic quality. Moreover, we must remember that the aim of this research is to move towards the design of an audiofit for use with non-linguists, and competence in Laver's framework is a matter of skilled training. Therefore, settings such as 'tense' and 'lax' for tension settings or 'mild palatalization' are likely to be beyond the grasp of earwitnesses. However, the distinctions between phonetic quality and voice quality have been noted and further reference to this distinction or articulatory settings will be made in this study, but will be reserved for the methodological discussion of the design of the study, rather than presented to any non-linguists in audiofit form.

Taking a step towards an arguably more accessible system for linguists and non-linguists, Hollien (1990) in his work in the field of forensic speaker identification, produced what he describes as a 'reasonably robust basis for aural-perceptual speaker identification' (Hollien 1990:61). His categorisation is, on the one hand, perhaps less detailed than Laver's approach, making fewer fine distinctions; however, it is more oriented towards forensic linguistic purposes, using distinctions that are designed to be tangible to wider audiences and presentable as evidence in court. Hollien does not distinguish features along the lines of phonetic or voice quality, but provides discursive categorisations for describing the *aspects of the speech signal* - voice and pronunciation:

i/. **Speaking fundamental frequency, or heard pitch**: general pitch level (high, medium, low) and its variability
ii/. **Articulation**: Any idiosyncratic pronunciation that sets the speaker apart from others
iii/. **General voice quality**
iv/. **Prosody**: timing, speed, rhythm. For example, how slow, or fast, a person talks or how smooth or choppy is his or her presentation.
v/. **Vocal intensity**: how loud a person speaks and how they vary their vocal intensity when they talk.

vi/. **Speech characteristics**: (1) dialect, (2) unusual rise of linguistic stress or affect, (3) idiosyncratic language patterns, (4) speech impediments and (5) idiosyncratic pronunciations.

(Hollien 1990:60-61)

Discussing his framework in the context of forensic speaker identification, Hollien reminds us that the listener may take into account all, or only some, of the six sets of attributes listed above (along with a number of subfactors) when they make a judgement, a design which is arguably more ecological in the forensic context. This is an important point to be borne in mind when working in the field of forensic speaker description. Any future audiofit design must allow for earwitnesses to describe as little as one component from the framework, with each piece of descriptive evidence being treated with its own integrity.

One of the most distinctive characteristics shared by both Laver's phonetician's approach to describing voice quality and Hollien's forensic speaker identification approach is their focus solely on objective measures. However, there is still the problem for speaker description, in that a high number of these objective categories may not be cognitively digestible for non-linguists or may be misinterpreted. For example, Niedzielski and Preston (1999:4-7) have noted that non-linguists use the term 'nasal' for inappropriate amounts of nasality on either end of the scale, whereas linguists use the similar term 'nasalised' for only the excessive end.

Given these limitations for non-technical respondents in the forensic context, we should give consideration to an alternative, smaller, less technical selection of descriptive features provided by Levin et al. (1994). In this study, a cohort of non-linguists were asked to rate the 'speakers' in a matched-guise format, reading English which contained lexis from Latinate versus Germanic origin, in an RP accent and a regional (south-east Welsh) UK accent. The rating categories suggested were:

Formal/informal speech?
Fancy words?
Simple grammar?
Trying to impress the listener?
Colloquial speech?
Long words?
Paying attention to their speech?
Unsure?
Relaxed?
Difficult to understand?
Trying to be friendly?
Easy to listen to?
Pleasant voice?
Was the listener distracted while listening?

In a second example of this type, Preston (1999:363) elicited respondents' mental maps of regional areas of the US but did not provide a set of descriptors. The most frequently mentioned labels provided by the non-linguists were:

Slow-Fast	*Educated - Uneducated*	*Formal - Casual*
Smart - Dumb	*Snobbish - Down-to-Earth*	*Polite - Rude*
Nasal - Not Nasal	*Normal - Abnormal*	*Friendly - Unfriendly*
Drawl - No Drawl	*Bad English - Good English*	*Twang - No Twang*

Preston then converted the labels into a six-point semantic differential task, affording the opportunity to perform a factorial analysis to investigate whether the evaluation samples could be reduced. What emerged were two robust factors:

1/. 'Standard': shows loadings from the categories that one associates with education and majority norms, e.g. formal, fast, snobbish
2/. 'Friendly': shows loadings from affective categories, including down-to-earth and casual

The opportunity to observe factorial loadings gives us some sense of the cognitive shape of non-linguistic comment, and factorial loadings will be considered later in the design of the methodology and subsequent analysis. Moreover, the descriptive categories offered by Levin et al. and the labels provided by the non-linguists in Preston's study are less technical than those provided by Laver and Hollien, and one can imagine that frameworks with less technical labels could be administered to respondents/earwitnesses with minimal additional explanation. The problem for this approach is that, in contrast

to Laver and Hollien's descriptors, those provided by Levin et al. and offered by Preston's non-linguist respondents, with the exception of Preston's 'slow-fast' and 'nasal-not nasal', refer almost entirely to subjective measures of voice, proffering value judgements, which, whilst interesting to explore, may not be consistently interpreted or mobilised by all speakers. It is not difficult to imagine that such subjective measures may not be considered sufficiently robust to be relied upon as part of the investigative process.

There is however, a compromise option that sits somewhere between the technical and the lay, between objective and subjective measures. The design, exemplified below, again resembles Laver's descriptive categorisations and does not explicitly distinguish between phonetic and voice quality. However, it does include both objective and subjective measures. Moreover, it arranges the descriptors as bipolar scales and the notion of applying contrasting perceptual semantic labels to voice features has a long tradition in the field of language attitudes studies. Van Bezooijen (2002) asked respondents to rate different dialects of Dutch, providing them with the following descriptors:

VOICE:		
Not husky	-	*husky*
Not creaky	-	*creaky*
Not nasal	-	*nasal*
Deep	-	*shrill*
Low	-	*high*
Soft	-	*loud*
PROSODY:		
Monotonous	–	*melodious*
Slow	–	*quick*
Calm	–	*agitated*
Legato	–	*staccato*
ARTICULATION:		
Slack	-	*precise*
Voiceless	-	*voiced*
Fronted	-	*backed*
Closed	-	*open*
EVALUATION:		
Ugly	-	*beautiful*
Repulsive	-	*attractive*
Exaggerated	-	*moderate*
Masculine	-	*feminine*
Rational	-	*emotional*
Serious	-	*funny*

Van Bezooijen (2002:25)

The list gives us an interesting and detailed range of semantic possibilities for perceptual voice description, including a focus on objective labels for different aspects of articulation, plus subjective labels relating to personal impression and value judgements. We can also see that some of the elements originally described by Laver (1980) resurface in van Bezooijen's suggestions – 'creakiness', 'pitch' and 'huskiness'. Also, the presentation of these descriptors in the form of bipolar semantic scales appears to have been a user-friendly and meaningful way of presenting the descriptors, with van Bezooijen reporting no problems with respondents grasping the format presented. However, with regard to the semantic referents of the descriptors, questions still arise regarding firstly, whether the descriptors such as 'legato-staccato', 'voiceless-voiced' and 'closed-open' are sufficiently transparent for a non-linguist audiofit; and secondly, whether the range of descriptors is comprehensive enough to cover all of the voice features that non-linguists may wish to remark on.

It is probably self-evident that descriptors devised by linguists, which serve as a useful orientation towards individual speech characteristics, are designed by those whose primary motivation is to technically analyse speech rather than describe it in layman's terms, and such terms are unlikely ever to be sufficiently meaningful to the non-linguist. The question for this research is how we bridge the gap between the assortment of categories that have been suggested by phoneticians and forensic linguists, and those categories that are salient in the minds of the non-linguist. The suggested line of investigation in the present study is to conduct new research on what non-linguists notice, how they conceptualise what they notice, using objective and subjective measures, and what vocabulary they choose to express what they notice. These data should then be interpreted within the context of the established phonetic, sociophonetic and forensic linguistic work. This is a theme that will be returned to in the design of the current research in Chapter 3: Method.

2.4 Attitudes

People can and do hold attitudes about just about anything (Smith and Mackie 2000:247), and attitudes can be a convenient way of explaining consistent patterns in behaviour (Baker 1992:11). Sections 2.4 will explore the nature of attitudes, how they are formed and conceptualised. Section 2.5 will then focus on how the findings of language attitudes have illuminated our understanding of how linguistic and social phenomena interact in the process of perceiving and describing voices and how they offer us access to individuals' patterns of belief, predictability of response, social stereotypes and prejudice.

2.4.1 Why do attitudes form?

> *When we talk about attitudes, we are talking about what a person has learned in the process of becoming a member of a family, a member of a group, and of society that makes him [sic] react to his social world in a consistent and characteristic way, instead of a transitory haphazard way.*
>
> *Sherif (1967:2), in Garrett et al. (2003:4)*

We develop attitudes because they are useful to us (Katz 1960). Attitudes help us to master our environment, assisting us in the negotiation of our interactions with *attitudes objects* in two ways: Firstly, the object appraisal function of attitudes helps us organise and simplify our experience, orientating us towards the important characteristics of an object so we can deal with it efficiently. The second is the utilitarian or instrumental function, steering us towards objects that will help maximise our rewards and reach desired goals while avoiding undesirable situations or events (Smith and Mackie 2000:250).

Attitudes have further dimensions, also. They help people express their real selves, voice their convictions, show what they stand for and affirm the significant relationships. The social identity or value-expressive function of attitudes helps people gain and maintain connectedness with others. Attitudes on environmental, social or religious issues help us to define who we are.

2.4.2 What is an attitude?

By their very nature, attitudes cannot be directly observed. A person's thoughts, processing system and feelings are hidden. As Allport comments:

> *Attitudes are never directly observed, but, unless they are admitted, through inference, as real and substantial ingredients in human nature, it becomes impossible to account satisfactorily either for the consistency of an individual's behaviour or for the stability of any society.*
>
> *(Allport 1935:839)*

Attitudes therefore are latent and can only be inferred from the direction and persistence of external behaviour (Baker 1992:11). But how do we define something that we cannot directly observe? This can be done by making a separation between attitude and other related terms, such as *belief* and *attitude* (Edwards 1982). For example, a subject's response to, "Is a knowledge of French important for your children, yes or no?" indicates a belief. To gauge an attitude would require further enquiry into the respondent's *feeling* about this expressed belief. Garrett et al. (2003:10) cite Oskamp (1977) in distinguishing between *values* and *attitudes*: Values are 'the most important and central elements in a person's belief system of attitudes and beliefs', and are more global and general than attitudes.

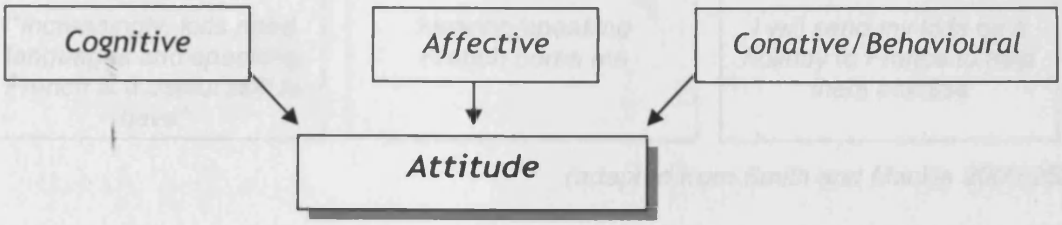
Baker (1992:14) contrasts *opinions* and *attitudes*, which he judges to be used synonymously in everyday speech. Baker observes that the difference between attitude surveys and opinion polls is that opinion polls tend to locate community or group preferences and wishes, providing indicators of population viewpoints, for example, who respondents intend to vote for in an election; whereas, attitude surveys focus more on the relationship of attitudes to a variety of other variables seeking to understand human functioning; for example, how attitudes in the United Kingdom towards a United States ruled by George Bush may influence the choice of holiday destination – should I go to Disneyland, USA or avoid the US, stay in Europe and go to Eurodisney?

Attitudes are commonly defined in terms of an evaluation of the relationship between a particular entity and attitude object. The attitude objects under evaluation are composed of complex social phenomena, interpreted

subjectively by the onlooker, e.g. the self, other people, things, actions, events or ideas (Smith and Mackie 2000); Attitudes then, are viewed as a cognitive representation or summary of our evaluation of an attitude object, expressing some degree of favour or disfavour (Eagly and Chaiken 1998).

The thrust of many definitions including those above is towards a multi-dimensional view of attitudes. This is offered through the commonly described *tripartite* structure of attitudes (e.g., Edwards (1982) and Rosenberg and Hovland (1960)).

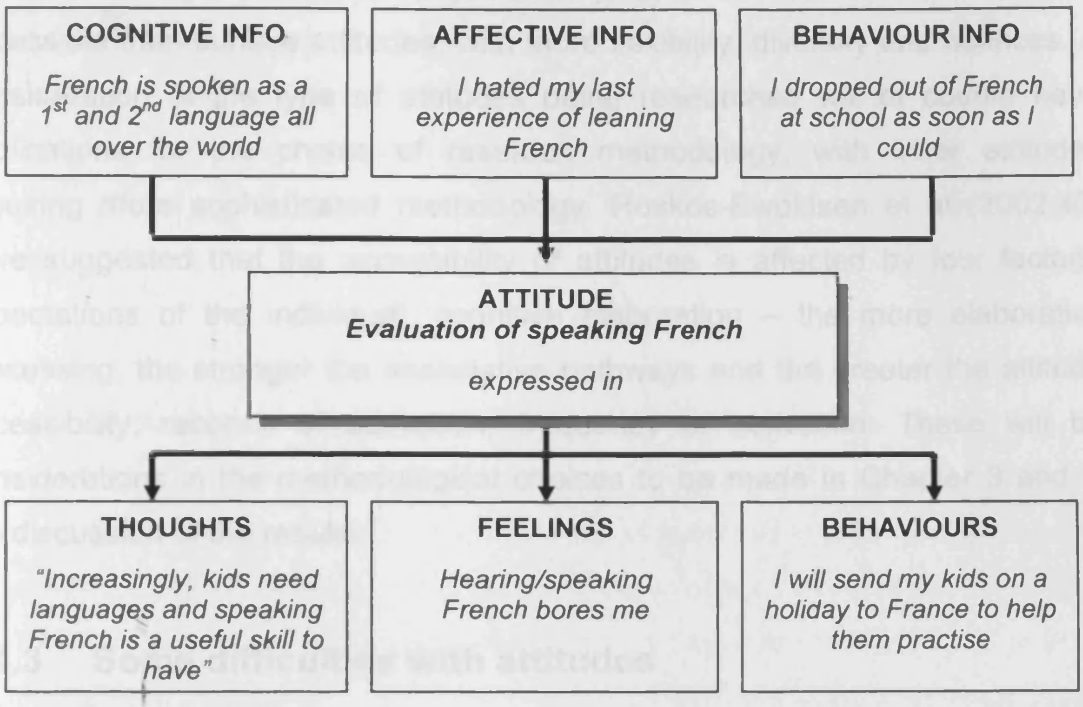
Figure 2.4.2.1 Tripartite structure of attitudes



In this model, attitudes are seen as *cognitive* because they entail thoughts and beliefs about the world; they are *affective* because they involve feelings towards an attitude object; they are *conative* or *behavioural* because they encourage certain action (Cargile et al. 1994). For many, exclusion of any of these parts does not equate to research into attitudes. For example, Edwards (1999:109) is critical that many "language attitudes" studies are in fact studies of belief.

To illustrate this point, we can refer back to the earlier example regarding attitudes towards French; it is possible that one might believe French to be of importance, while heartily disliking the language and its speakers. An added *affective* dimension is required in such questions for a fuller *attitudinal* evaluation. We can exemplify the value of a tripartite model of attitudes and how it might manifest itself by expanding Edwards' (1999) comments onto an example of attitude to French:

Figure 2.4.2.2 Illustration of tripartite structure of attitudes: attitude to French



(adapted from Smith and Mackie 2000:252)

Using a tripartite model allows us to view how different pieces of information accumulate and how they do not always imply a consistent evaluation of an attitude object (Smith and Mackie, 2000:252). Many people will hold such *ambivalent* attitudes (Katz 1981) as the one illustrated above, with their attitudes having many facets, including both positive and negative reactions to attitude objects. It would be unwise, then, for researchers to infer from intended future behaviour alone that a fully positive attitude is held towards the object of learning French. That is, without an examination of the affective and cognitive modes, a researcher could seriously misunderstand the shape of the attitude. A similar example might be that a speaker has negative feelings towards the social group they identify with the speaking of RP, but they themselves modify their accent towards RP for instrumental reasons, borrowing perceived prestige in order to be seen as educated and intelligent in the work place.

Furthermore, if we think of attitudes multi-dimensionally, we should also acknowledge that some attitudes are more easily accessed than others. Stalans (2002:18) points to the view from cognitive psychology regarding people as having at least two 'layers' of attitudes: firstly, *surface* attitudes, which will have been discussed more and are more likely to be aligned with widely known

political or religious values, and secondly, *inner* attitudes, which are less accessible than surface attitudes, with more flexibility, diversity and nuances. A consideration of the type of attitudes being researched will of course have implications for the choice of research methodology, with inner attitudes requiring more sophisticated methodology. Roskos-Ewoldsen et al. (2002:46) have suggested that the accessibility of attitudes is affected by four factors: expectations of the individual; cognitive elaboration – the more elaborative processing, the stronger the associative pathways and the greater the attitude accessibility; recency of activation; frequency of activation. These will be considerations in the methodological choices to be made in Chapter 3 and in the discussion of the results.

2.4.3 Some difficulties with attitudes

The previous section illustrated a popular tri-partite *internal* structure to attitudes, although this model has its critics (Potter and Wetherell 1987). Nevertheless, the point being made by drawing on a tri-partite model is that it is important to resist the temptation to pigeon-hole 'individual' attitudes for ease of clarification, comprehension and generalisability, and to see attitudes as more than isolated single fixed entities, static singular phenomena, organised in our brains like individual files in a filing cabinet (Stalans 2002:17). Attitudes are connected to images, beliefs, emotions and experience and as such are embedded in other attitudes and stored in the memory in associative networks (Fazio 1989), affecting and being affected by numerous elements in a virtually endless, recursive fashion (Cargile et al. 1994:215). To see attitudes as individual entities is to ignore the complexity of human experience.

Much as it can be tempting to see attitudes in isolation, so too can it be tempting to neglect to acknowledge the changeable nature of attitudes. Attitudes may be stable and enduring but also variable and short-lived (Stalans 2002). Garrett et al. (2003) for example, identify research from the persuasion literature which points to differing levels of commitment in attitudes (Sears and Kosterman 1994); more recently acquired attitudes being less stable compared with, for example, those acquired early in the lifespan (Sears 1983). These

observations would seem to imply that dwelling, for example, on the 'strongest' evaluation response in order to illustrate a viewpoint, could compromise any subsequent research claims. The strength of the attitude may simply be down to the day's newspaper headlines, interpersonal dynamics or simply the respondent being in a good or bad mood. Follow-up data would be required to check for stability.

Other responses may simply not be indicators of real attitudes. For example, an interviewer having to conduct research on a fairly obscure point may be impressed by the strength or forthrightness of a response, inferring that the respondent is aware of the subject matter and may even have some personal investment in it. Here again, caution must be exercised. Ostrom et al. (1994) illustrate this by discussing responses to the "Metallic Metals Act", a non-existent piece of proposed legislation that featured prominently in an experiment conducted in 1947 by Sam Gill in Tide magazine. Gill asked respondents the following question:

Which of the following statements most closely coincides with your opinion of the Metallic Metals Act?

- It would be a good move on the part of the US.
- It would be a good thing, but should be left to the individual states
- It's all right for foreign countries, but should not be required here.
- It is of no value at all

Of those asked, 70% expressed an opinion despite the fact that no such act existed and, therefore, the respondents could have no actual knowledge. The responses (for those 70%) were:

- It would be a good move on the part of the US. (21.4%)
- It would be a good thing, but should be left to the individual states (58.6%)
- It's all right for foreign countries, but should not be required here. (15.7%)
- It is of no value at all (4.3%)

Ostrom et al. point out that some people who have no response on an issue pretend to have an opinion, concocting a response on the spot, based on little

or no information about the issue. Their evaluations are *non-attitudes* and do not reflect a pre-existing view the respondent held on the subject. They may do this, because they feel some pressure to portray themselves as thoughtful and knowledgeable or as the result of social desirability or acquiescence biases. However, their responses can be purely random choices, and they would probably give a completely different answer were they to be asked the same question a few weeks later (Ostrom et al. 1994:28).

A knowledge of the potential for respondents to provide a simple psychological reflex as an evaluation should inform all attitudes research designs, and attitudes researchers are generally aware of the ebb and flow, environmental susceptibility and general 'flakiness' of attitudes, anticipating and differentiating between the range and strength of responses that may be elicited, and are wary of generalising claims about individual attitudes to a wider population, a danger that can be encountered by using single representative tokens of speakers or accents as attitudes objects. All of these will be considered in the Discussion chapter (6) of this research.

One final point to make is that at no point is the current research engaged with persuading or effecting attitude change, or coercing respondents to provide or alter their statements. This study in forensic speaker description is limited to the observational rather than investigating ways of influencing the attitudes of those respondents encountered.

2.5 Attitudes and language

So far, this chapter has mapped the general territory around attitudes. Forensic speaker description specifically involves the observation of responses to linguistic stimuli and so the next step is to explore the issues as they relate specifically to *language* attitudes and issues that may impact upon this research. The next sections will exemplify language attitudes in the public domain and then detail some of the major findings of language attitudes works over the last thirty five years.

2.5.1 Language attitudes in the public domain

As we saw in the last section, the study of attitudes affords the researcher a unique view on human interaction and reasoning, and on the issues of the day. A casual journey through the British media can illustrate some of the points that make it into popular discourse. Below is a selection of stories collected from two of the most popular news websites in the UK, the BBC and The Guardian. They illustrate attitudes towards written language, spoken language, non-native English speakers, bilingualism, native English speakers, and borrowing foreign words. This is by no means an exhaustive list and it is not difficult to find stories reporting on non-linguist attitudes towards language in the public domain, as is illustrated in Figure 2.5.

Figure 2.5 Examples of language attitudes in the public domain

'English crucial' for immigrants

More than half the members of ethnic minorities resident in Britain believe those settling in the UK should learn to speak English, a survey suggests.

More than 90% of 3,000 questioned said those arriving in Britain should not expect special treatment.[...]

A significant minority of ethnic minority respondents said they had a low opinion of asylum seekers and other recent immigrants, as did 50% of white people taking part in the survey.

http://news.bbc.co.uk/1/hi/uk/3648768.stm

Springer protests pour in to BBC

The BBC has received more than 15,000 complaints about its decision to air the musical Jerry Springer - The Opera.

Campaigners Mediawatch UK have railed against the show going out on BBC Two on Saturday because of the high level of swearing.

http://news.bbc.co.uk/1/hi/entertainment/tv_and_radio/4152433.stm

Texting 'is no bar to literacy'

It's gr8 news 4 skools. Claims that the explosion in text messaging among children is eroding youngsters' literacy skills appear to be unfounded, according to research.

A study comparing the punctuation and spelling of 11- and 12-year-olds who use mobile phone text messaging with another group of non-texters conducting the same written tests found no significant differences between the two.

http://education.guardian.co.uk/schools/story/0,5500,1378951,00.html

Watching their language

Secondary schools are to receive guidance on how to teach children "standard" spoken English.

The intervention by the Qualifications and Curriculum Authority (QCA) comes amid concern over the effects of the media and text messaging on the quality of teenagers' conversational skills.

http://news.bbc.co.uk/1/hi/education/3551406.stm

Students offered language lessons

Record numbers of first-year students will be offered crash-courses in basic English and grammar when they start their university courses.

Nearly half of Scotland's universities have been forced to provide remedial fast-track classes because of plunging literacy levels in schools.

They claim undergraduates are not ready for the linguistic and literary demands of degrees after leaving school. Some students could not write, spell or punctuate simple sentences.

http://news.bbc.co.uk/1/hi/scotland/3692092.stm

Welsh proud of 'unpopular' accent

Welsh accents are among the least popular in the UK, according to a BBC poll.

Many who took part in the survey believed that having a Welsh accent could hinder a career. [...]

In Wales, 56% of respondents said they were proud of their accents, more than the figure for the UK as a whole.

http://news.bbc.co.uk/1/hi/wales/4179629.stm

Language impurities and how to enjoy them

More diktats from cultural police has Luke Meddings arguing for a more playful, even messy approach to teaching English

A language brouhaha has appeared over the channel, where French cultural mandarins have slammed the use of English names for imported television programme formats such as Star Academy (currently rendered as 'La Star Academy') and Popstars (rendered as, erm, 'Popstars').

http://education.guardian.co.uk/tefl/comment/story/0,,1425287,00.html

The selection above illustrates just a few of the thousands of news stories that touch upon language attitudes in some way or other. Issues such as:

i/. Written language: Standards of spelling and articulation ('*Students offered language lessons*'); the role of teaching of standard English versus dialect forms and regional variations; the effect of mobile phone text messaging on the language and its users ('*Texting "is no bar to literacy*"')

ii/. *Spoken language*: Inter-generational changes/variations; 'standards'; usage in the media (on radio and television) ('*Watching their language*'); swearing and blaspheming ('*Springer protests pour in to BBC*'); politeness; pronunciation and notions of correctness ('*Watching their language*'); laziness; why people 'persist' on using regional accents ('*Welsh proud of 'unpopular' accent*')

iii/. Non-native English speakers living in the UK ('*English "crucial" for immigrants*'); teaching and learning English

iv/. Native English speakers learning foreign languages

v/. Language learning in the classroom

vi/. Bilingualism: bilingual education; bilingualism in government and public spaces; bilingualism in the media; language rights and planning

vii/. The borrowing of 'foreign' words: In the UK, the borrowing of American vocabulary, and what is perceived as the Australian rising intonation at the end of sentences; outside the U.K. and native English speaking communities, the borrowing of English vocabulary into another language ('*Language impurities and how to enjoy them*')

Other examples of language ideology, stereotypes and myths are addressed by Bauer and Trudgill's (1998) *Language Myths* and Andersson and Trudgill's (1990) *Bad Language*. It is interesting to note that many attitudes portrayed in the media would appear to reflect a broader opinion that there is an inexorable decline in language standards, as appears to be the case in the stories of '*Students offered language lessons*' and '*Watching their language*', or the fears

that form the backdrop to '*Texting "is no bar to literacy"*. This echoes Cameron's (1995) observation that conformity to the laws of grammar signifies for many people conformity to the laws of society, and disregarding the one leads to disrespect for the other. Milroy (1998) notes that such a view of the decline in moral and linguistic standards was exclaimed by Jonathan Swift as far back as 1712, decrying the 'Licentiousness which entered with the Restoration [1660]' that infected our morals and then corrupted our language.

Of course, this is not to suggest that it is 'wrong' to hold attitudes towards language or any other attitude object and indeed, one may argue that it is impossible to be a language user *without* holding a set of attitudes based on beliefs about language. However, there is no doubt that very few members of the general public or the media are acquainted with the concept of a 'language attitude'; there is scant evidence of an awareness in the media, for example, that the views that abound in the public arena should be qualified as expressions of a language attitude, rather than statements of *linguistic fact*. Does this virtually ubiquitous, poor grounding in linguistics matter? Compared with the buffet of everyday educational concerns portrayed in the media such as poor standards, failure, truancy and dropout rates, the inability of the average British citizen to tell an isophone from an isomorph, or an accent from a dialect does not seem to carry the same gravity. There are, however, plenty of examples of the media peddling as linguistic fact what may be termed by professional linguists to be 'linguistic urban myths'. A few examples of this are alluded to above and in the context of forensic linguistics, there is research to substantiate that language ignorance and prejudice does matter, as it has serious implications for social justice, as was discussed in Section 2.3.1, Voice Evidence (Allport 1935; Lind and O'Barr 1979; Seggie 1983; Dixon et al. 1994; Dixon et al. 2002; Dixon and Mahoney 2004).

The burden has lain with linguists and social psychologists to raise language myths as issues and devise and conduct appropriate research into them, and over the past thirty five years in particular, a significant body of research has addressed many and more of the language myths we see above, contributing to the distinction between language myth and language reality. The next section

will set out some of these findings as they pertain to forensic speaker description.

2.5.2 'Good English', 'Correct English' and RP

Professional linguists (especially in the UK) will be aware that for non-linguists, there is an almost global assumption that there is a 'correct' English; an English from England that should be aspired to at all times; an uncorrupted English that has a correct and unvarying structure and vocabulary, has connotations of standards, politeness, good education and social status, and is inherently 'right'. The non-linguist perception of 'better' English is central to any research on language attitudes and is likely to be reflected in any descriptions of speakers they provide, and evidence of this belief abounds in the media and public discourse, and has also materialised in language attitudes research over the last thirty five years.

A number of studies (Giles 1970; Trudgill and Giles 1978; Garrett et al. 2003; Garrett et al. 2004; Bishop et al. 2005; Coupland and Bishop 2007) have shown that within an intra-cultural setting, linguistic varieties of English are evaluated differently in terms of aesthetic, intellectual and social value. Generally, the evaluations of non-linguists place RP and/or Standard English at the top, rural accents in the middle, and urban accents at the bottom. RP or standard English or Queen's English (Bishop et al. 2005) is, it seems, a highly exalted speech variety amongst non-linguists.

This view of a 'correct' English is manifest in many examples of public and media discourse and one simply has to scratch the surface to encounter views reflecting the assertion that those who use regional accents or dialects, or who use slang and/or other forms of 'incorrect' English do so out of poor education/educational achievement or as a result of simple laziness, feebleness or sloth. Moreover, those who persist in using alternatives to 'correct' English – or other languages within the British context, (e.g. Welsh in Wales, Arabic in West Yorkshire, etc) – do so out of political or social intransigence,

demonstrating a deliberate intention to create or maintain linguistic and cultural ghettos and reject English-speaking culture.

Giles et al. (1979) identified the apparent anomaly between the opinion of the masses (non-linguists), who tend to evaluate varieties and their variants as 'good/correct', 'bad/incorrect', and the minority (professional linguists), who view language varieties in terms of their functional equivalence. Giles and Powesland (1975) devised an empirical study that would scientifically investigate how varieties attained their good/bad status amongst non-linguists. They tested two explanations: the *'inherent value' hypothesis* - some language varieties are intrinsically more pleasant to listen to than others, an argument that matched match folk belief; alternatively, the *'imposed norm' hypothesis* - the standard variety has gained acceptance as a result of cultural norms, and because of social pressure to emulate the standard, it has come to be regarded as superior on many dimensions (van Bezooijen 1994). Here was a clear opportunity to test the non-linguist assertion that there is one correct and superior way to speak English, or any other language.

Giles and Powesland's conclusions point away from any credibility for the inherent value hypothesis. What was clear to linguists was that the judgements made by most people (i.e. non-linguists) are largely arbitrary and based on social rather than linguistic criteria (Davies 2007:136). Welsh and French Canadian respondents, when played examples of Quebec French (Giles et al. 1979), supplied very different responses, with those from outside the language community – the Welsh respondents – unable to differentiate the varieties on measures of prestige and aesthetic qualities. Evaluations of varieties of languages by those who are able to correctly identify them – the Quebec French speakers - were based upon social connotations.

Despite the research findings of linguists, thirty five years later the non-linguist view of good/bad English in the UK appears unchanged: In their study designed to contrast findings from their research with those of Giles' (1970) influential study on social evaluation of some of the major English accents relevant to the UK, Bishop et al. (2005) describe their results as showing a 'remarkable

similarity to those of Giles.' (Bishop et al. 2005:131). Bishop et al. point out that the favouring and disfavouring of language varieties appears to be a stable part of the conceptualisations of language variation. They also note that the absence of significant shifts in attitudes over thirty five years is, from the progressive libertarian stance, 'disappointing'. It appears that little headway has been made in dispelling language myths and illuminating non-linguists' murky understanding of how language works. For this reason, when planning research centred on non-linguists' perceptions and representations of unknown voices, it is vital to be aware of the factors that might mediate non-linguists' responses. Moreover, linking non-linguists' views, beliefs and attitudes to the forensic context affords the researcher the opportunity to investigate how decisions such as the attribution of guilt may be influenced by non-linguists' own prejudicial attitudes towards language varieties, as was discussed above in Section 2.3.1.

2.6 Planning research on forensic speaker description: methodological choices

Research into forensic speaker description inevitably asks questions around how non-linguists view, perceive and evaluate different accents and voices. Whilst research into forensic speaker description is novel, research in the field of language attitudes over the last thirty five years has much to offer the development of a new methodology. Language attitudes research has a rich history of designing research experiments to investigate community-level responses to systemic language variation within a given community. The basic paradigm is one whereby a cohort of non-linguist judges directly evaluates or indirectly responds to a sample of attitude objects in the form of recorded or live voice samples, or a non-vocal mode, employing a sample of geo-social labels (Williams et al. 1996). Thus, language attitudes research has advanced our understanding of the attitudes, attendant beliefs and values held by non-linguists around the world; be it through responses to direct questioning or indirect responses to requests made apparently by different voices with differing accents. Exploring how people evaluate accents and respond to voices has evolved into a research aim in itself.

However, as extensive as the research has been, one can nevertheless identify methodological gaps which have yet to be addressed. Over the last thirty five years or more, there has been a preponderance of research, especially in the *elicitation* of attitudes, conducted in a largely abstract research environment in which the presence of the real-world contexts has often been sacrificed for the sake of the aims of the research. It appears that in the quest to isolate the accent judgement from other forms of attitudinal judgement, attention has often been focused on ensuring the homogeneity or range of the cohort of judges, but less emphasis is placed on the interactional context of the attitude object or the relationship of the judges to the attitude object or the mode within which they encounter the stimulus.

Like the surgeon who wishes to understand the structure of an organ, language attitudes researchers have tended to remove the attitude from the body of real-world context, and laid it on the slab. And by doing so, our understanding of the structure and direction of language attitudes has increased. But what happens when language attitudes research is not conducted acontextually? We know, for example, that studies into language attitudes and attributions of guilt (Lind and O'Barr 1979; Seggie 1983; Dixon et al. 1994; Dixon et al. 2002; Dixon and Mahoney 2004), have demonstrated that attitudes are subject to change when judges are supplied with differing contextual (legal) information, and there appears to be a strong case for advocating a more applied direction in language attitudes research, ensuring that the sample is provided to judges within a social context in real-world fields. After all, if the samples are not presented within a social context, what social context are they generalisable to?

One such field is forensic speaker description, where abstraction from context serves no purpose, but explicit research within a given 'criminal' context assists the aim of ecologically valid outcomes. One of the major aims of this research into forensic speaker descriptions is to utilise language attitudes methodology in an applied context. The research seeks to explore the benefits of, to return to the analogy, replacing the tissue into the organism to see how it functions in context, informing the understanding, and possibly practice, of all involved in the legal context. Ladegaard echoes this sentiment:

> *Stereotypes used about language are inseparable from stereotypes used about people and therefore, we should be careful not to rely exclusively on an approach to language which presupposes the separation of speaker from situation and context, and reduces the speakers to automatons who deliver a specific linguistic product depending on fixed social variables such as gender, social class, region and age. The real world in which our subjects live is much more complex and dynamic and will inevitably be perceived differently by different people. Therefore we need an approach to language which will take this complexity into account.*
>
> *(Ladegaard 2001:36)*

Chapter 2 continues, therefore, by setting out considerations for language attitudes research in order to inform the design of new research into forensic speaker description: how do non-linguists perceive and describe the voices and accents that they hear and how are their perceptions and descriptions mediated by their own evaluations of the individual accents and voice characteristics that they encounter, by their commonly-held beliefs about language, by the researcher's selection of attitudes objects, by the selection of respondents, by the selection of donor voices, by the instrument used for eliciting evaluations and descriptions, and by methodological considerations? In setting out the methodological consideration for research into forensic speaker description, two marked terms are used, *system variables* and *estimator variables* (Broeders 1996). System variables are those that are under the control of the researcher or forensic linguist, such as the line up of voice samples, how they are administered to earwitnesses, etc. Estimator variables are those which relate to the validity of earwitness performance, such as, witness familiarity with accents, accuracy, etc. The following sections serve as a form of checklist of issues, variables and choices that anyone wishing to conduct research into forensic speaker description may like to consider.

2.7 System variables: The selection of language attitudes objects

All experimental designs in language attitudes research entail the selection of the object(s), to which we wish to gauge the attitude. The selection of attitude objects and research design in language attitudes work over the last thirty five years has followed a distinct theme, derived largely from Giles' (1970) influential study. Giles chose a between-speakers design, contrasting English accents

relevant to the UK context, which he presented to a cohort of young British people for evaluation.

Regional English-speaking accents	International accents:	Non-regionally identified accents:
Cockney	French	Indian
Birmingham	Italian	W. Indies
Somerset	German	R. P.
Yorkshire	N. American	Affected RP
Liverpool		
Scottish		
S. Wales		
Irish (It is not clear if the speaker was from Northern Ireland or Eire)		

(After Giles 1970: 215)

Among Giles' aims was the goal of empirically and quantitatively documenting for the first time contrasting patterns of evaluation that speech varieties in the UK attract, establishing that the accent you use in the UK may affect the way that others evaluate you. Giles' research design has been highly influential over the years and a rich tradition of language attitudes research has built up in its wake, replicating Giles' research design. However, a critical evaluation of the research design and its emulation throws up a number of problematic issues:

(i) Equality of attitude object: Simple comparisons between area and population reveal that Giles' selection of accents for evaluation had little to do with representing areas of equal size, demography or status, Giles chose, for example:

London, capital city	population: approx 8,000,000
	area: 617 square miles
Scotland, country	population: approx 5,000,000
	area: 30,420 square miles
Liverpool, regional city	population: approx 450,000
	area: 68 square miles

Giles does not state that the aim was to elicit attitudes towards accents on a basis of national, regional or social equivalence; the aim was neither

based on symmetry of geography, population size or concept and it is likely that the fundamental goal of demonstrating the contrasting patterns of evaluation of varieties of English took priority. It may be that to non-linguists, a label is a label, regardless of the size of community it is attached to. However, the point remains that in Giles' and a raft of subsequent studies, equality of attitude objects does not appear to be a consideration. Language attitudes researchers over the last thirty five years have tended to present a between-speakers design, vocally or conceptually, utilising a smörgåsbord of higher profile, regional and social speech varieties and stereotypes in order that lay people would be certain to recognise and evaluate them. Language attitudes researchers have not yet presented arguments that all labels are labels regardless of the size of the community they represent. In order, presumably, to yield more colourful and vivid results, researchers have undoubtedly been conscious that the deeper the contrast between the accents, the clearer the projection of the attitudes and the greater the claim for methodological validity. But if no measures have been taken to select the attitude objects according to any particular geographically or socially equivalent criteria, how can we be sure that the accents played or suggested to the respondents are cognitively represented at the same level? It may be that judges are responding simultaneously to speakers defined by a local term, a regional term or a national term, a stereotype of a medium-sized city or a somewhat more diffuse stereotype of a five-million-inhabitant country, with the implication that the researcher is not comparing like with like.

A few exceptions exist where either a geographical approach focusing on one complete area has been used, for example, the Wales-wide research conducted by Garrett et al.(2004), or research on dialect change in lesser-known towns such as Kerswill and Williams' (2002) investigation of dialect levelling in Reading and Hull. However, for the majority of language attitudes research, unless the researcher has acknowledged the disparity of attitudes objects as part of the intended design, there are implications for the interpretation of the data in terms of its ecological validity.

(ii) **Researcher-centric conceptualisations of language variation:** As discussed, above, there appears overall to be a tendency throughout the history of language attitudes research to include varieties that are adjudged by the researcher to be in contrast, a decision taken from a technical, informed helicopter view of language variation across the whole community/nation. The selection of contrasting varieties has been a matter of researcher judgement based on which varieties the researcher believes are socially or geographically marked and the majority of respondents will be able to recognise and index. But how can we be sure that all the phonological characteristics present in the researcher's choice of voice sample have the same salience to individual respondents? With the exception, perhaps, of Dennis Preston's work in folk linguistics, virtually no research has been conducted where the selection of attitudes objects has been based on contrasting language varieties that the folk have first identified, and little research has examined language variation from the local community level outwards, allowing the folk to identify what linguistic juxtapositions are salient to them. The taxonomy that has dominated previous research, conceptualising, framing and presenting researcher-selected, high-contrast varieties to respondents, has perpetuated the presumption that respondents too view their speech communities as either regional or posh. This may be the case but it should be substantiated through research that includes investigation of the non-linguists' conceptualisations of language variation.

(iii) **The respondent's estimation of accent provenance:** As has been mentioned on page 39, Preston (1996) observed that few examples of language attitudes research have asked respondents where they *think* a voice being played to them is from. Without asking, how do we know that the judges identified the origin of the voices correctly, and that their responses are based on evaluations of the researcher's intended variety? The judges may be responding to their perception of where the speaker is from, which may differ from the researcher's intended attitude object. Without confirming where the judges perceive the speaker to be from, the researcher may be unaware of the possible invalidity of later claims.

Turning more specifically to the accents selected as attitude objects, in the British context, one accent has been conspicuously absent from language attitudes work: the frequently-used folk concept of 'no accent'. In the British context, 'no accent' is a description that is likely to describe someone who (probably through years of English-based political and media control) sounds 'normal' to many southern English, 'southern' to those from the north of England and 'just English' to Welsh, Scottish, and Irish speakers. Outside these descriptive parameters, 'no accent' is unlikely to be sharply geographically or socially marked and unlikely to attract the same strong stereotyped reactions as other accents commonly found in the south of England such as Estuary English, West Country or RP. Alternatively, 'no accent' could also be used by some non-linguists who do not regard themselves or their community as having a marked regional or social accent to mean 'an accent similar to mine'. Niedzielski (1999) in her perceptual study of accents variably labelled as from Detroit or Canada found that most Detroit residents do not recognise the distinctiveness of their own speech and hence harbour preconceived notions that their speech is unmarked for dialect features (Thomas 2002:124). Both 'a standard accent of English' and 'an accent similar to my own' both receive similarly strong positive evaluations in language attitude studies, see for example Coupland and Bishop (2007) and non-linguists' understanding of the term 'no accent' warrants further investigation.

Although the rather linguist-oriented term 'standard English' and the more user-friendly 'an accent similar to my own' have emerged in later work, such as Bishop et al. (2005), most studies have steered away from investigating the non-linguist concept of 'no accent'. Linguists recognise that the term 'no accent' is, of course, a misnomer, for the very act of producing comprehensible speech entails the use of a meaningful pronunciation system and it is clearly not possible to speak without an accent. Yet 'no accent' is a concept that is readily provided by non-linguists. Its absence from the research most probably reflects the fact that linguists are aware of the complex relativity of the term and that to offer the term 'no accent' as a response category in fieldwork will entail further, possibly qualitative, investigation to establish what the term means to individual non-linguists, which may be impossible within practical research constraints.

The current situation in the UK is that despite years of research on non-linguist evaluations, we know far less about 'no accent' than probably any other regional or social accent in the UK; we know very little about the possible intended referent of the term 'no accent' and we know little about how respondents describe and feel towards the 'no accent' voices whose provenance they are unable to estimate. Regardless of the rather nebulous and complex relativity of the term, it seems inconceivable that research on non-linguist audiofits can proceed without acknowledging the existence and prevalence of the term in non-linguist conceptualisations and in their lexicon or investigating how, when and why the term is used. If we do not know how people feel about or describe an accent, how can we rely on forensic voice descriptions provided by the general public? It is essential to investigate how we establish community authenticity in the description of accents, in order to be certain what it is that the non-linguists believe when they provide their descriptions and evaluations. 'No accent' should therefore be considered for inclusion as one of the attitude objects under investigation.

2.8 System variables: Indirect and direct measures

The choice of research instrument in this field will be dictated by whether the researcher chooses to employ a direct or indirect approach. The two approaches can be characterised thus:

2.8.1 Indirect measures

The primary research tool for indirect methods is the matched guise technique (MGT) (Lambert et al. 1960). The use of MGT is an attempt to control linguistic factors by use of, for example, audio recordings of one bidialectal/bilingual speaker reading aloud a prepared text in both available dialects or languages – two guises. This creates the illusion to the listeners that more than one person is present (for example Lambert et al.'s use of a bilingual speaker of both Canadian French and English). Responses and speaker evaluations are then considered to reflect the listeners' underlying attitudes, eliciting responses that are relatively low in stereotypy (Cargile et al. 1994:214) and high in spontaneity

and sincerity. That is, responses are less likely to be artefacts of response biases (discussed in Section 2.8.4, below), and more likely to reflect respondents' private views.

Generally speaking, MGT has been used where it is anticipated that it will be very difficult to get honest data from people who are aware of what is being researched. This is particularly relevant in communities where the language issue may be one of high sensitivity, for example, French-English in Quebec (Lambert et al.1960) or Welsh-English in Cardiff (Bourhis and Giles 1976), where the use of MGT may be an effective way of concealing the real purpose of the research.

The MGT has attracted criticisms:

i/. Neutral texts: Giles and Coupland (1991) have argued that there will always be textual and contextual connotations, and that the texts themselves can never be neutral as there will always be a symbiotic relationship between language varieties and language attitudes in interpretive processes.

ii/. Evaluating the speaker/the speech/the stereotype: The use of decontextualised recordings, with repeated messages focusing the respondents' awareness on the differences in accent exclusively may encourage the use of stereotypes that may not figure as a part of the respondents' every day response to language variation. Edwards (1982:22) argues that it is an evaluation of the representative stereotype of a speech variety that is being provided rather than a language attitude. If this is the case, researchers must be cautious with regards to the claims of validity they make with research based around voice samples or conceptual labels.

iii/. Respondents' areal taxonomy: As has been mentioned previously, Preston (1996) has argued that there has been a large number of examples of language attitudes research where those whose opinions are being sought (rather than any pilot group) are not asked to describe where they believe the speakers to be from. MGT may exacerbate this problem by its inherent need as an indirect

technique to work 'secretively' and without explicitly asking questions as to the respondents' areal taxonomy.

iv/. Authenticity: The practice of speakers providing voice samples in two or more dialects, e.g. Seggie (1983) using RP, Australian and 'Malay Chinese' in an MGT, raises the question as to the authenticity of the speaker's accents. How many speakers are there, for example, who are genuinely, symmetrically bi or tri dialectal/lingual, in for example, Australian English, RP and 'Asian' accents? Van Bezooijen and Ytsma (1999) have suggested that using a matched guise technique, one accent may realistically be compared only to one other as it is virtually impossible to find speakers sufficiently fluent in more than two varieties. The question arises, therefore, whether in the case of three differently accented speakers being used, the voice donors are examples of natural speakers of all the selected dialects/languages or they are providing impersonations of the main characteristics of each dialect/language, renderings or caricatures that are hopefully believable to their selected local folk (in this case, Australian) audience. If the aim is a conceptual one, establishing how speakers from a certain regional, social or ethnic group are evaluated, we might argue that the recorded voice is redundant and that conceptual labels could be used in its place. If the aim is to evaluate the performing of an accent, then any caricatured features present in the individual speaker's performance of the voice may mediate the evaluations of the voices, compromising any claims for generalisability of these evaluations to the wider social group. Garrett et al. (2003) add the following summary comments on problems with the matched-guised technique:

v/. Community authenticity: the labels used for the speech varieties in published studies are sometimes too vague to be meaningful. For example, 'Welsh English'. Garrett et al. (2003:74) have complained that such labels do not objectively capture the varying degrees and qualities of varieties of, for example, English in Wales. They do not represent the structuring of rural and urban communities or the variable symbolic associations between the Welsh regions or their speech (actual or stereotyped). To say to a Welsh person that somebody has a Welsh accent is a description of little depth or helpfulness.

This point will be returned to in Section 2.12.5 in the discussion on respondent provenance.

vi/. Style-authenticity: Reading aloud has its own norms and rhetorical conventions. For example, Labov (1972) distinguished between 'reading style' and 'careful/casual/spontaneous speech' and even reading fluency varies between speakers and may become a source of spurious variation (Thomas 2002:130). Can the attitudes elicited from an MGT with speakers reading aloud be meaningfully extended to unscripted and/or natural use language use?

An alternative variant of MGT is the 'verbal-guise' technique (VGT) (Gallois and Callan 1981). The VGT is a technique often employed out of necessity when it has not always been possible to find a single person who can competently produce varieties required for the study. The use of such speakers can also be used to defend against the charge of artificiality (Garrett et al. 2003:53-54). For example, Garrett (1992) in a study of reactions to the English pronunciation of foreign learners of English, was able to find French speakers of English who could produce English spoken with 'broad', 'mild' and 'hypercorrect' accents, and for comparison, Spanish speakers who could do likewise with Spanish-accented English. It would have been futile and detrimental to the realistic rendering of the variants to search for a single person to aim at producing all of the six recordings. The use of VGT may be seen as having great potential for addressing some of the shortcomings of MGT, in the light of forensic research where legal outcomes are at stake.

2.8.2 Direct measures: benefits in relation to MGT

Direct measures are characterised by interviews or questionnaires designed to enquire explicitly about the respondents' attitudes towards various language behaviours. These can include be: focusing the respondents' attention on their own speech habits/preferences (Labov 1966); or asking respondents to evaluate recorded speech samples of genuinely different speakers (van Bezooijen 1994; Kerswill and Williams 2002; Garrett et al. 2004); or asking respondents to undertake conceptual accent evaluations in the form of

researcher-prescribed labels of speech varieties (Williams et al. 1996; Bishop et al. 2005); or asking people to draw upon a map of the dialect areas of their community as they see them and provide their own labels range and name (Niedzielski and Preston 1999; Preston 1999). It should be noted that there is of course room for MGT within a direct framework if the researcher asks the judges first to identify the accent and then fill in scales, although the great majority of the literature does not follow this route.

In terms of benefits, direct measures afford the researcher the opportunity to reassure the respondents that there is no right or wrong answer and to furnish the respondents with as much or little data and preparatory guidance as they see appropriate, according to the research aims. Whilst this is also possible with, for example, MGT, it is limited by the researcher's choice of context. For example, Bourhis and Giles' (1976) use of a voice guise delivering theatre announcements in Wales or Kristiansen's repetition in a Danish context (1997) do not afford the researcher the opportunity to guide the respondents away from misunderstanding. The researcher is unable to clarify the task given or correct any respondents who have misunderstood the instructions.

Furthermore, and unlike the MGT, the voice samples can be examples of genuinely different speakers delivering non-content-neutral texts (as opposed to factually neutral), avoiding the need for one speaker to impersonate the speech of another variety or re-use the same speaking passage, for the sake of controlling all variables but the speech variety. Such a technique, avoiding verbatim repetitions, can reduce listener fatigue and is also able to accommodate Preston's (1996) criticism, in that the researcher can ask directly where respondents believe the speakers are from, and underpin any conclusions as to what the respondents were evaluating.

2.8.3 Direct measures: methodological considerations - biases

Direct measures not only give researchers explicit methodological 'permission' to provide respondents with attitude rating scales, e.g. bipolar semantic differential scales, Likert scales, but they also afford researchers the opportunity

to elicit evaluations by qualitative means, and both approaches can serve to be illustrative of each other. For example, a rating scale followed up with questions such as 'Why do you think this?', 'How important is this for you?' can often have the effect of opening the aperture through which we may see the attitudes, affording a fuller view of the shape and depth of respondents' attitudes. Such questions can also serve as a tool for enhancing the validity of the research approach by increasing the effect of the *accountability motive* (Chen and Hanson 2004).

The accountability motive, reflects the implicit or explicit expectation that one may be called on to justify one's beliefs, feelings, and actions to others, and has the effect of increasing the likelihood of respondents devoting more cognitive resources to the evaluation process. This may be disadvantageous if the judges become, for example, so acutely aware that the researcher is attempting to elicit socially stereotyped responses, that they divert an ecologically disproportionate amount of cognitive resources to the suppression of stereotypes, thereby becoming overly preoccupied with their responses and jeopardising the validity of the research. Moreover, when people know that they will be accountable for their views on certain topics, a parallel *accuracy motive* may exert an influence on persons' estimations, based on the fear or desire to not be perceived as wrong or inconsistent, or simply to have the easiest reason to hand should they need to justify their choices. The influence of our motivations should therefore not be under-estimated, as they guide what evidence we look to and remember, and how we interpret or scrutinise the evidence that we encounter.

A number of other problems may be associated with employing direct measures, some of which may be identified from the existing literature; others have been identified as part of the planning for the present study:

i/. Invented opinion: Judges supply an opinion where they have none, e.g. the earlier example (p.59) from Ostrom et al.(1994), describing the elicitation of non-existent opinions on the Metallic Metals act.

ii/. Acquiescence bias (Ostrom et al. 1994): Judges may be eager to 'please' and invoke opinions, modify their responses or offer agreement, supplying what they see as the desired stereotypical responses that they believe the researcher wants to hear in order to gain researcher approval. Moreover, in such cases, or where judges suspect that they can see through the task, they may provide responses which fulfil the task but may not reflect their own private views.

iii/. Social desirability bias (Cook and Sellitz 1964): Judges may be aware of the social undesirability of their answers and modify their responses to be socially more acceptable; for example, where the judges' private attitudes align with a controversial stereotype under investigation, they may focus away from the stereotype, if it is socially undesirable, focusing on more socially desirable responses. This is particularly relevant to direct questions involving controversial or sensitive subjects.

iv/. Ego-achievement bias: Where an acquiescence bias is based on the judges' eagerness to 'please' the researcher, an alternative, which I put forward here as a consideration in the planning of the current research, refers to the potential for respondents to feel that participating in the research is much like participating in a quiz, placing themselves in the role of contestant, and the researcher in the role of 'quizmaster'. For the respondents, the motivation to perform well in the quiz may not be for the benefit of the researcher, but for their own self-fulfilment. For example, participating in research into language attitudes or speaker description research involves listening to 'mystery voices from different regions/countries' or responding to conceptual labels based on regions/countries. This may create the impression in the minds of the respondents that the 'quizmaster' holds a set of pre-determined 'correct' answers that the respondents must try to identify; the task being for the respondent to draw on knowledge from the popular culture/dominant ideology and spot, match and supply the 'correct' popular social stereotypes attached to each label that the 'quizmaster' is asking for. If this is the case, the judges' responses would have less to do with their own evaluations of speech varieties or speakers, or their own inner or private attitudes, and more to do with performing well in the quiz by getting the answer 'right'. That is, judges'

responses may be mediated by the use of direct questioning and pre-labelled voices and reflect what they *know* about the stereotypes attached to this label as opposed to what they actually *believe*. An explicit test of this theory is suggested for future work; however, it will, for now, remain a consideration in the design and interpretation of the research.

In the event of a possible re-evaluation of previous research, we could ask: have responses in previous studies consistently been indicators of the respondents' own stereotypes and attitudes, or have they been invoked from the dominant culture to produce the requisite 'answers' they believe the quizmaster (in their role as the 'more knowledgeable' party) already knows? Given this possibility, researchers must also be cautious in interpreting the respondents' provision of the expected social stereotypes as being evidence of the validity of their research methodology.

v/. Respondent induction: A possible flip-side of the above is to avoid, as the researcher, actively 'guiding' the respondents; that is, researchers may be guilty of inducing the responses they want from respondents through a combination of offering to a willing, co-operative set of judges a set of overtly socially marked labels with stereotypes attached.

vi/. Parity/pity: A name I have attributed to the observation from Brown et al. (1985), whereby judges may be motivated either to back a winner, regardless of having no particular connection with any of the objects, or to give the underdog some more positive ratings. This may be of particular relevance in the context of, for example, attributions of guilt and will be returned to in the Discussion of the results (Chapter 6).

Other weaknesses that are not based on respondent biases include:

vii/. Weakness of judges' areal taxonomy: I identify this as a possible consideration for interpreting the data in this research. Respondents' areal taxonomy, their ability to describe where someone is from may be so weak that the intra and inter-rater reliability is compromised. Ask the same respondent on another day, or a different respondent on the same day, and there may be little

consistency or agreement in evidence if the areal taxonomy is weak. This will be dealt with in greater detail in the 'task' selection below.

viii/. Use of unrelated voice recordings: A 'sound-only' condition can heighten the effects of accent (Elwell et al. 1984) and when a between-speakers design of different voice donors is used for the samples, inter-speaker variations in verbal style may influence evaluations: breathiness, throatiness, pause durations, pitch, linguistic hesitancy, politeness (Street et al. 1984; Dixon and Mahoney 2004).

ix/. Ecological validity: In their study on 60 undergraduates at an English university, Brown et al. (1985) recorded a bidialectal R.P./South East Wales English speaker to read a passage to be played in a pseudo-health education context. They concluded that the design of the research - especially the commonly used, matched-guise, repeated measures design with direct comparisons of a number of speakers - can lead to accent being too prominent in the respondents' focal awareness to give ecologically valid results (Brown et al. 1985:218). This may be of particular concern in the context of forensic speaker description, where the role of accent may be over-represented in relation to the large number of real life cognitive tasks that may occupy a respondent.

A second concern of ecological validity is that of the researcher selecting voices for description tasks that have been chosen because of their high recognisability. It is not uncommon to read research in which the voice donors were selected for their distinctive accents, typical of a given city, region or country, with the accent authenticity of researcher's choice being corroborated by other professional linguists. The issue here is how strongly marked a speaker's dialect is. Thomas (2002:130-131) points out that although it may be advantageous to use speakers who show strong traits of the selected varieties under study, such speakers may not typify their speech communities, compromising the ecological validity of the study. Conversely, speakers who are most easily accessible to researchers, such as university students, may be

exposed to outside influences and may not be especially vernacular because of their socio-economic background.

x/. Verbal overshadowing: First introduced by Schooler and Engstler-Schooler (1990), verbal over-shadowing relates how giving a description of the visual appearance of a perpetrator can have a negative effect on subsequent identification. Perfect et al. (2002) extended this to the speaker identification domain and observed that verbally describing a to-be-recognised stimulus leads to a decrement in identification accuracy (Perfect et al. 2002:979). There are many considerations here which go beyond the scope of this research, such as the whether quantity and type of description, provision of frameworks to assist description or spontaneously encountered vs. laboratory encountered voices will affect performance. This is undoubtedly an area worthy of further study and whilst it will not be pursued in this research, its potential relevance to the field of forensic speaker description is acknowledged.

2.9 System variables: The questionnaire

This section discusses the range of tools that the researcher may use, such as the design of the questionnaire and the use of voice sample. In this section, the focus will be on evaluating these research instruments as they pertain to voice identification in particular, rather than encompassing questionnaire design and making sample recordings in their broadest sense.

The questionnaire offers the advantage of enabling the researcher to capture a lot of data from a large cohort of respondents with a good degree of systemic reliability and speed. It can be used in preference to face-to-face interviews and has been an essential tool throughout the history of social research. For the aims of asking respondents to describe where they think speakers may be from, asking for evaluations of the speakers' voices and eliciting attributions of speaker guilt, a questionnaire can be deemed the most appropriate instrument. Following a long-standing tradition of language attitudes research, the preparation of the questionnaire should include consideration of the following:

i/. Context: Of consideration firstly is the context for the research. Brown et al. (1985) and Giles and Coupland (1992) have suggested that the (sociolinguistic) context may mediate the attitudes of the respondents, and its context can serve as a reason for the judges completing the questionnaire (even if this does not entirely match the researcher's ultimate aim). Some contextual considerations might be:

- What, *for the respondents*, is the context of the research?
- Why should they take it seriously?

ii/. Content of the questionnaire: when considering the content, a fundamental question is how to engage with the respondents to meet the research aims. This could be achieved by means of closed-ended responses, such as tick boxes, yes/no, true/false response to closed questions or multiple choice selections. Alternatively or additionally, the questionnaire could include rating scales, which are mentioned in more detail below, to allow respondents to display their degree of feeling. The researcher must be aware, however, that if judges are only able to express their views through semantic differential scales, this may limit the amount of information that can be elicited or skew the responses by frustrating the respondent. One possible solution to this is to allow respondents space to provide their own descriptions. It is highly likely, in the context of an audiofit, that respondents will have comments and intuitions that they would like to add, clarify or explain. Finally, the researcher must decide whether to include perceptual maps on which respondents can mark their intuitions regarding the regional, locational, ethnic or social background of the speakers, as the aims of the research dictate (Preston 1996; Williams et al. 1996; Niedzielski and Preston 1999).

iii/. Speaker origin: The researcher must decide if the respondents are to be told where the speakers are from and whether this will information be given through labels to be read whilst listening to the voices (e.g. Giles (1970); Bishop et al. (2005)). The second option is to provide no information regarding the speaker origin and rely on the respondents' own judgement to evaluate the speaker origin. It must be remembered that not asking for their estimations of speaker

origin is a practice criticised by Preston (1996) and considered of vital importance in establishing exactly what the respondents are describing and evaluating. However, asking respondents for their estimations of speaker provenance can be facilitated in two ways – by providing a space for them to write the estimations, or alternatively, by providing a map, onto which the judges can mark their responses (Preston 1999; Garrett et al. 2003), a practice which simultaneously provides data regarding the strengths of the respondents' areal taxonomy.

iv/. Rating scales: How should non-linguists describe their intuitions regarding the speakers' voices? Should they be provided with a range of Likert-type scales (Likert 1932) on which to express the strength of their agreement or disagreement? Alternatively, should they be provided with semantic differential scales, usually on a scale from one to seven indicating, for example, how intelligent/unintelligent a speaker sounds? As discussed earlier in Chapter 2 (p.35), Zahn and Hopper (1985) conducted a review of many of the papers that had speech evaluation measures in the form of semantic differential scales, stating that the number and diversity of speech evaluation measures had been an impediment to the integration of research findings. They devised the Speech Evaluation Instrument (SEI), which 'concurrently assesses the dimensions of listeners' evaluative reactions to speakers and their language'. The SEI identifies three main dimensions: *superiority*, *attractiveness* and *dynamism*. In their research, Zahn and Hopper found that these three factors accounted for 64.5% of the variance in the subject ratings and as such, are widely accepted as being the underlying basis for a large number of semantic differentials. However, Zahn and Hopper also state that it remains to be seen whether different segments of the population utilise different dimensions of evaluation. Garrett et al. (2003:64) agree, stating that language attitudes research cannot always assume that the same set of universal dimensions will always be salient. However, the argument can be extended here, when one considers the context of the present research. For example, given the criminal context, and the questions regarding how respondents' evaluations of potential guilt may interact with their descriptions of unknown speakers, it may be profitable to consider an alternative set of dimensions from the field of credibility. Perloff (2003) suggests

that credibility is more than a simple unitary dimension and that it also contains three primary dimensions of evaluation, *expertise*, *trustworthiness* and *goodwill*. Yet when one aligns Zahn and Hopper's and Perloff's dimensions, one can argue that they do not represent identical evaluations.

Figure 2.9.1 Comparison of Zahn & Hopper's and Perloff's dimensions

attractiveness	-
dynamism	-
superiority	expertise
-	trustworthiness
-	goodwill

It may be argued that 'superiority' and 'expertise' represent the same evaluation, but that neither 'trustworthiness' nor 'goodwill' align with 'dynamism' or 'attractiveness'. The proposal here is that some kind of blend of the two sets of established dimensions may be more suitable to the context of the present research. Moreover, given the need for any questionnaire to be semantically transparent and relevant to the task respondents find themselves completing, it is suggested here that the combination of the new set of dimensions should be relabelled. The following is suggested:

Figure 2.9.2 Comparison of dimensions

Zahn & Hopper	This research	Credibility
dynamism	aggression	-
superiority	intelligence	expertise
-	honesty	trustworthiness

I would propose that the term 'dynamism' be replaced with the more violent crime-related and semantically transparent 'aggression'; I would argue also on the grounds of semantic transparency and contextual relevance that 'intelligence' be used in place of 'superiority', and 'expertise'; I would propose that 'honesty' replace 'trustworthiness' as a label to be presented to

respondents as it more directly related to a criminal context; finally, I suggest that 'attractiveness' and 'goodwill' are inappropriate for the context of evaluating potential criminals and should not be included here.

Finally, it is worth observing that the use of multiple evaluation scales can be illuminating, not just from the perspective of principle components/factor analyses used by Zahn and Hopper, which allow us to identify the most common factors that underline accent evaluation. Such scales may also highlight correlations between speech style and social evaluations, For example, van Bezooijen and Ytsma (1999) invited respondents to provide subjective evaluations of speakers' personalities. Speakers of Standard Dutch were regarded as sounding more arrogant, whereas speakers of regional dialects were rated more positively. Van Bezooijen's (1988) investigated respondents[1] evaluations of speakers of various socio-economic status from the city of Nijmegen and found correlations between prosody and strong personality, and also between intellectual/socio-economic status and segmental speech features.

v/. Degree of accentedness scale: A particularly common feature of lay descriptions of voices is to state how strong an accent is, which we saw evidence of in the questions asked by the South Wales Police DC, in Chapter 1. This is often articulated in terms such as a speaker having a 'strong', 'broad' or 'thick' accent. If this is a common feature of non-linguist observations, then it should have its place in any research pertaining to speaker voice descriptions. Van Bezooijen & van Hout (1985) in their research on evaluations of Dutch include a 4-point scale in which 0 = standard Dutch, 1= slightly accented, 2 = fairly accented, and 3 = heavily accented.

2.10 System variables: Voice samples

It is clear that at the heart of research into forensic speaker description will be the stimulus voices. Issues to be considered in selecting the voice samples and their texts are set out in this section below.

Within this context Braun and Künzel (1998) have observed that there is great value in including normative observations - comparing a heard voice with another to illustrate characteristics, especially to highlight any peculiarities or defects in the pronunciation of certain speech sounds. In this way, speakers may be described or identified not only by which regional or social groups their accents fit into, and not only by the dimensions for describing voice quality that are common to all, but by those salient idiosyncrasies that mark the speaker apart from a perceived norm. Such an approach not only extends the possible identificatory criteria, but also entails more expanded descriptions, taking us from one-word or semantic bipolar descriptions to more extensive description of the range of possible interpretations for each available possible item.

Finally, before expanding the discussion of voice samples, we should acknowledge that whatever the voice sample used, there will always be the possibility of intra-speaker variation. No speaker is free from variation in the speech and all will be subject to changes in delivery depending on physical conditions, acoustics, emotional states and relationships with the audience. Given this, we must expect that the features that make a voice distinctive on one occasion may not appear in all contexts. Reciprocally, not all contexts will induce the same speech variations. And what makes a voice distinctive on any one occasion may be a combination of features, each of which may be present in any given instance (Nolan 1997). Such an understanding should inform the selection of voices and the interpretation of any findings.

2.10.1 Voice samples: matched-guise

In the section above I detailed many of the main strengths and weaknesses of language attitudes research, and noted that many researchers have selected the matched guise technique thanks to its facility to control variables and manipulate the linguistic aspects. However, when considering the selections of voice samples, if a matched guise technique is to be employed in order to emulate a between-speakers design, a fundamental difficulty is likely to be recruiting a symmetrically bi/tri or multi-dialectal or lingual speaker, the difficulty of the task being commensurate in the number of voice samples used. When

the research aims are extended to include forensic speaker *description*, however, the debate over whether or not to employ a matched-guise technique ceases to be: in research where the aim is also to elicit a range of qualitative descriptions not only of the speakers' accents, but also of their voice qualities, the logical decision must be to employ a different speaker for every sample, in order to elicit comment on the maximum possible range of physiological renderings. Only if the specific aim is to evaluate a single speaker's ability to display a range of physiological features might there be a need for a matched-guise technique in forensic speaker description. Research in forensic speaker description is compelled, therefore, to recruit voice samples from different donors to be played to respondents.

2.10.2 Voice samples: accents

We know that language varies regionally and socially, and that different regional and social accents provoke a range of predictable responses from non-linguist judges (Giles 1970). However, we also know at the level of user that two people from the same area may not speak the same way (Labov 1966), and that speakers style-shift according to their audience (Bell 1984; Bell 1997; Wolfram and Schilling-Estes 1998). It may be wrong therefore to assume homogeneity within any one region or social speech community or to assume consensus model of the speech community, i.e. a model of agreement on the social value of speech varieties. Given this, the use of accents in research into forensic speaker description must be treated with caution, and two distinct approaches can be identified:

i/. Top-down: The question here is whether the researcher is interested in the associations and evaluations that particular regions or social groups provoke. If the researcher wishes the respondents to identify the attitude objects and tap into the social stereotypes attached to these groups/labels and the responses they elicit, the researcher needs to identify speakers from these regional and social groups who use marked accents, record them, and verify independently with a pilot group or panel of expert phoneticians that these accents are acceptable examples from the speech varieties being investigated. Such work

has been undertaken by many researchers into language attitudes and has resulted, according to Edwards (1982:22) in respondents providing judgements that allow the expression of stereotyped reactions to the language varieties rather than the speakers. Edwards' assessment is supported by work using only conceptual labels rather than voices, which elicit similarly stereotyped judgements (Giles 1970; Coupland and Bishop 2005), without real voices being heard by the respondents.

ii/. Bottom-up: Alternatively, the researcher may be interested in the respondents' own perceptions and estimations of speaker origin. Here the research aim may be to understand within the given context (e.g. narrating a story, responding to questions) what the respondents decide is salient for them within the accents and voices in the recording, and the subsequent associations that arise from their evaluations. If this is the case, the researcher may have less need for finding and recruiting speakers with stereotypically marked accents only and verifying them with independent groups beforehand. Such an approach is more likely to lend itself to a verbal guise technique, allowing for a potentially wider cohort of candidate voice donors, and avoiding the potential contrivance inherent in the matched-guise technique and the effects of respondent induction (p.89).

2.10.3 Voice samples: male versus female voice donors

The choice of whether to use all male, all female or mixed sex voice donors is bound to the aims of the research. We know that the sex of the voices may provoke differing evaluations (Street, et al. 1985). However, there is a lack of research to indicate what differences there may be between forensic speaker *descriptions* of each sex, given the physiological differences between male and female vocal apparatus. Nevertheless, the majority of research has tended to concentrate on male voices, in the forensic context especially, which is most likely to be a reflection of the higher number of male criminals to female.

2.10.4 Voice samples: ethnicity

Research into Black American English in the language attitudes field (Lind and O'Barr 1979; Niedzielski and Preston 1999) has suggested that ethnicity is an area for study in itself. The researcher must decide whether to include more than an example of ethnic voices as an independent variable or whether ethnic accent as a cue is worth research in its own right (Riches and Foddy 1989; Dixon et al. 2002). The potential relevance to future forensic speaker description research is acknowledged.

2.10.5 Voice samples: the use of actors

Broeders (1996) is quite emphatic regarding the use of actors, stating that they should be avoided at all costs. He argues that linguists and phoneticians will be fully aware of some of the features that are likely to manifest how the speech of actors and 'professional speakers' can differ from that of other language users Broeders (1996:9). Broeders is presumably referring to elements of speech delivery associated with rehearsed texts and performances for public display, and the point that staged voices may sound unnatural to a lay audience.

Given the serious legal consequences involved in not providing extremely convincing renderings of 'natural' sounding voices – misidentifications and missed identifications – the argument for avoiding professional speakers is strong and this study follows that argument.

2.10.6 Recording voice samples

One of the key points that distinguishes forensic speaker description from more general language attitudes research is the forensic context. Fundamentally, the research aims to not only collect data regarding non-linguists' speaker evaluations and voice descriptions, but also investigate the influence that the forensic context may exert upon these evaluations and descriptions and vice versa. With this in mind, the researcher needs to convey some level of forensic context with the recordings. Three possibilities are outlined below:

i/. Recordings of crime in action: The researcher could use as the voice sample recordings of crimes in action, asking respondents for their 'online' impressions. This option would be high on ecological validity with regards to the forensic context and include entirely natural delivery by the voice sample, avoiding the issue of using actors or reading texts aloud. This route is, however, fraught with constraints: firstly, access to such material is particularly difficult, in the UK especially, as the police only publish evidential materials in selected circumstances; secondly, the recording of voice data of a crime in action relies on the available technology, and at the time of writing, despite the plethora of CCTV cameras in the UK, few of them record sound, and those that do, do so with such poor audio quality that they may prejudice the impressions of the speaker's voice, or may require subtitles; thirdly, there are acute ethical constraints upon presenting recordings of crimes in action due to the trauma and upset that may be caused to the respondents.

ii/. Simulated crimes in action: Television programmes reporting on crimes frequently use re-enactments as a device to prompt viewers' memories or to assist them with understanding the anatomy of a crime. With the appropriate preparation and debriefing of respondents, this route may be ethically less complex. Careful consideration must be given, however, to the fact that such programmes are designed for multi-media presentation with the story being told simultaneously through images and sound. The presentation of visual data may prejudice the processing of aural data, and the isolation and presentation of only aural data may give an unbalanced account of the speakers pragmatic performance, (e.g. appearing to violate Grice's (1975) maxims or language expectancies (Burgoon 1995), whereby the language used does not conform to the social norms, but violates the receiver's expectations by deviating from the expected behaviour). This, again, may again prejudice the respondents' evaluations.

iii/. Suspects talking in contexts other than crime performance: An alternative to the recording of crimes in action may be to present respondents with recordings in which the speakers are only *suspected* criminals and are heard speaking but not enacting a crime (Lind and O'Barr 1979; Seggie 1983; Dixon et al. 1994;

Dixon et al. 2002; Dixon and Mahoney 2004). There is the danger that absence of gunshots and screams, for example, may dilute the forensic context as the respondents do not hear for themselves that the speakers are criminals. In the absence of a recorded crime, it would certainly be the researcher's responsibility to establish in the minds of the respondents who these speakers were, why they were recorded and what their purpose is for speaking. Whether this detracts from respondents' attributions of guilt is a matter for future research.

One possible way of accounting for the suspects having been recorded is to suggest to the respondents that these recordings are police recordings, made in police stations, on police recording equipment, made with a number of individuals apprehended on suspicion of committing a crime. This approach offers some distinct advantages, in that it is ethically less contentious; the laboratory-based allusion to real-life conditions facilitates greater experimental control (Bull & Clifford 1984:93); also, following the tradition of language attitudes research, such an approach affords the researcher the opportunity to in some way engineer the recordings to be centred on the same topic and be more content neutral. That is, it does not involve divergent accounts of a crime in action, the differences between which may be more of a predictor of guilt attributions than the evaluation of voice or accent. Content neutrality is important in research such as this. We know from studies such as Garrett et al. (2004) where Welsh teenagers listening to other Welsh teenagers' narratives judged the Cardiff speaker, with the most entertaining story but disfavoured accent, most positively, that when content neutrality is not observed, there is likely to be an effect on the way that the speakers are evaluated.

With regards to the quality of the recordings, if recorded digitally, the researcher may care to slightly degrade the quality of the final recordings through digital manipulation, in order to endow a feeling of genuine recordings 'in the field'. However, the recordings must not be degraded to the extent that parts may be too difficult to hear, and if different accents are to be used, the quality should not impact on recognisability.

2.10.7 Selection of voice samples

We know that any texts used as prompts must be sufficiently extensive to convey details about the voice (van Bezooijen & van Hout 1985) for listeners to form opinions and provide valid evaluations. The decision regarding the content of the text may, however, present difficulties. What needs to be decided is how much between-speakers textual diversity to admit onto the voice samples. If the researcher opts for one of the first two examples above, of using recordings of live or simulated samples of crimes in action, the researcher is unlikely to have any influence over what language is included in the recording, due primarily to the need to be faithful to the facts, and due also to the unlikelihood that a researcher's interest matches those of TV production companies. Therefore, the only control that the researcher is likely to be able to exercise is editing whatever recording they are able to access, reducing the data to the best match for their aims. The fact that there is variation between the speakers is not in itself a problem, depending on the research aims. For example, Garrett et al. (2003) conducted attitudinal research related to narrative performance, at the very heart of which was the investigation of how the speakers within a Welsh context performing a short narrative to their peers were evaluated by other speakers within Wales. By necessity, the research involved a number of diverse texts.

With regard to the number of voice samples to use, as there is no previous research that has attempted to elicit voice descriptions from non-linguists with the aim of deriving an audiofit, there are no established guidelines to follow. As a reference point, there is strong support in the forensic voice identification field in the UK for the McFarlane guidelines (Nolan 2003), drawn up by a police officer dealing with a case of arson in London which led to the death of a young woman. The McFarlane guidelines recommend, amongst other detail, that a maximum nine voice samples be used in a line-up (suspect + eight foils). However, the present study is not recreating police line-ups and is not a study into forensic speaker identification; as such, it is not involved in researching passive recognition, but rather in eliciting active descriptions from respondents. It is highly likely then that any study of this type will have to consider using

fewer than nine voice samples in order to keep the cognitive load for the respondents to a manageable size, avoiding listener fatigue.

With regard to content of the voice samples, there is some evidence that differences in vividness of content influence judges' evaluations, especially in the legal context. That is, evidence from witnesses that is emotionally interesting, concrete, image provoking or proximate in a spatial, temporal or sensory way may lead respondents to draw inferences about greater credibility of the communicators (Bell and Loftus 1985:660). With this in mind, we know that between-speakers content variation has long been a controlled variable in language attitudes research, the aim being to direct the respondents towards the linguistic variation in order to illuminate the effect that the manipulation of linguistic variables has upon respondents' evaluations, rather than focusing their attention on the message. The decision for the researcher of forensic speaker descriptions, therefore, is to evaluate to what extent content-neutrality, the lack of marked differences in the content of each voice sample, may throw into relief the contrasts in speaker accent. Garrett et al. (2005) point out that the repetition of the same message content can make accent more salient than it would be outside the experimental environment and over-amplifying the effects of speaker accent, may encourage ecologically invalid attributions of guilt. If the researcher decides that content neutrality is essential to the research aims, the choice above as to which type of voice recordings to play to the respondents is made simpler: Eschew the first two options of using pre-recorded texts, and opt for recruiting new *voice donors* with varied accents, who are provided with researcher-engineered content-neutral texts to deliver.

The between-speakers accent contrast may be softened by the use of context. As discussed in Section 2.6, there is an argument against the abstraction of attitude evaluation from a real-world context. The forensic context may have a critical role to play here as it potentially provides researchers with a real world context for when and why one might ask five people the same questions and receive very similar answers. If the researcher suggests that the recordings that respondents are going to hear were produced in a police station, and that the voices on the recordings are those of, for example, five suspects, then the



similarity between the speakers, their purpose for speaking and the presence of an official interlocutor (police officer) is implied, endowing the existence of five content-neutral texts with a sense of plausibility.

It will be understood by most respondents that the reason for recording suspects speaking in a police station is for the purpose of an official interview, pending an investigation. By its very nature, an interview centres on the story of the individual, which is elicited by an interlocutor. It is the interviewees' account and the faithfulness to the material facts that is central to the investigation. It follows, therefore, that if five people are interviewed, in relation to the same crime, it is likely (especially if all are innocent) that they are able to report similar accounts of the facts, building the expectation that texts may be content-neutral.

2.10.8 Voice samples: texts and the Picture Memory Interview

Central to the formulation of plausible police interviews is the maintenance of content-neutrality throughout yet not allowing the voice donors to sound scripted. We know that reading has its own rhetorical conventions (Garrett et al. 2003), moreover, the same speaker can be evaluated differently when reading a text than when speaking spontaneously (Labov 1966), with the listener's attention being drawn to the incongruence between formal reading features and informal dialect features (van Bezooijen & van Hout 1985). The use also of verbatim texts, replicating parts of a crime in action may give rise to emotional reactions on the part of the respondent which may unfavourably affect their judgements (Komulainen 1988).

Given this tension, a form of visual prompt is advocated for the purposes of this research. Visual prompts are not without precedent in the forensic literature. The DyViS project at Cambridge University (Nolan et al. 2008) has, at the time of writing, an ongoing project to create a forensically oriented speech corpus by interviewing 100 male speakers of 'standard southern British English', using visual stimuli (http://www.ling.cam.ac.uk/dyvis/database/database.htm - accessed 10/9/08). For this research I have developed a technique entitled the Picture Memory Interview (PMI), the characteristics of which are as follows:

i/. The PMI is based on the need for spontaneous-sounding, naturally delivered speech samples.

ii/. Specific to the need to create the illusion of a real interview, the PMI provides the voice donors with pictorial answers to questions, rather than long and complex statements.

iii/. The answers to the questions are provided only in pictorial form, avoiding giving the voice donors prescribed sentence structures or requiring them to read sentences which may re-appear in their responses, thereby minimising the appearance of the researcher's voice in the voice donors' texts. This also prevents deviation and ad-libbing on the part of the voice donors.

iv/. The voice donors are provided with a series of pictures and a practice interview is conducted. The researcher asks questions, and the voice donors interpret the visual cues as their own responses. If content-neutrality is required, each voice donor is given the same picture cues. If the requirement is to provide voice samples with differing content, a variation of pictures can be utilised.

v/. Once the researcher has checked that the voice donor has interpreted the picture cues correctly, a preliminary recording is made, with the picture cues covered. The voice donor provides their own answers from memory.

vi/. If the quality, naturalness, spontaneity and content-neutrality of the recordings are acceptable, no further recording is necessary.

vii/. If necessary, the researcher can make second or third recordings until satisfactory levels have been achieved.

viii/. The recordings are then edited to remove the voice of the researcher.

The PMI offers several advantages: firstly, the voice donor will be engaged in online verbalisation of memorised items of their experience, replicating a process that forms part of a police interview and imbuing a sense of ecological credibility; also, the voice donor is not reading but using entirely their own phraseology and co-producing natural pauses, fillers and other discourse

markers, adding to the authentic character of the recordings; finally, following the tradition of much MGT work, we know that by conducting PMI's, we can be confident that voice donors are delivering practically the same information, barring individual phraseologies that are a natural part of speech delivery, the only variation being limited to individual pronunciation and manner of delivery. Furthermore, the use of the PMI to direct voice donors to deliver a form of narrative of events has the potential advantage of reducing the likelihood that the rater will be distracted by a sequence of disconnected fragments that make no sense to the judges (van Bezooijen and van Hout 1985). It may be assumed that the evaluation of spontaneous speech is more representative of reactions in everyday life and therefore carries greater ecological validity within the forensic context.

2.10.9 Voice samples: length of recording

As with the number of voice samples to be used in forensic voice description tasks, there is no precedent or agreed ideal length of voice sample to be used for eliciting speaker descriptions, and this study is navigating somewhat unchartered waters. We know that any voice sample played to respondents should be of sufficient length for the judges to identify any voice features that may be salient to them, yet we do not want to risk listener fatigue by using an overly long sample. Also, we wish to recognise the need for some level of ecological validity - exposure to criminal voices is recognised as being usually a brief experience, so we do not wish to give respondents an unrealistic opportunity to listen to the voice samples in depth.

Views from neighbouring academic fields provide diverse advice on this matter. Bull and Clifford (1984:107) in their research on forensic speaker identification found no significant difference between recognition rates when adult respondents were exposed to 1, 2, 3 or 4 sentences; Bull and Clifford (1999) report that the number of utterances heard from a target speaker in a voice line-up was found to have hardly any effect on listeners' performance; Yarmey and Matthys (1992) in their study of 576 undergraduates, exposed to a recording of a mock abductor for periods of 18, 36, 120 seconds and 6 minutes, report not

only higher hit rates in terms of correct identifications, but also considerably higher false alarm scores, when exposure to the target voice was increased to between two and six minutes from 18 to 36 seconds. Blatchford and Foulkes (2006:247), in their investigation of the recognition rates of shouted utterances found a significant difference between speakers recognised after two syllables (52%) and after 12 syllables (81%). The McFarlane guidelines mentioned in Section 2.10.7 (p.94) recommend that selected speech samples should be about one minute long and can comprise various fragments of speech joined together.

The topic of sample length is even less conclusive in language attitudes research, with a long tradition of eliciting evaluative responses by playing recorded voice samples; most language attitudes researchers do not detail the exact length or number of sentences that have been used in their studies, and certain studies have consciously selected long speech samples as part of the methodology, for example Garrett et al.'s (1999) study of evaluative responses to young Welsh English speakers' narratives. It appears that consensus in the forensic linguistic and language attitudes fields as to the ideal sample length and whether it is of relevance to the research results is a while away.

The decision for the researcher of forensic speaker description is at what point have respondents heard sufficient to make their descriptions, and at what point is continued exposure having a negative effect, through listener fatigue, on the respondents' ability to perceive and report what they here. For forensic speaker description research, it appears that the decision regarding sample length will be an experimental one, involving informed trial and error.

2.11 System variables: The task

There is a long tradition of language attitudes task design that suggests that the use of semantic differential scales and Likert scales, as mentioned above, can assist respondents in articulating their impressions and evaluations. In terms of the question types, the researcher should, however, be aware of the following:

i/. That there are sufficient scales for judges to plot their responses.

ii/. There may be too many scales leading to a cognitive overload for the respondents and a fatigue effect.

iii/. That supplying only semantic differential scales employing adjectives through which respondents may express their views/attitudes may limit the range of information that can be elicited and restrict the respondents to operating in an evaluative vacuum (Brown et al. 1985), or encourage respondents to expand their views to make judgements they would not normally think of. This might otherwise be described as an example of a *notational* bias, whereby pre-coded responses are restricting the researcher's ability to view the possible wider variation in non-linguist responses. (See Section 2.8.4 for a list of other biases in this respect.) This may, however, be unavoidable; to a certain extent, any study will inevitably need to restrict data in some way in order to focus on specific research questions, and respondents too may restrict themselves to providing what they consider to be enough information without having to write at great length.

iv/. There is much value in supplementing quantitative scales-data with at least short, open-ended, free-response items, asking respondents to elaborate on their judgements (Garrett et al. 2003). This may not only illuminate the reasons for the evaluations, but also encourage respondents to feel accountable for their evaluations offsetting some of the response biases that may arise. On the other hand, too much pressure may lead the respondents to provide the response that is easiest to justify (invented opinion/acquiescence bias, p.79). The researcher should consider carefully how accountability items are worded.

2.11.1 Outlet for respondents' attitudes – maps

As mentioned above, evaluative scales and qualitative descriptions are not the only way of eliciting respondents' evaluations and estimations. Perceptual mapping, developed by cultural geographers Gould and White (1986) has been used in both British and American contexts in order to elicit a sense of

respondents' areal taxonomy of the region/community under research, making use of respondents' mental maps of regional areas and examining the labels assigned to various regions. Respondents may, for example, be presented with a blank map upon which they draw dialect boundaries around where they believe regional speech zones to exist, which they are then invited to label in their own terms. Examples of this procedure include Preston's work asking respondents to label regional variation in the US (Preston 1989) and Inoue's perceptual mapping task of the UK (Inoue 1999), discussed on p.44 of this thesis, the map of which is reproduced on p.132. Other tasks have required respondents to listen to recorded voices and then try to locate them on a map, such as Preston's (1996) work inviting respondents to plot heard voices on a north-south US transect (mentioned above, p.40). On one level, the use of perceptual mapping differs greatly from the 'traditional' language attitudes approach of presenting speakers from different pre-selected, prescribed labels determined by linguists.

One of the great advantages of this approach is the opportunity for the researcher to see not how language actually varies, but how it is *perceived* to vary, by examining and comparing the respondents' sketches; for forensic speaker description, it is the perceptual reality that is the goal, rather than the reality of the distribution of features.

For example, since the 18th century, RP and most dialects in England and Wales have undergone a process of *r dropping*, deleting pre-consonantal and absolute final position [r] (Wells 1982:218). However, the rhotic accent which preserves [r] in all environments where it occurred historically, is still found in various areas of the UK - the West Country (South West England), amongst older speakers in the rural South East of England, on the Welsh borders, in the North West of England and in most of Scotland. However, one would not employ a perceptual approach to establish which areas in the UK are rhotic speaking areas, i.e. in which area the post-vocalic [r] is realised. This is because it would be highly likely, certainly in the context of England and Wales, that non-linguist respondents would have little idea of the 'real' distribution of such features. Most respondents would be likely to associate the rhotic accent

with high-profile social and media stereotypes such as a farmer from South West England or the urban accented Bristolian (also South West England) 'chav' character, 'Vicky Pollard', from the popular British comedy series Little Britain, a derogatory parody associated with low socio-economic class, juvenile delinquency, fixated on counterfeit goods and criminal activity. On the other hand, when considering the uses of perceptual research, one *would* employ a perceptual approach to establish where people *think* rhoticism takes place; information that could convey interesting details to the linguist folk beliefs and social stereotypes.

The arguments in favour of perceptual dialectology, in particular the use of hand-drawn maps on which the respondents themselves identify, plot and characterise those major towns/regions/countries/areas that are most salient, are many. For example, it may facilitate a better insight into people's representations of language variation and what is salient to them; it may assist in the articulation of their beliefs about language, language use and language users, all the comments being contextualised within the respondents' speech community. Preston (1996) has maintained that offering respondents the opportunity to mark the evaluations and beliefs on to real maps facilitates the interpretation of language attitudes data.

It must be made clear that the aim of the exercise is not to measure non-linguists' knowledge of areal language against the results of professional linguists' research. Moreover, in an area such as forensic speaker description, where we need to understand language variation from the non-linguist's point of view, perceptual mapping may offer the non-linguist an outlet to complement or support their verbal description. Perceptual maps have the potential to allow us to view the respondents' world in terms of their own geo-referential categories, using their own geo-social shorthand.

2.11.1.1 Limitations of perceptual mapping

Whilst perceptual dialectology may be seen as a valuable methodological departure from traditional language attitudes approaches, we must consider its limitations:

i/. It is spatial in its orientation; the use of geographical stimuli – maps – has the potential to corral the respondents into conceptualising, visualising and reporting, through patchworks of isoglosses and geographical language variation, to the exclusion of social dialects and inter-personal contextual variation. How does one report on a map, for example, that their next-door neighbour is Scottish, but by virtue of going to a private school in the UK, has an accent that approaches RP?

ii/. Respondents may vary greatly in their knowledge of spatial geography – distinct towns, areas and regions. While some may be well-travelled, others may not; where some may have an idea of the range of areas within their nation or region, others may not. Where some may be aware of systemic regional and social accent variation, others may be poor at relating it to regions, cities or countries. Various researchers have alerted the academic community to this gap. Preston (1989), for example, complained that language attitudes research did not confirm if respondents had a mental construct of a place where a voice could be from, e.g., their mental maps of regional speech areas might not include one with a sample voice that could be identified and so there is little point in claiming that any reaction they do give is an indication of how they judge voices from this geographical area.

iii/. Respondents may vary greatly in their knowledge of visual geography. Where some may have need of a mental map (a long-distance lorry driver, for example) others may not; whilst some may see the weather map on television and be aware of the component parts of the country, others may just see a shape with random towns within.

iv/. Respondents may vary greatly in their ability to relate their knowledge of visual and spatial geography; i.e., they may know what someone sounds like if they are from Sheffield or the wider county of Yorkshire, but be unable to locate either place on a map; or they may be able to define the area of

Yorkshire on a map, but have no real knowledge of how natives to that area speak.

v/. Relying on a combination of untrained areas, where respondents have potentially weak knowledge of the sounds and geographical boundaries of their community and a weak ability to plot their perceptions spatially on a map, has the potential to dilute perceptions to the point of irrelevance.

vi/. Much perceptual dialectology is based on the respondents identifying the linguistic communities as they see them, rather than responding to a set of prescribed variants given by the researcher. However, being free to identify the relevant geo-linguistic communities, the communities that the respondents identify may not reflect the true range of linguistic diversity to be found in the community. The focus of their attention may well be drawn to those most widely discussed, more socially prominent or politically powerful characters in the community's geo-linguistic soap opera with less prominent but statistically or socially relevant variants being under-represented cameos or over-looked completely. There is the danger of ultimately eliciting as much about social myths and legends as there is about the linguistic diversity of the real life characters that inhabit the speech community.

The responses may tell us what people think (or what they *think* they think), and we may even elicit consensus amongst our respondents. However, we must remember that perceptual dialectology approaches carry with them the danger of eliciting responses that are likely to be restricted to comments on the most familiar or popular (geo-linguistic) social stereotypes, rather than information regarding how they view the full range of language variation in the communities that surround them. The aim with the current research is to employ a bottom-up approach and investigate what non-linguists can provide, rather than what they can say about prescribed and labelled accent samples.

It is vital, then, that the researcher is clear as to the aims and intended interpretations of the data when considering the deployment of perceptual mapping tasks. Is the aim to evaluate non-linguist respondents' knowledge? If

so, assessing their conceptual representations on a map may not be an entirely valid approach, as limitations on their areal taxonomy may not reflect their ability to distinguish and identify accents at the aural level or vice versa. Likewise, representing evaluations on a map may not reflect the reality of language variation in a given community. It is perhaps safer, if utilising a perceptual mapping task, to interpret the data as evidence of, 'if this is what you ask non-linguists, this is typically what they say and how they represent it', rather than, 'this is what non-linguists know or can tell you.' This allows us to set aside any issues regarding competence or accuracy, which can be compared to Preston's (1996) work on availability and levels of prompting (see below, p.39)

In summary, when considering the use of perceptual mapping, one must decide on the purpose of the activity. Non-linguist intuitions about geo-spatial language variation are clearly not as informed, detailed or accurate as those of a trained, professional linguist, and the descriptions elicited from the folk are clearly not intended for comparison with the professional view. However, if one has an interest in how, at a community level, non-linguists perceive the language around them to vary, and in what patterns, perceptual mapping may be regarded as a useful instrument in the researcher's toolkit.

2.11.2 Guided perceptual mapping

The preparation for this research has identified a potential adaptation to perceptual mapping. Utilising common, culturally-located conceptual frameworks of spatial geography, respondents may be assisted in plotting their perceptions by the provision of maps that are overlaid with simple labels and boundary lines that reflect generalised speech communities. *Guided perceptual mapping* allows respondents to work within an assisted conceptual framework, of, in the United Kingdom for example, Wales, Scotland, Northern Ireland, North England, South England, etc, demarcating visually for the folk approximately where regional speech areas 'begin and end'. Not only does this 'pre-indexing' offer the benefit of reducing the amount of geographical information that the folk themselves must provide, but by marking towns that may be conceptually equated with prestige speech in the mind of the non-linguist, e.g. Oxford or

Cambridge, the respondent is able to use a geographical tool to represent social intuitions in addition to regional ones.

2.11.3 Playing the samples

With regard to the playing of the voice samples, the researcher should ensure that all samples can be heard by all respondents in the room, and of vital importance is to ensure that same sequence is not repeated to all respondents, in order to counter any order effect. Furthermore, sufficient time should be provided to all respondents to ensure that they are able to consider the evaluations. Unless specifically intended, this should, however, also be limited so as to avoid extensive time to re-consider and reformulate their responses. Respondents should be advised that there exist no correct answers, that it is their own personal evaluations that are of interest and that copying from each other is neither necessary nor helpful. Finally, respondents should be clear that all responses given are provided anonymously and that their answers cannot be traced.

2.12 Estimator variables

Estimator variables (Broeders 1996) relate to validity and earwitness performance. In this section examples of estimator variables such as respondent age, sex, experience, etc, will be considered as they relate to forensic speaker description.

2.12.1 Estimator variables: salience and speech sensitivity

If one of the aims of forensic speaker description is to elicit from non-linguists linguistic information regarding the accents encountered in the field, it is essential to establish how salient these accents are to them. In consideration of how to quantify what is salient to non-linguists, Preston (1996:40) proposed four types and levels of folk awareness (detailed on p.39), *unavailable, available, suggestible* and *common*. The question for research into forensic speaker description is how available the accents and voices encountered are to the non-linguists; if the accents encountered in the research or in the real world are

unavailable to the folk, there may be little point in attempting to read attitudinal preferences in the data. In contrast, should the accents be anywhere between *available* and *suggestible*, there is a surely a social responsibility upon the field of linguistics to explore how the folk can be engaged in order to best elicit what available detail they have and assist the criminal investigation process. The concept of availability will be a regular reference point throughout this thesis.

In the field of forensic speaker identification, research has shown that there may be a small amount of between-listener variation in the task of speaker recognition, on a measure of 'auditory speech sensitivity' (Pahn and Pahn 1991). This is a test designed to investigate an individual's ability to perceive and produce different elements of speech, that is, tonal movement, voice onset, pitch, rhythm, dynamics, nonsense syllables and their combinations (Köster et al. 1998:23). There is, then, the possibility that the ability to describe voices and accents heard may be mediated by individual variation across respondents. An example of this is provided by Bortz (1970) cited in Bull and Clifford (1984:100). Bortz investigated how a witness's own voice may influence his description of others' voices and found that raters with high-pitched voices judged others as having low-pitched voices, and vice-versa.

2.12.2 Estimator variables: respondent memory

With regard to the delay between exposure to a voice and the description, we have no substantive evidence in the context of forensic speaker description to draw upon. What we do have is a substantial body of work on the effects of delay between exposure and participation in a voice identification parade. Bull and Clifford (1984:119) conclude that the safest, most valid conclusion is that although delay does have a detrimental effect, that effect is far from catastrophic. In the legal context, an upper limit to any possible delay suggested in the McFarlane guidelines (Nolan 2003:289) is that voice identification parades should take place within 6-8 weeks of exposure. However, although judges' perceptions may be influenced by the recency and frequency of exposure, experimental results demonstrate a lack of consensus on the subject. It is the author's feeling that the study of memory and its effects on forensic

speaker description warrant their own respective research. With the exception of alternating the sequence in which the voices are heard to combat any ordering or memory effects, memory as an independent variable is not included as a consideration in test design in this context as it would be beyond the scope of this research.

2.12.3 Estimator variables: respondent sex

Data from the field of forensic speaker identification suggest that men and women perform equally well in identification tasks (Braun 1996; Hollien and Schwartz 2000), suggesting that they are indexing and recoding voices in a similar way. Given this, there appears no immediate need to strictly control this as a variable, other than as an area of investigation in itself.

2.12.4 Estimator variables: respondent age and age estimations

The twin issues of respondent age and the ability to estimate age are closely linked. Success in age estimation does not appear to be related to matters of training, as Braun (1996) found in research that compared trained and untrained listeners' age estimations. She found that the success of age estimation did not did not significantly differ between the trained and untrained listeners, concluding that age estimation is a task which does not require any phonetic skills, let alone training in forensic phonetics, but rather, is based on the everyday experience (and often on stereotypes) of a listener within a speech community. It appears that stronger factors that mediate the success of age estimation are the age of the speaker themselves, the age of the voice they are trying to evaluate and the physical condition of the speaker.

Braun (1996) found that older listeners tended to over-estimate speaker age, whilst younger listeners tend to under-estimate it. This finding is quite consistent with tendencies reported by other researchers: Bull and Clifford found in their research from the voice recognition field, that where a witness is under 16 years of age or over 40 years, extra care should be exercised in cases where voice identification is deemed to be of importance, due to the drop in recognition

accuracy noted for the older and younger groups (Bull and Clifford 1984: 97) Furthermore, Huntley et al. (1987) found that listeners tend to focus their estimates around their own calendar age. There is, then, the real possibility that the non-linguists' may be providing age estimations that are unrepresentatively emblematic of older or younger speakers. With regards to physical condition, Braun (1996) points out that estimations of age may also be mediated by non-linguistic factors physical factors, such as those in good physical condition being seen as younger.

As a precautionary measure, and given the difficulties associated with estimating speaker age, Braun recommends that, it is better to use age spans or even general descriptions like 'very young' or 'middle aged' rather than predict specific ages.

2.12.5 Estimator variables: respondent provenance

Respondent provenance may be seen to affect the data in a number of different ways in terms of regional affiliations, linguistic experiences and psychological dimensions:

i/. Defining respondents by speech community: Much dialect geography research, dating back to early variationist studies, has employed researcher-selected communities such country, region, town, village or social class for categorising respondents, with the researcher determining at what level a region or speech community/social class is defined, and which label to use when referring to it. As Britain (2002:606) argues, the traditional variationist approach sees geographical space simply as a container (or a series of containers). For example, a traditional approach to language geography might focus on Wales and identify, Welsh English speakers, within which are South Welsh English speakers, South East Welsh English speakers or South Glamorgan (a county in South East Wales) Welsh English speakers, or Cardiff English speakers (Cardiff sits in South Glamorgan) or Splott versus Ely Cardiff English speakers (both are working class suburbs on the opposite sides of Cardiff). This *matryoshka paradigm* results in treating one speech community as

if it were nested inside a larger community, which is in turn nested in a larger community, ad infinitum, with each component speech community being assumed to be a group of homogeneous monolects.

However, speech communities are rarely homogenous or discrete; they are heterogeneous and highly complex. If we re-consider the example above, an arguably more realistic way, we should start by acknowledging the physical, social and perceptual factors (Britain 2002:609) that forge speech communities and shape their evolution. Language variation is mediated by more than mere Euclidean space. For example, geologically, Wales is almost entirely mountainous with severe restrictions on north-south urban connections. Where one might travel 150 miles east from the Welsh capital city of Cardiff to London in two hours by train, the 150 mile journey from South to North Wales is not possible by train, only by a four-hour car journey. The terrain is inevitably reflected in the language variation: the ancient Welsh language has never been standardised, with the dialects in the north and south showing very distinct syntactic, lexical and phonological features, to the extent that they lack a degree of mutual intelligibility. Moreover, other social factors such as the influence of Welsh monolingualism in the western half of the country, English monolingualism along the Wales/England border and Welsh/English bilingualism throughout much of Wales leads to a disparate linguistic repertoire for the population across the country, in both the English and Welsh languages (see, for example, Crystal (1995)).

Finally, colonisation and immigration into Wales have added to the complexity. pattern of linguistic communities. From the Norse invasions on Anglesey and South West Wales, to the Norman English colonisations of the 11th and 12th centuries which have brought English beyond the Welsh-speaking heartlands to the western extremes of the country furthest from England, to the industrial revolution and flood of immigrant workers into the North and South Wales valleys to work in the mines and on the docks, the English language has repeatedly washed over Wales, with each wave bringing different dialects depending on the regional origins of the English-speaking incomers. There is, perhaps inconveniently for researchers, no linear evolution of English in Wales,

a neat set of matryoshkas nesting one inside the other, and it cannot be assumed that 'region' defined at political, historical or linguistic levels will entail a collection of discrete but homogenous varieties that form sub-strata of one linguistic entity, 'Wales'. Such labels as used by researchers do not objectively capture the varying degrees and qualities of varieties of English in Wales. Garrett et al. (2003:74), nor do they reflect community-salient language variation. Consequently, for researchers to define their own linguistic boundaries, prescribe a label to their selected speech community and then present it as an attitude object to a group of respondents may, at best, be thoroughly misleading and worst render any findings invalid.

There is, in research on forensic speaker description, a case for not prescribing labels, but instead allowing respondents to make their own community-salient identifications of speaker origin according to the town, city, county, region or state they believe the speaker to come from. Not only can this avoid inaccurate researcher prescriptions, but it also offers an insight into non-linguists' own geo-social identity and affiliations, it can provide the researcher with the opportunity to map the salient community varieties from the non-linguists' perspective, and it may ultimately provide sharper, more insightful data.

ii/. Recognition of community accents: With regard to how well respondents perform at identifying other speakers from within their own community, Kerswill and Williams (2002) in their research comparing new (Milton Keynes) and old towns (Reading and Hull) in England, show that the correct identification of other community members' voices is much higher in socially stable communities. That is, identification of community members is improved in those communities which are focused (Le Page 1980), where clearly patterned variation or little linguistic change is exhibited. Kerswill and Williams have suggested that it is extended social networks that will be a decisive factor (as opposed to social class) in recognition of others from the speech community.

By contrast, in socially and linguistically diffuse (Le Page 1980) communities, such as Reading, where an older town has witnessed a large amount of urban development leading to a fast growth in the urban population, young natives

simply fail to recognise older natives, associating them with rural counties to the west (Kerswill 2004).

iii/. Geo-social self perception Respondent geo-social self-perception is not necessarily predicted by their home region or where they went to school, etc. Some of the varied manifestations of geo-social self-perception are evidenced in our casual interactions with strangers that we meet. Anecdotally, many British speakers will be familiar with meeting citizens from the United States who will happily inform an audience that they are, for example, 'Dutch-Italian' through family history and subsequently reveal that they have visited neither of those countries nor do they speak the respective languages. Furthermore, it is common for example in many Mediterranean countries to encounter people who will talk of their 'home village', or in Italy, 'home country', which refers not to the place where they have grown up, but to a village or town where their grandparents grew up and may still live. Given this, we should not assume that those who identify themselves as living in one town, county, region necessarily align their identity with that region, and we should equally not assume that somebody proclaiming to be from a country, region, etc has any significant experience of that place, its communities or its language. This will be relevant when discussing the results of regional estimations of speaker origin in the Discussion chapter (6).

iv/. Recognition of non-community accents: Sociolinguistic evidence strongly suggests that respondents' own geographical provenance may influence their evaluations of non-community speakers: van Bezooijen and Ytsma (1999) note the difference in recognition rates of Netherlandic and Belgian accents of Dutch dependent on the origin of respondents across Holland and Flanders in (Flemish Dutch speaking area of Belgium); Preston (1996) found differences in respondents' abilities to place local accents on a north-south transect across the United States; Williams et al. (2003) point to findings of varied recognition rates of Welsh teenage speakers dependent on the provenance of respondents across Wales; Kerswill and Williams (2002) report differences in the recognition rates of three English accents (Milton Keynes, Reading and Hull). Kerswill and

Williams (2002) have suggested that exposure to and recognition of different varieties of English will be mediated by many factors including:

a/. The respondents' geographical and social mobility
b/. The life experience of the judges (such as open-closed networks, geographical mobility)
c/. The sociolinguistic maturities of the judges (mostly related to age)
d/. The salience of the differentiating features between the varieties, or in Preston's terms: their availability for perception and comment and the accuracy and detail with which they are perceived (Preston 1996)
e/. The degree of contact between the respondents and the broadcast media, by which familiarity with varieties is spread
f/. The degree to which these varieties feature in the broadcast media

We could also add to this:
vi/. The proximity of their own region of origin to other regions

It is the feeling of this researcher that studies into forensic speaker description must acknowledge where respondents are from and must also assess their performance recognising other non-community regional and social accents if an integrated understanding of how earwitnesses can perform is to be collated. Such data can also inform wider sociolinguistic practice and go some way to improving on the current paucity of research on the effects of respondent origin on non-community voice recognition in the UK.

2.12.6 Estimator variables: in and outgrouping

The basis for identification and evaluation of speaker and their origin is intersected by a set of cognitive biases related to in and out grouping. Respondents may claim ownership of or distance themselves from speakers, identifying them as part of their *ingroup* or distancing themselves as part of an *outgroup*:

i/. Ingrouping and outgrouping by social attractiveness: Subjectively perceived social attractiveness of the speaker due to suprasegmental factors (voice quality, tempo, pitch range) (Kerswill and Williams 2002), or speakers who are perceived as 'likeable', on the basis of the content of their speech sample, such as an amusing anecdotal voice sample, can be 'claimed' as belonging to the judge's own regional group regardless of the actual provenance, as was seen with teenagers' evaluations of Welsh peers (Garrett et al. 2003), discussed on p. 41.

In contrast, misattributions of voices to different regions may be the result of low social attractiveness or displeasing content on the voice sample, with respondents rejecting the speakers as part of an outgroup. Drawing on the literature from language attitudes research, it has been established that certain accents are frequently favoured or disfavoured. Should an accent be mistakenly identified for another area, it may well be the case that an error in the social categorisation of the speaker leads to the evaluations being derived from the invoked label rather than an evaluation of the accent itself.

ii/. Pronunciation-based in- and outgrouping and evaluative consequences: Alternatively, respondents may identify themselves as sharing a speech variety with the speakers. One of the possible effects of this respondent affiliation is that members of the ingroup may subsequently be evaluated more positively in terms of psychological dimensions; that is, in-group membership in itself has the power to make a voice appear more attractive. In-group loyalty can then be evidenced in the consistency of the positive evaluations across the evaluative dimensions such as intelligence, etc.

The opposite effect may involve speakers who will be happy to outgroup speakers who they know factually to be from the same region or to present a pronunciation that is clearly from their home region. A respondent with a strong regional accent may reject a speaker who they know to be from the same area on the basis of them presenting a strong RP accent and sounding 'posh'; alternatively, a respondent with a strong RP accent who recognises the speaker to be using the regional accent that is particular to

their home region may reject them on the basis of social class, perceived lack of education, sophistication. The subsequent evaluations of the speakers may be affected by an outgroup homogeneity bias. An illustration of this concept is the example from popular culture, where in science fiction universes, such as Star Trek, humans are almost always the only species with multiple cultures; alien species are almost universally depicted as having a single, species-spanning culture, seeing members of their own group as being relatively more varied than members of other groups.

Inverted examples of in and outgrouping biases come when speakers of one variety come to evaluate their own in-group speech more poorly than that of the outgroup. For example, (Plichta 2001), found that when rating European and Afro-American speakers reading a passage, European Americans rated the Afro-American as equal for standardness of speech, yet the Afro-Americans rated other Afro-Americans as lower than their European counterparts for standardness (Thomas 2002:130). This can be compared with the 'cultural cringe' identified in New Zealand by Bayard (1995), whereby speakers of New Zealand English came to dismiss their own variety of English as inferior to the varieties of other countries.

An outgroup homogeneity bias may have the consequence of facilitating the social stereotyping (positively or negatively) of the outgroup. Alternatively, in terms of, for example, making attributions of guilt, a group attribution error may occur, whereby group members are more likely to attribute a fellow group member's actions to their arbitrary circumstances, while attributing a non-group member's actions to something in that group's inherent disposition. The cognitive biases of in and outgrouping have a particular role to play in the design of a study and the interpretation and discussion of the results, and will be referred to in Chapter 6, Discussion.

2.12.7 Estimator variables: the role of social stereotypes

> *Although, all language varieties are in themselves equal, the patterns or features of a variety can connote an additional layer of social meaning for the lay language user. Linguistic forms, varieties and styles can set off intergroup beliefs about a speaker and their group membership, and can lead to assumptions about attributes of those members.*
> *(Garrett, Coupland and Williams, 2003:3)*

A theme that has run parallel to many of the above sections has been the role of social stereotypes, which may be evidenced particularly as an aspect of in and outgrouping as underlying bases for the affiliative and distancing processes. The questions addressed in this section relate to accessing and triggering social stereotypes in the context of language research, and the forms that social stereotypes might take, especially in the context of forensic linguistics, as they related to a speaker's regional or social background.

For both non-linguists and professional phoneticians involved with speaker description, the identification of a speaker's region or social background is unlikely to be based on the utterance as a whole. Identification is likely to focus on certain prosodic and segmental features and how, for example, selected consonantal or vowel realisations are distinctive (Kerswill and Williams 2002). Which particular realisations are prominent may be related to the strength or stereotypy of the accent. Which particular realisations are attended to, which are ignored, and the relative weighting of each feature are a matter of the non-linguist's personal experience. What researchers must bear in mind, however, is that there is no guarantee that, in terms of accent origin, the 'standout' features will be correctly indexed and identified by the listener. It may be that one particular realisation of a word may assist respondents in mobilising a stereotype and perhaps identifying the speaker's region. Equally, we must remember that what the researcher considers to be standout stereotypic features may be invisible to the non-linguist. Issues of Preston's (1996) *availability* will always pertain. Nevertheless, whatever standout/stereotypic features are salient to the respondent are likely to form the basis for subsequent evaluations of the speaker.

2.12.7.1 Conscious and unconscious processes

The triggering of social stereotypes in the minds of non-linguist respondents can be evidenced by the patterns to be observed in the way that non-linguists index, respond and behave towards what they hear, and can involve both conscious and unconscious processes, examples of which are:

i/: The respondent hears a voice and makes a conscious association between voice > speaker origin > speaker social group > social stereotype. This may be most manifest in what Devine (1989) refers to as high prejudice people – those likely to hold personal beliefs that overlap substantially with a cultural stereotype. Responses to such stereotype activation may present examples of what Devine refers to as automatic processes, whereby the stereotype has a history of recent or repeated activation in memory, is available or requires little conscious effort to be invoked and is inflated by the presence of stimulus cues.

ii/. The respondent hears selected parts of speech, vowels, consonants or paralinguistic features which are seen as emblematic of the speech of a particular social group, triggering social stereotypes related to that group, which are then attributed to the speaker as a perceived member of that social group. The subsequent judgements are not evaluations of the speaker per se, but responses to the characteristics of the social group to which they perceive the speakers to belong (Williams and Garrett 2002). For example, we can take the example of the well known shibboleth in England of [ɑ:] versus [æ] in such words as 'bath, grass, glass' which is seen to distinguish many southern speakers from northern speakers, and has been part of popular media comment and comedy sketches in the United Kingdom for many years, with speakers from the north and the south parodying and satirising each other's respective pronunciations. An example of this might involve a speaker from the south having their pronunciation of this phoneme seized upon in the presence of northern English speakers, experiencing accusations not only of being from the 'south', but with that accusation comes the secondary meaning of 'posh' = privileged social background, and therefore socially unattractive; this is a

frequently occurring example of geo-social othering, whereby persons tend to stress what makes them dissimilar from or opposite to another, carried over into the way they represent others, especially through social stereotypes.

Moreover, Milroy and McClenaghan (1977) note that accent is assumed to act as a cue in identifying a speaker's group membership but also suggest that this identification may take place below the level of conscious awareness, presumably by hearing similar accents very frequently or having heard similar accents recently, through which one learns to associate them with their reference groups. In other words, accents with which people are familiar may directly evoke stereotyped responses without the listener first consciously assigning the speaker to a particular reference group. This is closely related to *iconicity* (Irvine 1996), the transfer of linguistic features to social facts, whereby the presumed social attributes of a group are transferred to linguistic features associated with it, an occurrence of which may directly trigger recognition of those attributes without even being filtered through a conscious identification of the group. That is, stereotypes can be primed automatically through procedures that produce attentionless processing of primes (Devine 1989).

2.12.7.2 Acquisition of stereotypes as part of socialisation

Van Bezooijen (1994), in considering the acquisition of social stereotypes and their linguistic cues, investigated the context of Dutch in the Netherlands and revealed an apparent acquisition order of social connotations for various Dutch varieties: Seven year olds, whilst able to distinguish between the three Dutch varieties presented to them, decided that women were either housewives or farmers' wives regardless of the language variety they speak, making no aesthetic differentiation between varieties on the basis of social connotations of accent. Yet even at the age of seven, van Bezooijen observed children giving higher aesthetic ratings to the standard variety, which she attributed to some effect of the imposed norms hypothesis, which predicts that positive aesthetic evaluations of the standard variety are not the result of their inherent value, but the result of cultural norms. For 10 year olds, social connotations start to kick in

with farmers' wives being associated with 'rural' accents and a strong association observed between standard varieties and high status occupations. A comparison with adults showed that a full 'set' of social connotations for these language varieties is in place. This is compelling evidence that not only are social stereotypes more likely to be acquired than learnt through direct experience, but also that any language attitudes research involving adults is highly likely to involve investigating the effects of a range of social stereotypes upon evaluations of regional and social language varieties.

2.12.7.3 Accessing social stereotypes

With the above in mind, it is useful to the researcher to allow respondents in a forensic speaker description exercise to elaborate qualitatively on how they have arrived at their conclusions as to who the speaker is, where they are from and what this means to them. Some considerations when accessing social stereotypes are:

i/. Inter-respondent variation: Stereotypical evaluations may not be predictable or accessible from one person to the next (Preston 1996; Devine 1989; Gilbert and Hixon 1991; Ladegaard 2001). Researchers must be cautious in assuming that that all cultural ideas and stereotypes floating around any given society are complete units and identified in the same way by all members of society, or that possession of the stereotype is a binary choice, in that either a non-linguist completely has or completely does not have the stereotype. This is because awareness of certain aspects of a social stereotype does not automatically entail awareness of the complete stereotype; neither does it automatically follow that two respondents who see the same social object, have the same vantage point on it. For example, imagine that three non-linguists, 'A', 'B' and 'C', see social stereotype 'S', which is made of component 'X', 'Y' and 'Z'. It may well be that 'A' knows or believes 'X' and 'Y', 'B' knows or believes 'Y' and 'Z' and 'C' knows or believes 'X' and 'Z'. It would be convenient to treat the social stereotype as homogenous and embraced by all three respondents, yet we must remember that reality is often inconveniently more complex. In reports

relevant to forensic speaker description, from earwitnesses or the police, this may be of great relevance.

ii/. Intra-respondent variations: Much as the same people will speak in different ways according to their situation, intra-respondent variations in cognitive processes may affect evaluations (Devine 1989; Bradac et al. 2001). Within one respondent, attitudes and their attendant social stereotypes are not a static, singular phenomenon. Rather, they affect and are affected by numerous elements in a virtually endless, recursive fashion (Cargile et al. 1994). As above, we must remember that in the context of crimes taking place, stress and trauma may heavily influence the reports given by earwitnesses.

iii/. Cognitive biases: Speakers may behave in certain ways to fulfil certain self presentational goals (Cargile et al. 1994). The conscious suppression of stereotypes, the effects of personal impression management and other cognitive biases can skew responses, making prediction and interpretation of non-linguist responses difficult.

2.13 Estimator variables: language evaluations and the attribution of guilt

The final aspect of estimator variables to be discussed is central to research on forensic speaker description: the attribution of guilt. Language trait focused discrimination (Lippi-Green 1994) is a widespread phenomenon which permeates our day-to-day existence. Language may systematically (dis)advantage speakers in an institutional setting (Kalin 1982) and language prejudice is in no way discouraged by a judicial system that maintains a blind adherence to a standard language ideology (Lippi-Green 1994). Given that those charged with administering justice - legal professionals and the police - are drawn from the ranks of the non-linguists, it is reasonable to concern ourselves with the manifestations of non-linguist ideological evaluations in the legal context.

In this section, some of the findings of research into non-linguists' language evaluations in the legal context will be spelt out, illustrating some of the ways in which social stereotypes and non-linguist attitudes may impact upon attributions of guilt.

Dealing first with the context in the courtroom, in examining the effects of legal testimony on a court audience, Lind and O'Barr (1979) point to the reactions of jurors to variations in the manner of speech. They argue that jurors' acceptance of a communicator can be influenced by judgements regarding the communicator's competence and trustworthiness. 'Competence' they interpret as 'the ability to make valid statements in the legal context', and 'trustworthiness' as the intention 'only to give statements that they consider truthful'. Of note here also is that Thomas (2002) summarises much research in the psychology field, suggesting that impressions of competence are strongly related to speech rate - the faster rates being directly correlated with impressions of higher competence.

Lind and O'Barr also make mention of two other dimensions familiar to researchers into language attitudes, 'attractiveness' and 'social dynamism', with strong echoes of Giles' (1970) original study into British language attitudes through to Zahn and Hopper's (1985) Speech Evaluation Instrument and subsequent language attitudes research. If Lind and O'Barr's observations are correct, they argue, then considerable social evaluation occurs in the courtroom, and there is considerable justification for supposing that the evaluation of the speaker's credibility and ultimately guilt may be influenced by the manner in which they represent their communication to the court.

Returning to the issue of whether attributions of guilt may be mediated by social stereotypes and language attitudes, clearly the repercussions for the criminal justice system cannot be understated. Perhaps more significantly, if justice can be swayed in a carefully controlled court setting, in which evidence has been filtered, witnesses 'coached' and the communication between all parties overseen and strictly controlled, what of the build up to the trial, the preparation of the case, the compiling of the evidence, the arrest of the suspect and the

interviewing of the witnesses? Is it to be assumed that the speech evaluations discussed by Lind and O'Barr will only emerge in a court setting?

Lind and O'Barr's findings have been shored up by research focused specifically on attributions of guilt which has confirmed that language is one of the extralegal factors that may affect forensic impressions, attributions and decisions (Dixon & Mahoney 2004). Examples of this research include:

i/. Diverging from or converging to the standard: Dixon et al.'s (1994) study mentioned earlier in this chapter employed an MGT to researching the relationship between speech accommodation and guilt. They confirmed predictions that a suspect diverging from English to Cape Afrikaans would be rated guilty more often by white English-speaking judges than a suspect converging to English. This, the authors state may have reflected the judges responding to stereotypical perceptions about the 'Coloured' (the authors' term) group. Dixon et al. (2002) cited a possible relationship between non-standard speakers giving testimony and a reduction in self-assuredness, a corollary of which was increased attributions of guilt.

ii/. Causal attribution: Judges may evaluate and describe voices or attribute guilt as a function of causal attribution (Smith and Mackie 2000), whereby they seek to explain the reason for behaviour they encounter, often invoking common-sense explanations. For example, in the forensic research context, respondents' evaluations of speaker guilt may be mediated by their ability to invoke a reason for the speaker's behaviour. In countering this effect, researchers should consider the content neutrality of the texts used.

iii/. Crime type: The contextual variable of crime type can exert a significant effect on attributions of guilt (Dixon et al. 1994). Seggie (1983) found that accent interacted with crime type to shape assessments of guilt: He found through an MGT task that certain accents, such as Australian and RP, mediated attributions of guilt as a function of crime type. For example, the crime of embezzlement was associated by Australian judges with the RP speaker, and crimes of violence and vandalism were more strongly

associated with an Australian speaker. A third 'Malay Chinese' voice had less distinct social stereotypes associated it in the context of Australian judges.

There is sufficient evidence to suspect a definite relationship between speech varieties and attributions of guilt. However, there is still a need for greater breadth of research and a deployment of methods other than MGT, and in particular a closer examination of the relationship between speaker accent and attributions of guilt. With this in mind, there may be interesting benefits to investigating not only the judges' estimations of speaker origin and the descriptions of speakers in the judges' own words, but also how a variety of crime types and accents within a given community may interact, and what role other evaluative dimensions such as those used in general language attitudes research - social attractiveness, prestige - may play in this process.

2.14 Research questions

2.14.1 The known and unknown

The journey through Chapters 1 and 2 in this research has set the scene in terms of the context for forensic linguistics and forensic speaker description. Chapter 1 summarised the context, existing research and practice amongst forensic linguists and noted how, despite increasing interest and progress in the field of forensic speaker *identification*, research has largely stopped short of addressing forensic speaker *description*, a situation which serves no purpose to those who find themselves the witness to a crime. Chapter 2 drew, in the absence of any specific forensic speaker description literature, on the socioperceptual literature, presenting findings from research into folk linguistics, sociolinguistics, social psychology, phonetics and language attitudes, and the chapter set out issues for consideration when designing research into forensic speaker description.

The ultimate aim of this thesis is to further our understanding on what may form the basis for an audiofit. In order to achieve this, we would ideally have evidence that non-linguists are able to notice characteristics and provide a

description of the voice and accent of unfamiliar voices. Whilst Chapters 1 and 2 have presented and discussed findings from the diverse available literature, they have also highlighted the many gaps. It is clear that many aspects of the non-linguist perceptual and reporting mechanisms have yet to be explored and there are many unknowns to be attended to before the field can move forward. It is unclear, for example, what features of the voice and accent are salient to non-linguists (compared with, for example, the phonetic and voice quality features identified by trained professional linguists) and it is unclear how non-linguists might encode and index these features. There is little research on what vocabulary non-linguists might mobilise to report their observations. With no training in the specific terminology for phonetic and voice quality description, we cannot know whether non-linguists have sufficiently rich lexis to be able to convey their descriptions of phonetic and voice quality. Should non-linguists be given the opportunity for free description in an audiofit or do they require some kind of structured framework guiding them through points of the voice and accent to consider? Should non-linguists rely on their own lexis or should a typology of possible voice/accent qualities be provided to facilitate their descriptions? Should the descriptive terms be those used by the professional linguists or should more user-friendly terms be employed? If so, from where can we derive these terms? Are non-linguists' own terms sufficiently unambiguous and accurate to be meaningful?

The literature has also shown us that non-linguists are unlikely to encounter voices without making subjective evaluations of the speaker. Yet we do not know whether the phonetic and voice qualities of the speaker will take precedence over the non-linguist's evaluative judgements of the speaker's social group, affording a description of the individual, or whether the descriptions of individuals will unavoidably entail descriptions of the stereotypic voice, accent and personality of the speaker's perceived social group. It is unclear whether non-linguist evaluations impair the non-linguist's ability to perceive and report voice and accent detail. Do non-linguists' observations reflect subjective or objective observations? Can non-linguists provide linguistic description aside from evaluative judgements of the speaker?

With regard to accuracy of accent identification, the folk linguistics literature has shown that there is potential for non-linguists to provide useful estimations, although these will inevitably be weaker than those provided by the professional trained linguist. We do not know whether non-linguist estimations of regional or social provenance may be too vague or inaccurate to be meaningful. What is also uncertain is whether, with the non-linguist architecture of accents being constructed using a good/bad English paradigm with stigma attached to regional variation, non-linguist estimations will be sufficiently objective to be of any use in the investigative process. The existing literature has yet to address what effect, if any, the non-linguist's own regional and social background may have on the accuracy of their estimations for speaker provenance.

This chapter concludes by summarising these unknowns in to a set of research questions to be addressed in the following research sections of this thesis.

2.14.2 Research questions

Estimations of speaker origins and age

How well do a group of non-linguists recognise and estimate the linguistic backgrounds and ages of unknown speakers with selected British regional and social accents?

What factors may influence their estimations?

Evaluation of speakers

How do a group of non-linguists evaluate the selected speakers?

To what extent do their perceptions, descriptions, evaluations and attributions of guilt interact and what does this tell us about non-linguist ideology?

Evaluation of the speaker voice features

When invited to provide qualitative comments on voices, what descriptive vocabulary do non-linguists provide?

When provided with an audiofit of non-linguist derived voice features, how do respondents describe the selected speakers?

By what other means might non-linguists articulate their perceptions of voices?

3 - METHOD

3.1 Introduction

The research in this thesis is designed around eliciting respondents' impressions and descriptions of unknown voices, and how their subjective social evaluations may interact with accent and voice qualities. The underlying concept at the heart of the methodology for this research is the requirement to deploy direct approaches (for ecological validity), involving speech samples being played to respondents. The methodology attempts to control all but the manipulated independent variables that relate to speech variation in order to explain the patterns of response amongst listeners. At the same time, the research design sets out to assuage any potential response biases that may affect the validity and reliability of the findings.

This chapter sets out the design of the research, explaining the selection and background of the respondents (Section 3.2), the selection of the materials including the voices samples (3.3), and the procedure (3.4). The research was conducted in two Stages, Stage 1 and Stage 2. Stage 1 acted as the basis for Stage 2, with data gathered and processed prior to the design of Stage 2. I will give a general overview, addressing both Stages together where relevant.

3.2 Participants: Stages 1 and 2

The participants were all undergraduate students, living in South Wales and studying at Cardiff University. These people were not formally selected but were a convenience sample, based on availability and willingness to participate. The participants in Stage 1 and Stage 2 were selected from different cohorts one year apart, although they shared a similar profile, explained below.

3.2.1 Respondents' linguistic background: Stages 1 and 2

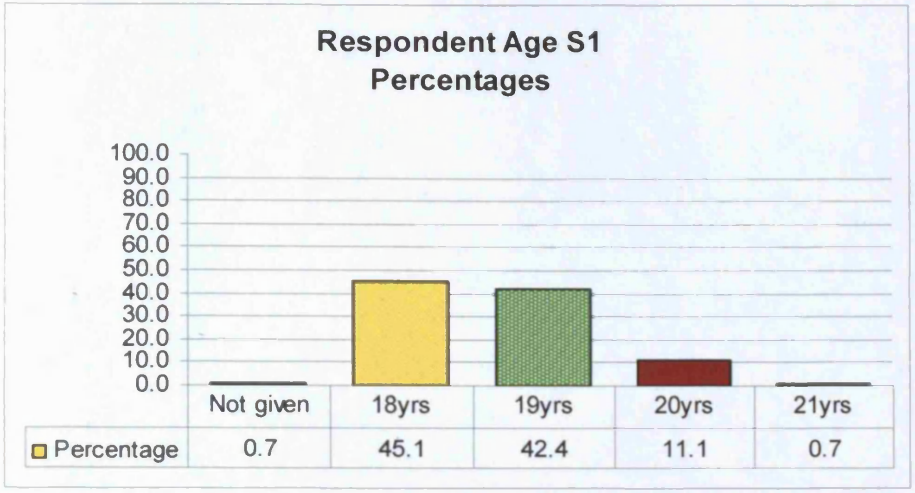
Those judges whose responses were included in the data were native English speakers from the United Kingdom who were just embarking on the first term of an undergraduate degree in Language and Communication. All respondents who wished to participate were invited to do so. However, those who were new arrivals to the UK were likely to have limited ability for recognising British regional and social speech variation, and perhaps greater difficulty associating British English speech patterns with the usual social stereotypes. Therefore, non-native English speakers from outside the UK were excluded from the final data analysis.

A few other participants were native to the UK and functionally bilingual, in English and Welsh. The multilingual status was not deemed to be preclusionary, as additional languages were unlikely to impinge on their native speaker competence in English. Likewise, respondents' ethnic background was not deemed of sufficient relevance to warrant investigation.

3.2.2 Respondents' ages: Stages 1 and 2

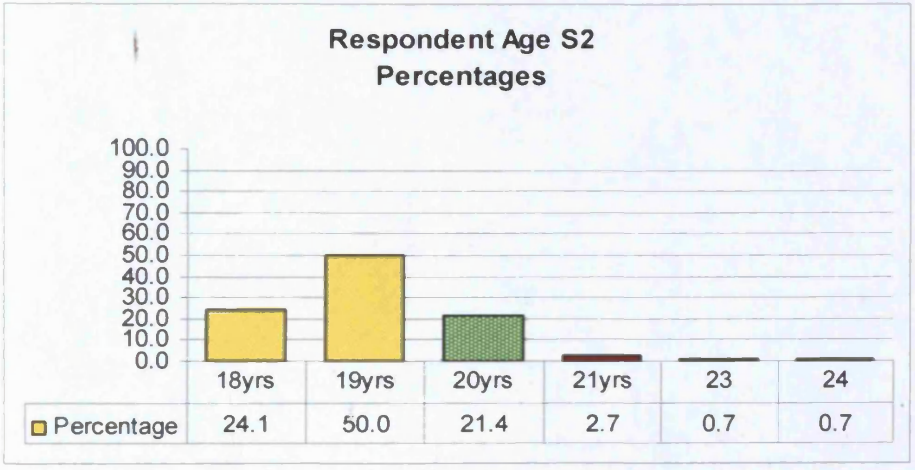
The findings from Braun (1996) and Bull and Clifford (1984) (Section 2.12.4) suggest that for speaker identification research, respondents over 16 and under 40 may yield the most reliable results. Given the lack of research in the field of forensic speaker description, this guidance was applied to the current research, and all participants were aged between 18 and 24. In Stage 1, 98.6% of the participants were aged between 18 and 21; in Stage 2 the figure was 95.5%. This would suggest that in terms of age, the participants in both groups were quite homogenous. A breakdown of the ages for respondents in Stages 1 and 2 is given in Figures 3.2.2.1 and 3.2.2.2

Figure 3.2.2.1 Respondents ages, Stage 1



N=144

Figure 3.2.2.2 Respondents ages, Stage 2

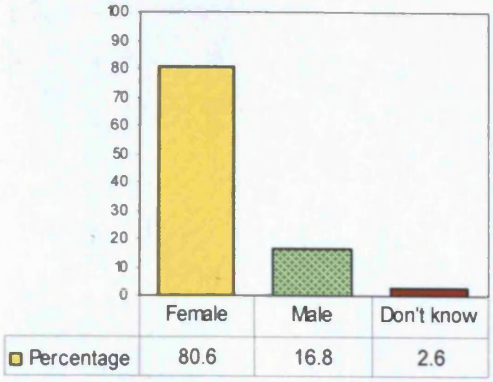


N=112

3.2.3 Respondents' sex: Stages 1 and 2

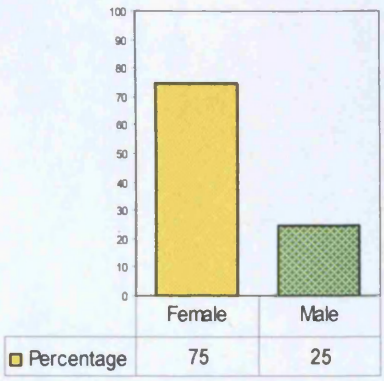
The convenience sample of respondents included a larger proportion of females to males. Following evidence from Braun (1996) and Hollien and Schwartz (2000), suggesting no difference in respondent performance in speaker identification tasks as a function of respondent sex, no distinction was drawn between the responses of either sex. A breakdown of the figures for respondent sex is given in Figures 3.2.3.1 and 3.2.3.2.

Figure 3.2.3.1 Respondent sex, Stage 1



N=144

Figure 3.2.3.2 Respondent sex, Stage 2



N=112

3.2.4 Respondents' educational background: Stages 1 and 2

Some of the students indicated that they may have had some experience in English language studies prior to entering university, for example, taking the English/Welsh 'A' Level examination in Communication/English Language, usually undertaken as one of the optional qualifications to gain entry to a university. It was decided that although these courses can act as a first step to a linguistics-oriented degree, including some training on awareness of language variation, the introductory experience would be too elementary to inform the respondents as to the forensic, sociolinguistic and methodological issues involved in this research. Therefore, their responses were unlikely to have been contaminated by existing knowledge from previous studies. Informal post-research interviews confirmed this to be the case.

3.2.5 Respondents' regional background: Stages 1 and 2

3.2.5.1 Stage 1

Section 2.12.5 (p.112) discussed respondents' regional background, with evidence from Preston (1996), van Bezooijen and Ytsma (1999), Kerswill and Williams (2002) and Williams et al. (2003) suggesting that respondents' origin may significantly influence their ability to recognise non-community accents. Accordingly, it was felt important in this research to log where the respondents were from. However, in a criticism of previous research, Garrett et al. (2003) observed that the researcher-provided labels with which respondents can

indicate their provenance are not always community-authentic (Section 2.8.1, p.76) and may therefore not be recognised or salient to the respondents. To accommodate this criticism, it was decided to allow respondents to self-define their origins, as geo-social self perception (p.112) is not always predicted by their place of current residence. On the questionnaires, each respondent was asked to provide anonymous information regarding where they were from. No place names were suggested, rather, respondents were invited to self-select their geo-social affiliations, allowing them to write the name of a town, county, region or country that they identified as their origin. Such an approach allows the researcher to categorise the data in terms of non-linguist perceptions 'on the ground' rather than imposing non-community-salient labels.

In Stage 1 (n=144), some 92.5% of respondents were from areas that are regionally adjacent to Cardiff and within two to three hours' drive from the city. These adjacent areas from which such a large majority of the respondents were drawn can be identified as the catchment area for Cardiff University. A further 6% from the North/North West of England/Scotland, who are not considered to be part of the expected catchment area for the university. 1.5% did not answer the question on regional origin.

The next stage was to decide how the catchment area could be sub-divided for analytical purposes, as the research intended to investigate any interactions between the non-linguists' responses and their regional origins. This is not a straightforward task. Aside from the countries that make up the United Kingdom – England, Northern Ireland, Scotland and Wales, there exist no official regions, only the administrative city, county borough and county division boundaries. Furthermore, the political boundaries have been realigned many times under the *Local Government Act 1926, the London Government Act 1963, Local Government Act 1972, Local Government Act (Wales) 1994* and from 1995-1998, there were created forty six new Unitary Authorities in England and Wales. (http:// www .lga .gov. uk/l ga/the_lga /Structure2.pdf). This has meant that for the purposes of this research, the current local government divisions are less than helpful for describing the linguistic and cultural regions of England and Wales as they are based on recent political decisions, and are unlikely to

represent cultural understanding of regional divisions, dialectological divisions or have linguistic community salience. Moreover, the alternative of removing modern political boundaries and focusing on dialect boundaries, presenting a map with only isoglosses reflecting historic linguistic communities is also likely to be of little community salience as non-linguists' perceptions of dialect variation differ from that of linguists (Williams et al. 1996, discussed in Section 2.3.1 (p.45)).

What is needed is a regional sub-division that is both community-salient yet reflects linguistic regional variation. The starting point is to consider media divisions in the country, particularly in television. Although not officially divided into regional political entities, the country's television media is subject to (ultimately government derived) regional divisions. These reflect long-standing community divisions and identities, and many linguistic realities (see below). Residents receive localised versions of the national media, with news programmes, and specially localised programming, being broadcast on both state (BBC) and independent (ITV) television stations. One could argue that for anyone with a television, the experience of living in a UK region is, to a certain extent, identified by the local news channel you receive and the geographical areas that your news refers to. Figure 3.2.5.1 (BBC) and Figure 3.2.5.2 (ITV) show the two versions of regional broadcasting in map form.

We can see from both maps that Wales is defined as a separate entity; adjacent to Wales, a West/South West region is identified, as are a Central/Midlands region and an East/Anglia region. For reasons of population density, London has its own regional service; however, it abuts the South East region and the two maps share a London/South East region of similar shape and size. These regional divisions are of course not strict dividing lines, but they serve as a useful starting point, employing regional divisions that are used in common parlance and easily recognisable as reference points throughout England and Wales. It must be acknowledged that one effect of partially localised media is that it may reduce awareness of other regions, concentrating the viewer's attention on local issues. However, the priority for the analysis of data is to

categorise the respondents' origins according to regions/labels that are salient to themselves and is not, therefore, considered problematic.

Figure 3.2.5.1 BBC Map of British Regions



http://news.bbc.co.uk/nol/shared/spl/hi/uk/05/born_abroad/around_britain/html/map.stm.

Accessed 21/2/07

Figure 3.2.5.2 Map of ITV Broadcasting Regions



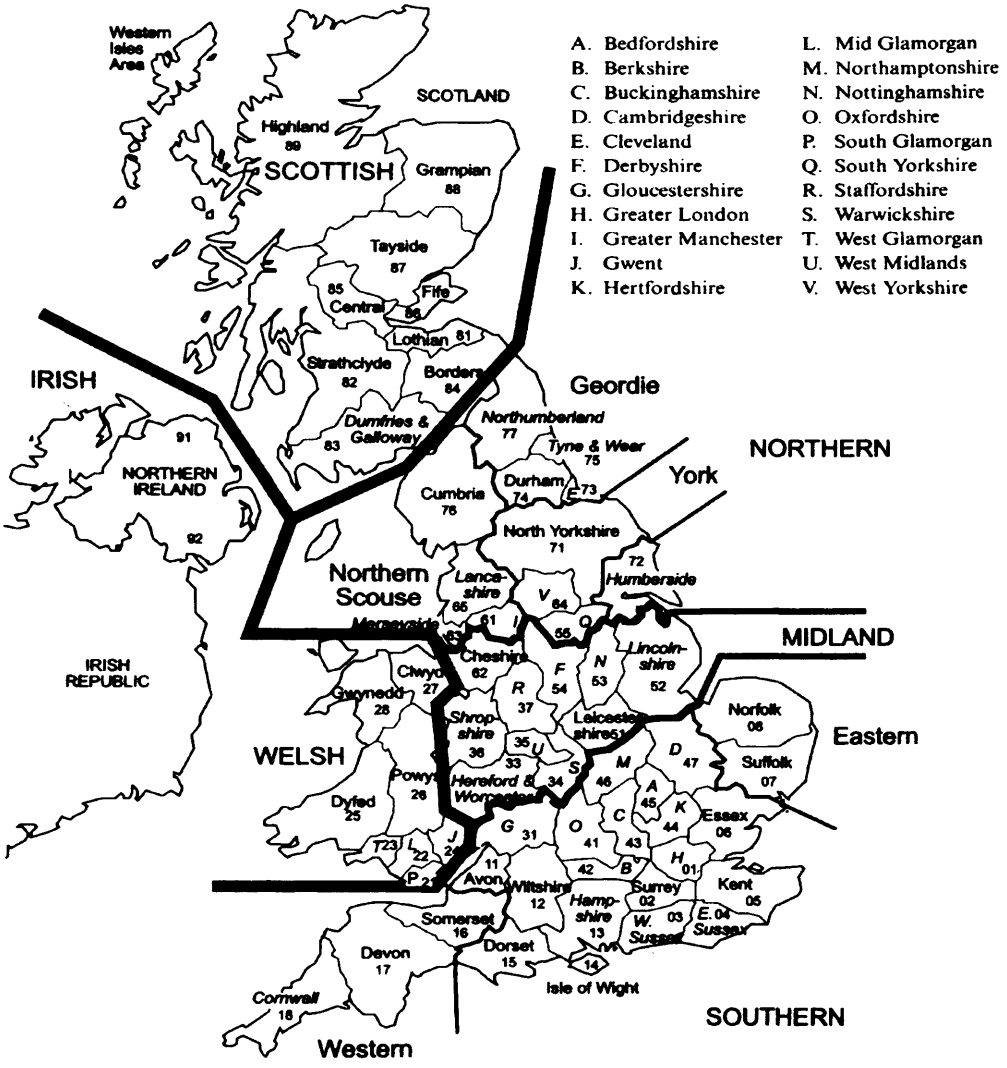
http://www.digitaltelevision.gov.uk/dig_switchover/map.html.

Accessed 20/2/07

In order to integrate non-linguists' perceptions of linguistic divisions with the media maps, the media-derived regional divisions were then compared with Inoue's (1999) perceptual mapping study (discussed on p.44), which asked students to describe where they thought linguistic boundaries may appear in Great Britain. Figure 3.2.5.3 shows the composite map for the whole group of Inoue's respondents. What we can see in Inoue's map is thick black lines delineating major dialect areas (Wales vs. England), medium thick lines delineating subdivisions of major dialect areas (Northern, Midland, Southern) and thinner lines delineating smaller subdivisions e.g., Western Southern, Eastern Southern. What is interesting is that the linguistic regions for his non-linguist respondents, in generalised terms, reflect the local media regions for the southern half of the UK. We see the identification of Wales, of a west/south west region, a central/midland region, a south/south eastern region, with an

eastern annex. We can note that although 'London' may exist as a broadcast region for practical reasons, in the eyes of the broadcast media, linguistically, it is not seen as an exclusively different region from its immediate surrounding.

Figure 3.2.5.3 Perceptual map of Great Britain



Source: 'Subjective dialect division in Great Britain,' in D. Preston (ed.), Handbook of Perceptual Dialectology, Vol. 1, pp. 161-76.

One can observe also, that the regions represent well-established regional varieties known to linguists: Wales, identified in the media divisions and by non-linguists, is known to linguists as containing distinct varieties; the identified South East region relates to the well-documented spread of Estuary English; the identified South West relates to a region that is distinct for its retention of rhotic pronunciation, amongst other features.

Observing the commonality between the media maps, Inoue's map of non-linguist estimations, and linguist-identified regional varieties, provided a basis for sub-dividing the Cardiff University catchment area. An approximated composite map was generated (Figure 3.2.5.4), focusing on the southern half of the UK, where 92% of respondents included in Stage 1 of the research had identified themselves as being from. The common regions to be identified and plotted on the map were Wales, South West, South East and Midlands, but the decision to be made was exactly where, given the variation between all of the contributing maps, to place the regional borders on an idealised map. Wales was in this respect simpler, as it is identified as a political entity and its borders were based on real historical, political, linguistic and cultural ones. However, for the English South West, South East and Midlands areas we recognise the sub-divisions are more complex thanks in part to the lack of official political regions.

The suggested resolution to this problem was to refer back to the respondent non-linguists themselves and see which regional areas they allocated their home towns to. For Stage 1 of the research, in their responses and post-response informal group debrief, respondents supplied detail not just with regard to their home towns, but also which region they identified themselves as coming from: as we move across the country, we find respondents from Cheltenham, Gloucester, Bath and Glastonbury indicating they were from the South West, but respondents from Southampton, Portsmouth, Winchester, Reading and Oxford indicating they were from the South East. (This division is again corroborated by our knowledge of the spread of Estuary English.) A regional border was therefore drawn between these localities to represent a generalised South West/South East isogloss. Likewise, respondents in Oxford and Cambridge identified themselves as being from the South East whilst respondents from Birmingham and Stratford identified themselves as being from the Midlands, and a South East/Midlands isogloss was plotted between the locations. Finally, the Midlands isogloss was extended to differentiate the Birmingham/Stratford respondents from those from Cheltenham and Gloucester.

The composite map in Figures 3.2.5.4 is a result of the subdivision process, derived from local media regions, non-linguist perceptions of regional variation in English and non-linguist reports of the regions in which their home towns belong.

Figure 3.2.5.4 Regional division of Cardiff University catchment area



A final few outstanding points of clarification remain: It is possible, given the northern limit of the Midlands border in both the BBC map and Inoue's map, that the Midlands region could be extended further northwards. However, as there were no respondents from this potential north Midlands to identify themselves one way or the other, the northern border will remain the subject of future studies. The respondent from Essex was from the southern strip which forms Greater London and so was identified as South East. Finally, although all of these isoglosses have loosely defined regional borders, it was anticipated that their spatial positioning adjacent to each other would alleviate any ambiguity if presented to non-linguists. Below is a table listing all of the provenances provided by the respondents in Study 1, after which, the numerical summary is provided. I have also included the names of the local regional broadcaster, to facilitate a comparison between media labels and 'folk' labels.

Table 3.2.5.1 Respondents' identified places of origin

WALES	
Town:	Cardiff, Barry, Swansea, Newport
County:	---
Region:	South Wales, Mid Wales, Welsh Border
Country:	Wales
Regional TV:	BBC Wales, ITV Wales
SOUTH WEST (ENGLAND)	
Town:	Bath, Bristol, Cheltenham, Glastonbury, Gloucester, Exeter, Plymouth, Western-super-Mare
County:	Cornwall, Devon, Dorset, Gloucestershire, Somerset,
Region:	South West, West
Regional TV:	BBC West, BBC South West, ITV West
MIDLANDS/CENTRAL (ENGLAND)	
Town:	Birmingham, Stratford
County:	Herefordshire, Shropshire, Worcestershire
Region:	Midlands, East Midlands
Regional TV:	BBC Midlands, Central
SOUTH EAST (ENGLAND)	
Town:	Brighton, Cambridge, Eastbourne, London, Milton Keynes, Oxford, Reading, Portsmouth, Southampton, Winchester, Windsor and Eton
County:	Berkshire, Essex, Hampshire, Hertfordshire, Kent, Surrey, Sussex
Region:	South East, Estuary, Home Counties
Regional TV:	BBC South East, Meridian
Outside Cardiff University's catchment area, not in final analysis (7% of respondents):	
NORTH WEST (ENGLAND)	
Town:	Liverpool, Manchester
County:	Lancashire
Region:	North West
NORTH (ENGLAND)	
Town:	Leeds
County:	Yorkshire
Region:	North
NORTH EAST (ENGLAND)	
Town:	Newcastle
County:	----
Region:	North East

SCOTLAND	
Town:	---
County:	---
Country:	Scotland

Of note here is, firstly, that those who chose to describe their origins by town seem only to choose towns, cities, or in one case a village which are well known. (The town of Glastonbury is world famous for its music festival). This suggests an awareness on the part of the respondents that only naming large/well-known places is a co-operative response, presumably in that it will assist the interlocutor in understanding their provenance.

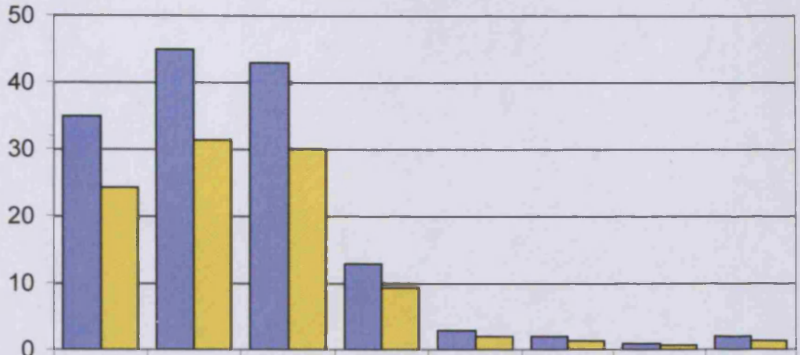
Secondly, the absence of obscure small village names, but inclusion of county names, especially in parts known to be rural (Cornwall, Devon, Gloucestershire, Shropshire, Sussex, etc), may suggest that where no large nearby conurbation exists, respondents co-operatively provide the next best-known entity, a county name. Thirdly, on a national level, the respondents from both Wales and Scotland use national labels as part of their self-identification. This may be national pride or the respondents accommodating the commonplace assumptions from the English side of the border that Wales and Scotland are homogenous entities. No English respondents used the label 'English'.

With regard to the division of the United Kingdom by regional television areas, there seems to be some strong correspondence between regional broadcasters' names and the labels invoked by respondents for self identification, especially where BBC broadcasting is concerned (e.g. Wales, South West, South East, and London). Is this because the long-established BBC chose labels that have been common region identifiers throughout recent history? Or have the labels chosen and used for over fifty years by the BBC come to influence regional identities to the extent that people identify with the region described by state broadcasting? It would be interesting to make a comparative study with other nations.

Finally, a breakdown of the figures for the regional origins of the speakers in Stage 1 is given in Figure 3.2.5.5 below. All respondents from the UK are listed.

Of relevance to this study in particular are the figures for the respondents from the catchment area of Cardiff, Wales, West Country, South East England and the Midlands (of England).

Figure 3.2.5.5 Stage 1 - Respondents by region



	South Wales	West Country	South East England	Midlands	North West England	North England	Scotland	Don't know
Respondents	35	45	43	13	3	2	1	2
Percentage	24.3	31.25	29.9	9.3	2.1	1.39	0.69	1.39

N=144

3.2.5.2 Stage 2

In Stage 1, the conceptual map was devised as a researcher tool for analysis of respondent region of origin data. For Stage 2, a perceptual map based on the Study 1 idealised conceptual map was presented to respondents, upon which respondents could indicate their origin, assisting them with their areal taxonomy of the UK (Figure 3.2.5.6). The rationale for this decision was to include in the research a bottom-up approach to accent labelling progressing in accordance with non-linguists' perceptions. This is seen as ecologically more valid in the context of the police interview of witnesses; for example, rapists do not tend to announce their regional origins whilst committing crimes, and there is no evidence to suggest that only those with strong socially marked accents are likely to commit crimes.

The map included the names of selected major cities, plus an invitation to mark their own town if it was not indicated. Alongside this, for respondents who preferred to indicate their origin through labels rather than maps, a list of the possible regional origins (Figure 3.5.5.7) was provided in order reduce the cognitive workload and likelihood of 'guesswork' taking place, again with the list of major towns within each region. In order not to focus respondents' attention

on the research focus of the southern half of the UK, all British regions were included, not just the southern half of the UK. The map was idealised with simplified straight-line borders, in order to be conceptually representative and salient to Stage 2's non-linguist respondents rather than geographically accurate.

In Stage 2, based on a second cohort of respondents (n=112) the adjacent southern UK catchment area for Cardiff University accounted for 98.6% of all respondents. This means virtually all respondents came from Wales, the South West, South East or Midlands.

Figure 3.2.5.6 Stage 2, idealised map of UK regions and cities



Figure 3.2.5.7 Stage 2 Geographical respondent/speaker locator

For example, if you're from Scotland and from Glasgow, write "me"

Scotland	me
Edinburgh	
Glasgow	me
Other (please say:)	

North East England	
Newcastle	
Middlesbrough	
Other (please say:)	

North England	
Leeds	
Sheffield	
Other (please say:)	

North West England	
Manchester	
Liverpool	
Other (please say:)	

North Wales	
Bangor	
Wrexham	
Other (please say:)	

Midlands	
Birmingham	
Nottingham	
Other (please say:)	

South Wales	
Cardiff	
Newport	
Valleys	
Swansea	
Other (please say:)	

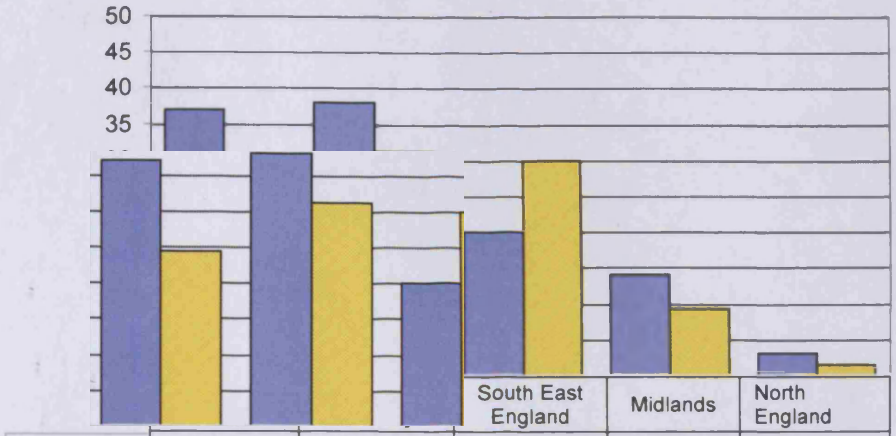
East Anglia	
Norwich	
Ipswich	
Other (please say:)	

South England	
Oxford	
Cambridge	
Other (please say:)	

South East England	
London	
Brighton	
Other (please say:)	

South West England	
Bristol	
Plymouth	
Other (please say:)	

Figure 3.2.5.8 Stage 2 - Respondents by region



			South East England	Midlands	North England
Respondents	37	38	20	14	3
Percentage	33.0	33.9	17.9	12.5	2.7

N=112

3.3 Materials

3.3.1 Suggested context of recordings: Stage 1

Given the forensic aims of the research, and the findings that accent and crime type can interact to provoke differing evaluations of guilt (Seggie, 1983; Dixon et al., 1994; Dixon et al., 2002; Dixon and Mahoney, 2004) all respondents were told that the voices to be heard had been recorded as part of a criminal investigation. Inspired by Seggie's choice of three different crime scenarios, (burglary/vandalism, assault and embezzlement) it was decided in Stage 1 to divide the respondents into two groups, each to be given a different criminal context for the crimes to have taken place in. Whilst both groups were told that the recordings were of suspects being interviewed in police stations after a crime had taken place, one cohort was told that the crime had been a serious sexual assault; the second half was told that the crime was a mugging. Both are violent crimes, but still markedly different in nature, given the sexual aspect of one of them. Respondents were told that we are interested in how they would describe a set of speakers in their own words and how they feel about them in a criminal context. In order to motivate respondents to answer seriously, respondents were told that their views would form a large part of some valuable research in their university department, which they could receive feedback on.

3.3.2 Suggested contexts of recordings: Stage 2

In Stage 2, a new crime was introduced for approximately 50% of the respondents, that of 'financial deception'. In order to act as a control against the forensic attitudinal conditions in Stage 1 and for half of the respondents in Stage 2, a second non-criminal context was provided for the remaining 50% of respondents in Stage 2 to act as a control against the previous examples based on crime and attributions of guilt. This cohort was informed that the evening may have ended in a successful 'romantic' liaison between one of the speakers and the girl identified in the voice samples.

3.3.3 Selection of voice samples: Stages 1 and 2

The voice donors were recruited through the researcher's personal and university networks and the same voices were used in Stage 1 and Stage 2. As one of the first studies of its kind, we have little guidance in the literature as to the ideal number of voice tokens to present in forensic speaker description research. Good practice used in forensic voice identification line-ups suggests a maximum of nine voices, as recommended by the McFarlane guidelines (Nolan 2003). However, the nine-voice recommendation pertains to the field of forensic speaker identification, where the respondent is passive and has only to recognise/identify the speaker. Forensic speaker description research is active by nature, and the potential for listener fatigue/cognitive overload must be acknowledged. It was decided therefore, to reduce the number of voice samples in order to reduce the cognitive load on the respondents. Five voice donors were selected.

Three notions were key in the selection of voices. Firstly, the speakers had to sound *spontaneous* and *unscripted* and not rehearsed or performing, avoiding issues of authenticity, community authenticity, and style authenticity discussed on p.76, and to reduce the chances of the respondents evaluating an emblematic stereotype rather than the speaker's voice. If respondents suspected at any point that the voice donors were simply that – people who had donated their voices for use in a social experiment – the validity of the research

would have been compromised. For this reason, actors were dispreferred. Secondly, it was important to ensure a level of *homogeneity*. If the speakers were seen to be markedly different people in terms of, for example, age, and more different than similar, again the validity of the research would have been compromised. Thirdly, it was essential for the voices to sound like genuine speakers of the dialects presented. As this would involve finding fluent speakers of five different dialects, the matched-guise technique was rejected and a verbal guise technique (Gallois and Callan 1981) (discussed on p.77) employed, whereby all samples were spoken by native speakers of each dialect. The use of actors was rejected (p. 91). Five authentic 'voice donors' were recruited, using the following criteria:

i/. As discussed on p.90, gender was identified as an area of research in itself, and to have both male and female sexes in the voice samples would put the research in danger of bloating beyond the limits for this study. It was decided therefore to stay with voice samples from one sex, male, as the social stereotype of possible male crimes may lend more credibility to the possibility that the voice donors were potential criminal.

ii/. Because of the suggested crime context, they should all be seen as roughly the same age and 'young.' In order to maintain credibility, the selected voices needed to plausibly be seen as 'young, carefree lads', the sort of men who respondents could easily believe were relatively free from commitments and able to go out with friends around pubs to socialise and get drunk and possibly engage in antisocial behaviour and mild promiscuity. The intention with this choice of social profile was to suggest in the minds of the listeners that the voice donors may be likely to put themselves in a position, through alcohol and or peer pressure, whereby their judgement might become impaired and they may commit a crime.

iii/. They spoke with different accents. The intention was to allow the respondents to identify the five voices as being similar, and the speakers as most probably part of the same homogenous social group. All would be saying very similar things (see 'script' below) but the biggest difference between them

would be their respective accents and voice features. The varied accents may or may not be apparent to the respondents at a conscious level, but plenty of research suggests that accent and voice features of a speaker could be largely responsible for the varied impressions that non-linguists may have of the same voice (Seggie 1983; van Bezooijen and Ytsma 1999; Dixon and Mahoney 2004). The intention was that the relationship between language varieties, social stereotypes and social judgements would be shown in relief against an otherwise socially homogenous background.

3.3.4 The selection of accents: Stages 1 and 2

Section 2.7 raised the criticism of previous language attitudes studies that have presented asymmetrically sized and conceptualised community and non-community accents to respondents, aligning towns, regions and countries alongside each other for evaluation. To accommodate this criticism, we can return to the sub-division of the Cardiff University catchment area highlighted in Section 3.2.5. This section noted how 93% of the respondents first surveyed in Stage 1 and 98% in Stage 2 were originally from the regions immediately adjacent to Wales – the Midlands, the South West – or from the South East area. From Cardiff, one can within two hours be in Birmingham, in the heart of the Midlands, in London in the south East, or in Bristol, the major urban centre of the South West. In identifying these four respondent regions, we have also identified four regions from which to draw the voice donors. It was felt that the voices to be played to the respondents should include accents commonly heard in Cardiff University's catchment area, especially as this may lend itself to greater ecological validity in terms of the voices that earwitnesses may encounter in their vicinity; such a design also affords greater experimental control in terms of relating evaluations of speakers from their own region with those of adjacent communities.

The first three voice donors recruited were selected on the basis of their regional accents, which in the judgement of the researcher were felt likely to be salient to the respondents. They were not, however, assumed to be absolute

tokens of the dialect of their origin, and this was specifically left to the respondents to evaluate:

Voice 1: 26 year old male: from Cardiff (*pop. 320,000*). He grew up in the city, attended a Welsh-medium state high school and universities in Cardiff and Bangor, North Wales. The speaker reported that he felt himself to have an identifiably Welsh accent, and in the judgement of the researcher, it was felt that there was a very high likelihood that he would be identified by non-linguists as being recognisably Welsh; it was anticipated that some respondents may even be able to locate him to the Cardiff area.

Voice 2: 27 year old male: from Bristol (*pop. 450,000*). He grew up in the city, attended a state high school and went to college in Bristol, where he has lived all his life. He reported that he felt himself to speak with an identifiably Bristolian/South West accent, and in the judgement of the researcher, there was a very high likelihood that he would be identified by non-linguists as speaking with a recognisably South West accent.

Voice 3: 30 year old male: from Hastings, (*pop. 85,000*). The oldest of the group, but in the judgement of the researcher, the one with the highest-pitched speaking voice. He attended a state high school in the coastal town of Hastings, Sussex, in the South East of England, some 60 miles south of London. He attended university in Brighton, in the same area. Outside of the South East region, he reported that he is frequently mistaken for a Londoner as the all-pervasive Estuary English has long since washed along the shores of the south coast and heavily influenced the speech varieties there. In the judgement of the researcher, it was felt that an identification of London/South East was a likely one amongst non-linguists.

These first three voices were selected from the same regions as 85% of the respondents in Stage 1 and in Stage 2. It was also felt, however, that it would be interesting to include an accent associated with more social than regional connotations.

Voice 4: 21 year old male: from Eastbourne (*pop. 93,000*), 15 miles from Hastings. The youngest of the group, this speaker was educated in a private school and speaks with an accent that he identified as 'public school' and in the judgement of the researcher an association with private schooling was likely. To a large extent, this accent is more a social than regional accent; one that has been transplanted to private schools around the world. Whilst it does have some social connotations of 'Englishness' it does not have any strong regional associations other than a general 'Southern English'. He was a final year university undergraduate student.

The fifth voice donor was selected for very particular reasons. Common to many studies is the dichotomous juxtapositioning of regional vs. 'high status' accent and a great many studies have investigated the contrast between the impression created by 'local' or 'posh'. It is this researcher's strong belief that there is an alternative speech variety that has gone largely over-looked, the category of 'no accent' described in Section 2.7 (p.73).

Voice 5: 28 year old male: from Birmingham (pop. 1,000,000) in the Midlands. The 'Brummie' accent (i.e. from Birmingham), has been featured in accent evaluation research since the early 1970s, and like other 'third-class' urban accents, it is an accent that is regularly parodied and stigmatised in the media and social discourse as being 'unattractive' or betokening a lack of education or sophistication. In fact, the Brummie accent is probably the most stigmatised accent in the UK (Dixon and Mahoney 2004; Bishop et al. 2005; Coupland and Bishop 2007). The speaker attended a local state high school in Birmingham and left home when he was 18 to attend university in the south of England and is educated to postgraduate level. He then moved on to become a practising lawyer, working in a profession that is criticised for having an unrepresentative dominance of Oxbridge people with middle class, RP accents (Darbyshire 2005).

Speaker 5's current accent has been influenced by social, professional and regional life choices that he has made. The speaker reports that due to his work as a lawyer, he has made a conscious effort to 'lose' the much stigmatised

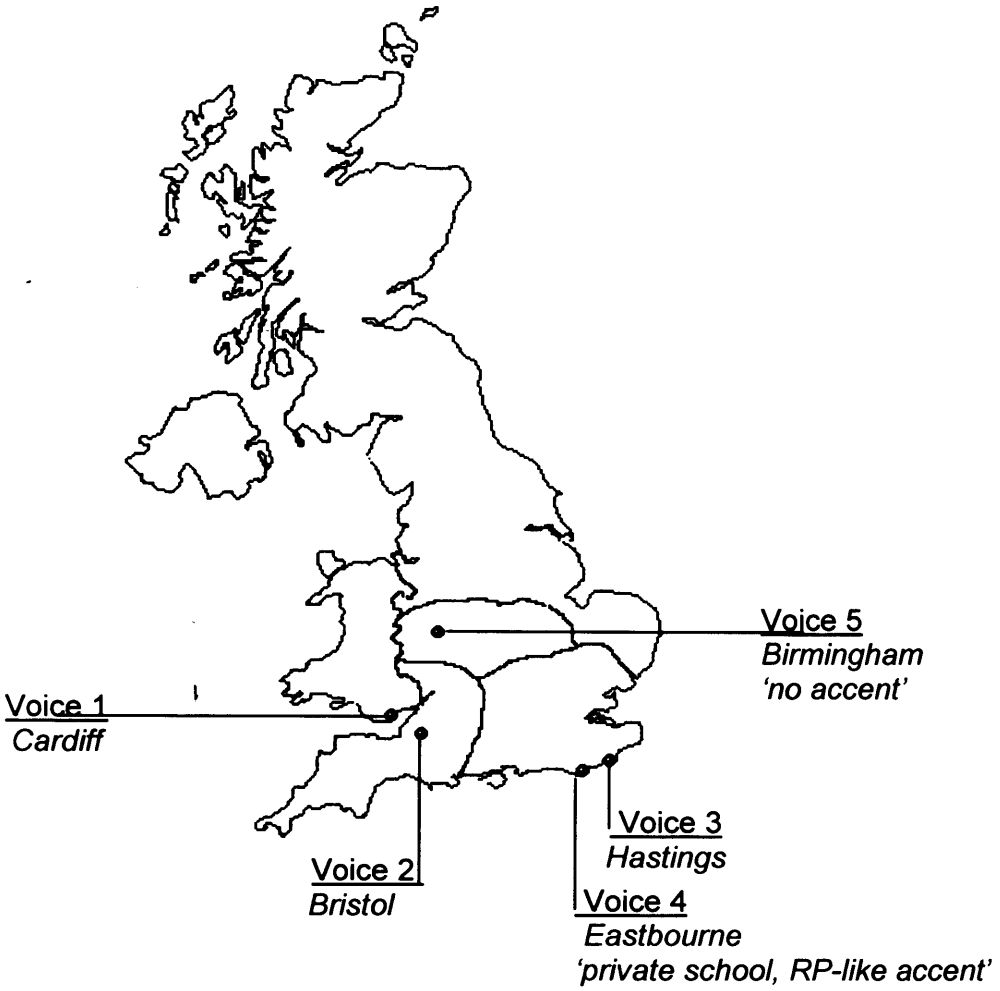
Brummie aspect of his pronunciation and he believes that to most, he simply sounds like he has 'no accent', although he is sufficiently self-aware to know that when he returns to his home town, a certain amount of bidialectal code-switching takes place. We may contrast Speaker 5's mobility and self consciousness regarding his accent with the speakers from Cardiff, Bristol and Hastings, all of whom still live in their home towns and report no issues with the way that they speak. The RP-like speaker from Eastbourne may not fit with this comparison, as, despite having left his home town to attend university in Cardiff, the accent that he uses appears to be more influenced by a sociolect associated with private schooling than by regional pronunciation. Speaker 5 is therefore the only speaker that is state educated, and has moved to another area for study and work. In the judgement of the speaker and the researcher, it was felt that very few respondents, and probably only those very familiar with the Birmingham accent, would identify his region of origin, and to most, hearing a recording made with a 'standard English' interlocutor, his interview would attract the description of 'no regional accent'. However, as no labels were prescribed by the researcher, this was left as a matter of respondent identification. Transcripts of all speakers are available in Appendix B. The geographical origins of the five voice donors used in the research are marked on the map in Figure 3.3.4.1 below.

3.3.5 Voice samples: Stages 1+2 - Picture Memory Interview

The same voice script was used for both Stages, the script for which was conceived firstly out of a requirement to suggest to the respondents that the recordings had taken place in a possible forensic context; secondly, the script needed to sound natural, unrehearsed and spontaneous. With these specific aims in mind, the Picture Memory Interview was conceived, and implemented in this research using the following steps:

i/. Format: within the scope of this research, it was decided that the most practical and efficient form of recording voices to present to an audience of non-linguist respondents was audio recordings, following the tradition of a great many direct and indirect approaches employed in language attitudes work.

Figure 3.3.4.1 Geographical provenance of voice donors



The recordings were made using the built in microphone on a Compaq N1015v laptop using a Realtek ALC202 soundcard, in a 3metre by 2metre office. With regards to the sampling quality of the recordings, although the technology used in police stations nowadays is of far higher quality than members of the public might expect (based on representations they encounter in television programmes and films), and although in the field of forensic identification, samples for a voice parade are selected and recorded at optimal quality by a professional forensic linguist for maximum clarity in playback for recognition purposes, the present research is neither intended as a display of the progress that has been made in the quality of police recordings, nor is it a study in forensic speaker identification.

The primary aim of this research is to create the illusion of a genuine forensic context - police interviews with possible suspects - and it was important to

maintain the illusion that the recordings had been made in the field, in police stations. It was felt that an optimal recording would risk shattering the illusion of a live police interview and arouse suspicion amongst the non-linguist audience that the recordings were not genuine. The decision was therefore taken to sample the recordings at 8kHz, rendering them almost telephonic in quality, well below what we might expect in other forensic contexts, yet clearly audible and more likely to match the public's expectations of the quality of recorded police interviews.

ii/. 'Criminal content of recordings': As discussed in Section 2.10.8, ethical constraints restrict the use of voice samples containing crimes in action as it would be potentially traumatic to play respondents recordings of crimes physically taking place, especially as, unlike television re-enactments, listeners would not have access to the usual visual cues that reassure viewers that the victim has subsequently survived. It remains to be seen what future developments there may be in this area, such as the DyVis project at Cambridge University, mentioned on p.96. On ethical grounds, therefore, it was decided that a context which suggested criminal activity without explicitly demonstrating it would be a more appropriate approach.

iii/. Implied suspicion: An alternative to playing crimes in action is to employ samples that contain interviews with people 'suspected' of having committed a crime. The decision in this study was to present five people's accounts of events *prior* to a crime taking place, but narrated after the crimes took places. This allows respondents the opportunity to form impressions of the speakers without them having to listen to explicit accounts of crimes, or to protests of innocence, both of which could invoke other personality judgements requiring explicit assessments of deception.

Respondents were told that the five voices they would hear belonged to five people suspected of being involved in a crime and to forestall any ego-achievement biases, in the form of zealousness to convict, respondents were told that one *or none* of the speakers may have been involved and they would not have to identify the guilty person as they may all be innocent. A potential

flaw in the design here would be that respondents could view the task as inherently competitive, in that if they selected one speaker to be guilty, they might, by default, decide all other speakers were innocent. However, the context of the provision of other scales in the questionnaire, and the encouragement from the administrator of the research from the outset for the respondents to mark their feelings in shades, for each question, including assessing each speaker's likelihood of being guilty, may have alleviated this problem. The responses collected indicate that the judges did indeed evaluate shades of possible guilt for each speaker, although this point will be returned to in the Discussion chapter and it is acknowledged here that future work may need to approach this question differently.

As part of the post-task de-brief, respondents were told at the end of the session that in fact none of the speakers had been involved in any crime, but only after the completed questionnaires had been handed in.

iv/. Factually neutral texts: It was clear that wildly differing stories could undermine the validity of the research, and so a factually neutral text was required, within which all speakers were required to be discussing the same series of events as if they had been co-witnesses. The material content would therefore be the same, whilst the voices telling the story would be different. This would allow the researcher to examine any variation in subjective evaluations amongst the respondents to the voices heard and to link them to the independent variable of mode of delivery, rather than to the material content of the samples.

vi/. 'Natural sounding': It was essential for the speech samples to sound unscripted and authentic, yet at the same time the content needed to be the same. As discussed above, criticism has been levelled at some previous language attitudes work, in that when samples are decontextualised, or samples use actors to mimic other accents the research may lack ecological validity, being isolated chunks of speech.

The PMI solution to this was to interview the voice donors with the researcher asking the questions, but giving the answers to the questions in pre-prepared picture form rather than in pre-written form. Reading aloud is, inevitably, likely to introduce several prosodic features and a greater frequency of spelling pronunciations (Garrett et al. 2003). An advantage of the PMI is that it allows the creation of not only factually neutral accounts across the five voice donors, but also affords a more natural delivery, as it is expressed using the voice donors' own words as natural responses to questions. This is also important in terms of ecological validity as respondents were being told that the speech samples were obtained from forensic *interviews* and the prosody and discourse markers should reflect this. In this respect, criticisms regarding ecological validity were accommodated in that the voice samples would be likely to be judged as meaningfully doing things with language rather than merely voicing utterances (Garrett et al. 2003). To further enhance authenticity, the voice donors were shown pictures (see Appendix A), which were then covered up when it came to answering the questions. The 'suspects' would then be recalling from short-term memory, with sufficient cognitive load to render a more genuine sounding recalled narrative. An alternative of using a video recording as the stimulus was considered, however, it was felt that this may adversely affect the required content neutrality across texts. The PMI is perhaps as close as we can get to an authentic-sounding police interview without access to the genuine material. (In the UK, police interviews are rarely placed in the public realm, although we await the outcomes of the aforementioned DyVis project.)

A story was created by the researcher around a fictional sequence of events that happened one evening. Seven pictures were provided with unambiguous content, depicting the events and material facts which would form the basis of the voice donors' accounts of the evening. Each voice donor was then asked a series of questions, to which they responded using their recall of the picture cues provided. Their responses were recorded. The question cues appear in Appendix A. The full transcripts of the recordings appear in the Appendix B. One of them is reproduced below:

Researcher/interviewer (R/I): Why were you celebrating?
Voice 1 (V1): Man U had just beaten Liverpool.. two-nil

R/I: Where did you go?
V1: Errr..we went to the Kings nightclub on Western Road....we got there around.. ten-ish

R/I: Who were you with and where were you?
V1: Ummm..I was with err.. a few of my mates.. and we were standing by the bar for most of the night.

R/I: Describe the girl. Who was she with and where was she standing?
V1: The girl..I think she was with err.. a few of her friends and she was standing .. by the dance floor

R/I: What was she wearing and if she had a drink, what was she drinking?
V1: She was wearing.a... short pink dress..umm.. I think she had on.. black knee-high boots and she was drinking.. a blue drink some. so was probably a cocktail.

R/I: How much had you drunk at this point?
V1: Ummm...probably around..4 or 5 pints of Stella

R/I: And what did you do next?
V1: We all.. went to the dance floor.. and we were stand er dancing around near the...near to the girl and her friends

Post recording, the interviews were edited to retain only the responses to the questions, and no responses were edited out. Given the lack of consensus across the fields of language attitudes and forensic speaker identification (see Section 2.10.9, p.98) on the length of recordings, or the number of sentences or fragments that the recording may comprise, it was felt that to loosely follow the McFarlane guidelines (Nolan 2003) with recordings of no more than one minute, and with no more than ten sentences used, would be appropriate. The five speakers' recordings were edited into samples of seven responses each. As a function of speaker memory and individual expression, the length of recording varied between speakers, from 28 and 52 seconds. The maximum

and minimum lengths are, in this first piece of research on forensic speaker description, informed by other fields but are nevertheless experimental. All recordings were screened to ensure that there were no material differences in content.

Having edited the recordings, it was discovered that there was slight GSM interference from a nearby mobile phone device on one of the recordings (Speaker 5). However, the interference was slight and was not deemed to impede upon the quality of the recording. The voices were then recorded on to cd and four separate versions were made, each with an alternative sequence, in order to reduce the likelihood of an ordering effect influencing the overall impressions described by the cohort. The recordings were played to the respondents in carpeted seminar rooms at Cardiff University, using substantial Sony portable cd players.

3.3.6 The questionnaires: Stage 1

The reader will note that this section is entitled 'questionnaires' rather than the singular. In line with standard methodological practice, the design of the study accommodated the possibility of there being an ordering effect, whereby the respondents' attention is drawn to certain features as a result of the sequence in which questions were encountered, affecting the impressions they have and the remarks they make. For Stage 1, then, a single set of questions was asked, but two alternative sequences were used. I shall first detail the content and then deal with the sequence of questioning. The issues that Stage 1 set out to address are as follows:

Firstly, each respondent was given time to supply information, anonymously, regarding their age, sex, first language and place of origin. They were then played each voice and given 30-45 seconds between each voice to address the following questions:

i/. "Which city or region do you think they are from?"
Following Preston's (1996) criticism of language attitudes research that has failed to ask the respondents where they believe the speakers to be from, an important part of this research was to elicit non-linguists' estimations, without giving respondents any labels or clues as to the speaker's origins. Not presenting the names of the regions where the speakers are from avoids the effects of respondent induction (p.81), whereby presenting socially marked labels may guide respondents into giving perceived preferred answers. In Stage 1, this was achieved by presenting the simple open-ended question regarding speaker origin above. The respondents had, therefore, been familiarised with the task by having already answered the questions with regards to themselves, and so had been cued to think about cities/regions of origin. Language attitudes results are open to debate when the respondents' areal taxonomy and identification of regional provenance of the voice samples are not known. Subsequent statistical analyses would investigate how the respondents as relevant community members defined these profiles (Garrett et al. 2003).

ii/. "Does this person come from your region/city?"
This gave respondents the opportunity to identify with or distance themselves from the voices that they heard, should the speaker be from an area that they recognise. This could be used as a useful cross-check of the community authenticity of the accents and the meaning of the evaluations.

For the first two questions, the respondents were able to simply circle 'Don't know' to prevent them feeling pressured into supplying information when they were unsure.

iii/. "How old is the speaker?"
Based on Braun's (1996) suggestion of supplying respondents with age spans, the questionnaire contained the following age options:

Under 18, 18-25, 26-35, 36-45, 46-55, 56-65 and over 65.

The decisions on the specific ages were researcher-derived, based on suspected categories of salient life-stage differences. It was felt that the results

would either justify or undermine this decision, signalling a lesson for future research.

iv/. "How would you describe each speaker's voice?"
This was an invitation for non-linguists to use their own non-researcher prescribed lexis to describe voice features.

v/. "How would you describe the speakers' personalities?"
Although there is a long history of research that indicates that non-linguists will infer a speaker's personality from their vocal characteristics or accent, this question was included in order to encourage respondents to differentiate the physical characteristics of the pronunciation/voice quality from the psychological impression made.

vi/. "How do you feel about each speaker?"
Long-established as an instrument for measuring subjective impressions, the semantic differential scale offers the respondents the opportunity to convey their impression where they may not be able to otherwise articulate them, giving them also the opportunity to indicate the strength of those impressions. Such scales are also easily processed statistically to efficiently reveal tendencies in the data. Moreover, the inclusion of bipolar semantic differential scales enabled the inclusion of both qualitative and quantitative research methods, avoiding the over-reliance on one research method alone (Garrett et al. 2003). The choice of dimensions to include was based on the relabelled hybrid of Zahn and Hopper (1985) and Perloff's (2003) rating scales suggested set in Chapter 2 (p.86)

The three selected factors were presented to the respondents on the following semantic differential scales:

1 = unintelligent and 7 = intelligent
1 = dishonest and 7 = honest
1 = aggressive and 7 = unaggressive

For the scale of aggression, the sense of 'lack of' aggression was reversed in order to counter any ordering or fatigue effects.

The choice of how many points to include on a scale is not one that is bound by any hard and fast rule, and it is established practice to see Likert-type scales with 5-level, 7-level or 10-level points. Dawes (2008) found that data from all three of these items showed very similar characteristics in terms of mean, variance, skewness and kurtosis after a simple transformation was applied. In this context, the choice of 7-level points on each scale was felt likely to provide respondents with enough atmosphere to express their shades of feelings.

vii/. 'Based on the voices that they have heard, which speaker(s) is/are most probably guilty of the crime of (delivered by questionnaire administrator, with half of the respondents being told the crime was serious sexual assault, the other half being told the crime was mugging) and which speaker(s) is/are probably not guilty?' Again, a semantic differential scale was provided upon which respondents could plot their judgements, ranging from:

1 = probably not guilty and 7 = most probably guilty

viii/. "Please give a reason for your choice"

Following the influence of the accountability motive discussed in Chapter 2 (p.79), it was felt that providing the opportunity for respondents to supplement their answers, elaborating on their judgements with short open-ended qualitative items has the potential to subdue some of the possible response biases discussed in Chapter 2, whilst assisting the interpretations of the quantitative results (Garrett et al. 2003).

3.3.6.1 The question sequence: Stage 1

As mentioned above, the research took into account the possible effect that question sequence may have on the responses, particularly the presentation of questions on accent/voice features first or a possible criminal context first. Would the suggestion of a criminal context from the outset condition the manner in which the respondents were listening? Would there be a notable difference between the judgements made by those who are introduced to a possible forensic context first compared with those informed regarding the forensic aims

of the research at the end, having already described their impressions of the voices?

To this end, two alternative questionnaire sequences were used. (The crimes of sexual assault and mugging appeared in questionnaires for Stage 1-1 and Stage 1-2.)

Stage 1(1)	Stage 1(2)
i/. Speaker origin?	i/. Speaker guilt
ii/. Same city as you?	ii/. Why?
iii/. Speaker age?	iii/. Speaker age?
iv/. Describe the speaker's voice	iv/. Describe the speaker's voice
v/. Describe speaker's personality	v/. Describe speaker's personality
vi/. Semantic differential scales for intelligence, honesty aggression	vi/. Semantic differential scales for intelligence, honesty aggression
vii/. Speaker guilt?	vii/. Speaker origin?
viii/. Why?	viii/. Same city as you?

3.3.7 The questionnaires: Stage 2

The format of the questionnaires for Stage 2 was broadly similar to that of Stage 1 in that it asked respondents to state where they were from, where they thought the speakers were from and to provide evaluations of the speakers' voices. However, it differed in the following key areas:

3.3.7.1 Perceptual maps

Following the results of Stage 1, a map was generated to assist respondents with their areal taxonomy of the UK (Figure 3.2.5.6) in Stage 2, based on media divisions, non-linguist descriptions and linguistic research. On this map, respondents were first invited to plot their own geographical provenance as a 'warmer' exercise, encouraging cognitive elaboration. Having oriented themselves to the task on a personal level, they could then plot their estimations of origin of the five speakers. To this end, the idealised perceptual map and list of origins (Figure 3.4.1.2) were presented afresh.

The introduction of the perceptual map has precedents in the literature and was discussed in 2.11.1 (p.100). In the current research context, it was anticipated that the use of perceptual maps would allow the researcher to view the respondents' areal taxonomy and assess whether such maps assisted non-linguists in their estimations in the forensic context. Moreover, it was felt that the maps might act as a support to respondents' open-ended estimations of speaker origin and support their verbal descriptions. It is also recognised that maps may present an opportunity for linguists to indicate their social stereotypes rather than voices, but it was decided that if such methods allow greater insight into the architecture of non-linguist perception of language variation, or allow us to see in the Prestonian (1996) sense, what is available or suggestible to them, and how global or detailed or accurate inaccurate their estimations might be, this can only be of benefit to the field.

3.3.7.2 Stage 2, Crime context

Respondents in Stage 2 were invited to mark their intuitions regarding possible guilt using a seven-point bipolar scale, indicating at each end 'probably guilty' and 'probably not guilty' (a slight change of wording from Stage 1).

In order to act as a control against the forensic attitudinal conditions in Stage 1 and for half of the respondents in Stage 2, a second non-criminal context was provided for the remaining 50% of respondents in Stage 2. This cohort was informed that the evening the recordings refer to may have ended in a successful 'romantic' liaison between one of the speakers and the girl identified in the voice samples. Again, plotted on a seven-point bipolar scale, the respondents were invited to indicate their intuitions regarding the potential success of each voice in their 'romantic' quest.

3.3.7.3 Rating the speakers' accents

Following the folk linguistic tradition of presenting linguistic concepts in lay terms, and in order to move towards greater ecological validity in the wider world of police interviewing of witnesses (personal correspondence with police

officers in South Wales Police), it was important that linguist-orientated concepts, such as regional and social accent or the acceptance that it is not possible to speak a language without an accent, were to be avoided. In preference, terms and concepts common to the non-linguist were employed. Respondents were therefore asked on hearing the voices to indicate on a bipolar seven-point scale presented for each of the voices, their feelings regarding the 'strength' of accent. This is inspired by van Bezooijen and van Hout's (1985) 'degree of accentedness scale', discussed in Chapter 2, but also combines three common non-linguist beliefs regarding English: firstly that some speakers have 'no accent', referring to a lack of obvious geographical markers; secondly, that some people are 'well-spoken' (RP), again with a lack of geographical markers, but with social markers of high status; and finally, that some people speak with regional accents that contain strong geographical markers. This scale presented to the respondents is essentially two scales put together and included such terms as, 'strong regional accent,' 'no accent' and 'well-spoken.' These terms were represented for each voice as below:

					no accent?					
Speaker	Strong regional accent |	1	¦ 2	¦ 3	| 4	| 5	¦ 6	¦ 7	| Well-spoken
	Anything distinctive about the accent?									

Again, in order to maintain ecological validity in terms of the police interviewing of witnesses, an additional question regarding the distinctiveness of the accent was included, in response to which non-linguists were able to report qualitatively, and in lay terms, any features which were particularly prominent for them. Respondents were also invited to mark on a fresh copy of the Study 2 map and in the geographical tick boxes, where they felt each speaker was from.

3.3.7.4 Describing and rating the speakers' voices

In Stage 1, respondents had been invited to provide qualitative responses regarding the speakers' voices. Following methodologies employed by van Bezooijen (2002) and Preston (1999) discussed in Chapter 2 (pp. 52 and 51), these data and the most common descriptive antonyms were plotted onto

bipolar semantic differential scales, which were then presented to the respondents of Stage 2. However, the bipolar scales differed from those of van Bezooijen and Preston, in that they focused exclusively on objective measures, rather than included subjective measures, as had been used in the previously studies.

In providing their evaluations, respondents were informed that they did not need to provide a comment on all aspects, but only on those which they felt to be relevant. These scales were:

Slow	-	Fast
Hesitant	-	Confident
Deep voice	-	High-pitched voice
Monotone	-	Melodious
Soft	-	Loud
Shaky	-	Steady

Some of the qualitative data collected from Stage 1 in which the respondents provided descriptions of the speakers' voice features were non-polar in quality, in the sense that there are no available or suitable antonyms. These non-polar qualitative comments were provided as separate tick boxes:

Anything distinctive about the voice?	❑ dry	❑ whispery	❑ sharp	❑ shrill	❑ effeminate
	❑ rough	❑ croaky	❑ raspy	❑ anything else?	

In order to accommodate issues of ecological validity, the question asked by the DC from South Wales Police (Chapter 1, p. 8) asking the witnesses to describe whether there were any speech defects was included. Those speech defects mentioned by respondents in the qualitative data provided in Stage 1 were included:

Any speech defects?	❑ lisp	❑ stammer	❑ stutter	❑ anything else?

It should be noted that 'stutter' and 'stammer' are most likely to be essentially synonymous for non-linguists, but as the questionnaire was derived from non-

linguist terms, the two were included in Stage 2. They will be treated as synonymous in the results section.

Finally, respondents were invited to comment on whether they felt the voices to be in any way familiar:

Does the speaker sound like anyone famous/anyone that you know?

3.3.7.5 Question order

The following question order was used for the Stage 2 questionnaires:

STAGE 2-1 (Crime) n=45	STAGE 2-2 ('Got the girl') n=62
i/. Where are you from? (Map+list)	i/. Where are you from? (Map+list)
ii/. The speakers' accents	ii/. The speakers' accents
iii/. The speakers' provenance	iii/. The speakers' provenance
iv/. The speakers' voices	iv/. The speakers' voices
v/. Speaker guilt	v/. Did the speaker 'get the girl'?

3.4 Procedure: Stages 1 and 2

All the participants were divided into six groups of between 15-25 respondents for the administration of the questionnaire. This was sufficient for a satisfactory range of voice sequences to be played to avoid any ordering effects. Each participant was told either by the researcher or by one of the three postgraduate student volunteers that assisted in the project, that they would be listening to some voices and that we were interested in their own personal reactions to the speakers. Respondents were reassured that there were no right or wrong answers, so there would no benefit in discussing with or copying from their friends; discussion between respondents was firmly discouraged as respondents may have influenced each other's views, heightening the effects of social conformity – not wanting to be the 'odd man out' (McGuire 1985), with such contamination having an impact on the validity and reliability of the data gathered.

The research took place in university seminar rooms where respondents were asked to sit apart from each other. Respondents were asked not to change their answers once completed, or to look forward through the questionnaire until requested to turn the page. No discussion regarding the voices heard or the attitudes that should be expressed was entered into by the researcher, or volunteer administrators.

The respondents heard the voices twice, once at the beginning and once in the middle of the questionnaire completion. It was felt that two listens would be proportionate to the size of the task and that a third listen may afford the listeners the opportunity to become over-attentive towards looking for signs of deception where the issue of guilt was concerned. Subsequent statistical checks were conducted to ascertain whether the sequence of presentation of voice and questions had any effect and no significant effect was found. Having completed the questionnaires and handed them back, all respondents were informed as to the true nature of the research in an informal de-brief. They were informed that none of the voice donors had been genuine suspects, but that we were interested in the impressions that people's voices could make on them, especially in a forensic context. They were also given the opportunity to submit an email address at which they could be contacted with the final conclusions of the project. The process of preparing and conducting the research lasted approximately forty minutes, after which feedback was given and a discussion based on the respondents' questions was held.

A first pilot group was used to check for any unforeseen problems consisting of an informal group of postgraduate student volunteers. The only issues that arose were around the positioning of the audio equipment (it had a short lead), which necessitated a slight re-arrangement of the furniture to facilitate a better acoustic transmission to all respondents.

4 - RESULTS: Stage 1

4.1 Introduction

This chapter reports the results of the first of the two studies conducted in this research, the details of which were outlined in Chapter 2. Based on the research questions Section 2.14.2, each section in this chapter represents an individual question given to the144 respondents. Where the questionnaire given to the respondents had items ordered according to methodological aims, the results presented here have been re-ordered for the benefit of the reader.

The chapter begins with what we might describe as 'what non-linguists can tell us about the speakers'; Sections 4.2–4.4 deal with the respondents' estimations of the speakers' origins: where do the respondents think the speakers come from (4.2)? Are there any associations between the respondents' own region and their identifications of speaker origin (4.3)? Which speakers were easiest/most difficult for respondents from each region to spot (4.4)? Continuing the questions regarding what respondents can tell us about the speakers, 4.5 moves on to asking the respondents to estimate the speakers' ages.

Sections 4.6–4.10 deal with what we might call less concrete, 'impressionistic' feelings focusing the respondents on aspects of the speakers' voices, vocal characteristics and how the respondents feel about the speakers: 4.6 reports results the respondents' qualitative descriptions of the speakers' voices; 4.7 reports results of respondents rating each speaker's intelligence; 4.8 reports on respondents' ratings for each speaker's honesty; 4.9 reports on respondents' ratings of each speaker's aggressiveness. Section 4.10 reports on respondents' attributions of speaker guilt. Finally, Section 4.11 takes all the quantitative data from 4.6–4.10 and performs principal component analyses and tests of correlation. The chapter concludes with a brief summary (4.12).

4.2 Where did respondents believe speakers are from?

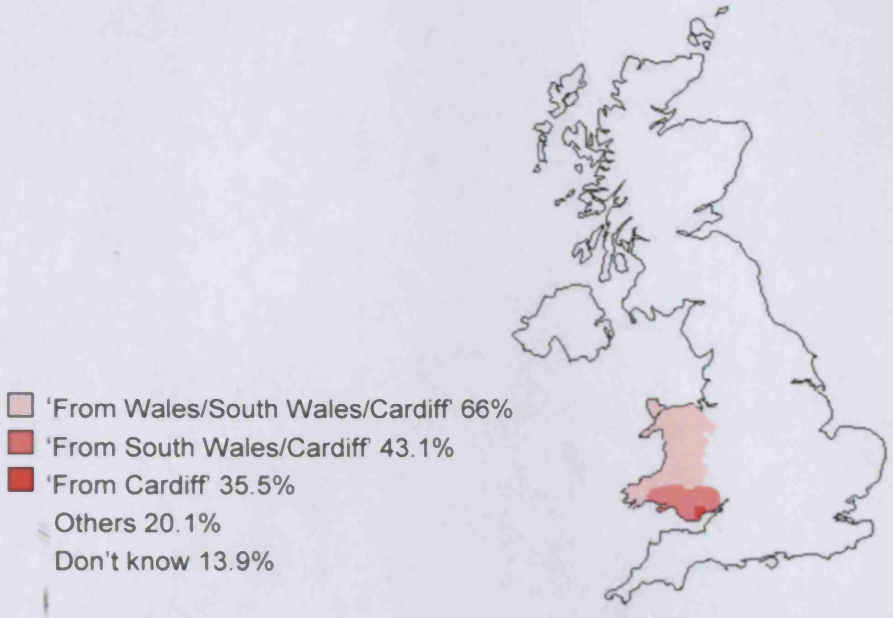
The following data show Stage 1 respondents' estimations of the five speakers' provenance. As described in Chapter 3, the respondents had first been asked to identify their own provenance by providing the name of their town/city/county or region, in order to familiarise themselves with the task and encourage cognitive elaboration. They then heard the five speakers' voices and were invited to state where they believed the speakers could come from in the space provided on the questionnaire. No maps or other clues were provided in Study 1.

The results presented in this section are given in percentage terms, detailing the *positive* identifications of the speakers' region of origin by town or city as given, but also the region where that the town or city is located, based on the regional divisions of the Cardiff University catchment area described in Section 3.2.5.1, p.135. For example, any estimations for Speaker 1 (from Wales) being from 'Newport' in Wales or 'Swansea' in Wales or 'Cardiff' in Wales are coded as 'Wales'. According to this coding, those estimations falling outside of the region of 'Wales' were simply noted as 'others'. For example, an estimation of Speaker 1 (Wales) being from 'London' would be coded as 'other'. Where respondents have left a blank space on the questionnaire, this is recorded as 'don't know'.

In order to assist interpretation of these data, they are visually represented in shaded form according to each speaker's home region, one map for each speaker. The intensity of colour/darker shade represents the greater specifity with which the respondents identified the speaker. The colouring does not represent statistically significant differences, but simply offers the reader a sense of the direction and gradation of responses.

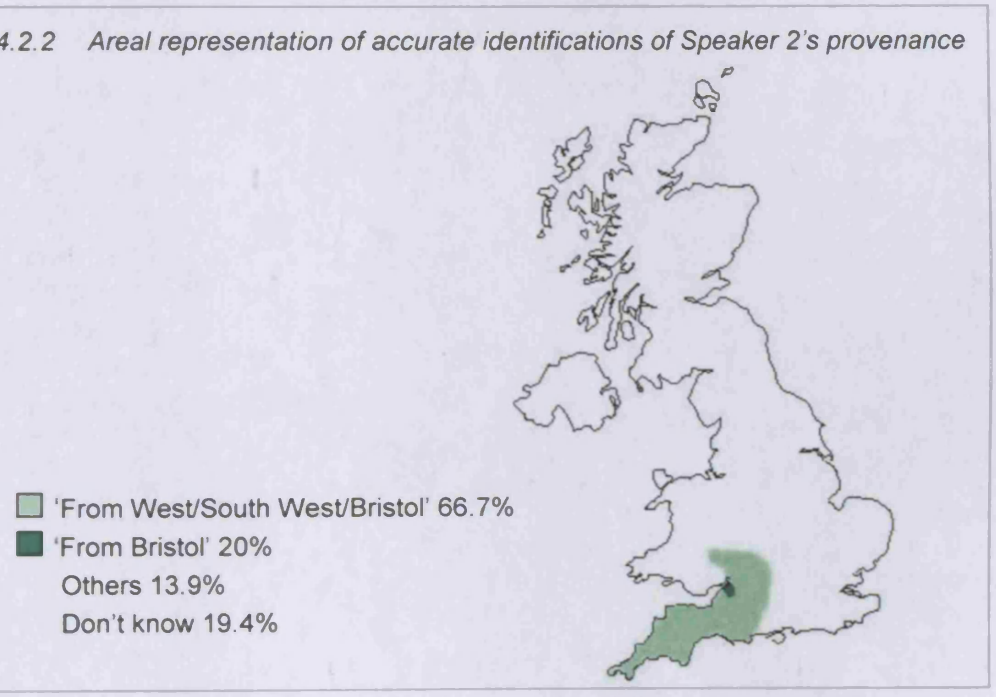
Speaker 1: From Wales > South Wales > Cardiff

Figure 4.2.1 Areal representation of accurate identifications of Speaker 1's provenance



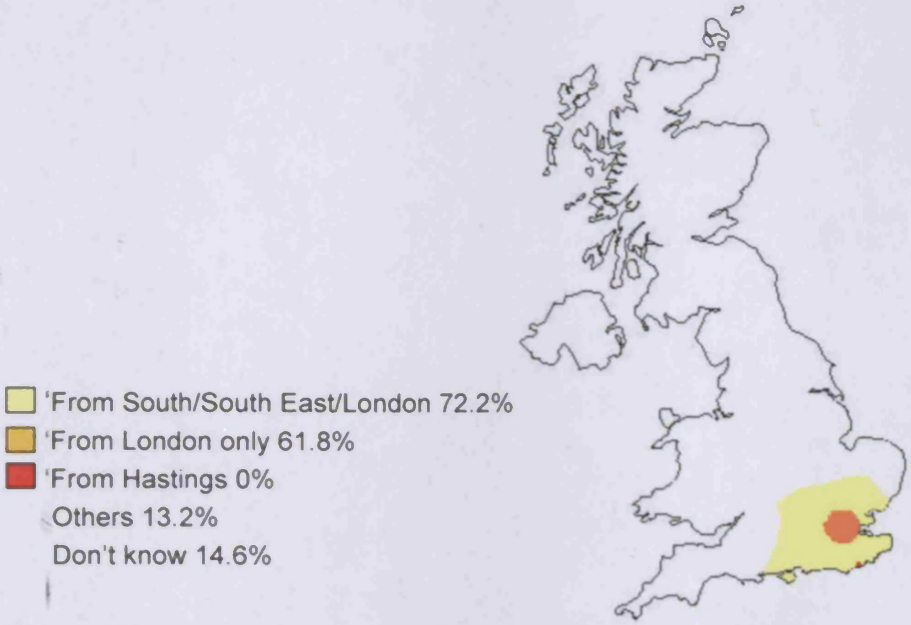
Speaker 2: From South West > Bristol

Figure 4.2.2 Areal representation of accurate identifications of Speaker 2's provenance



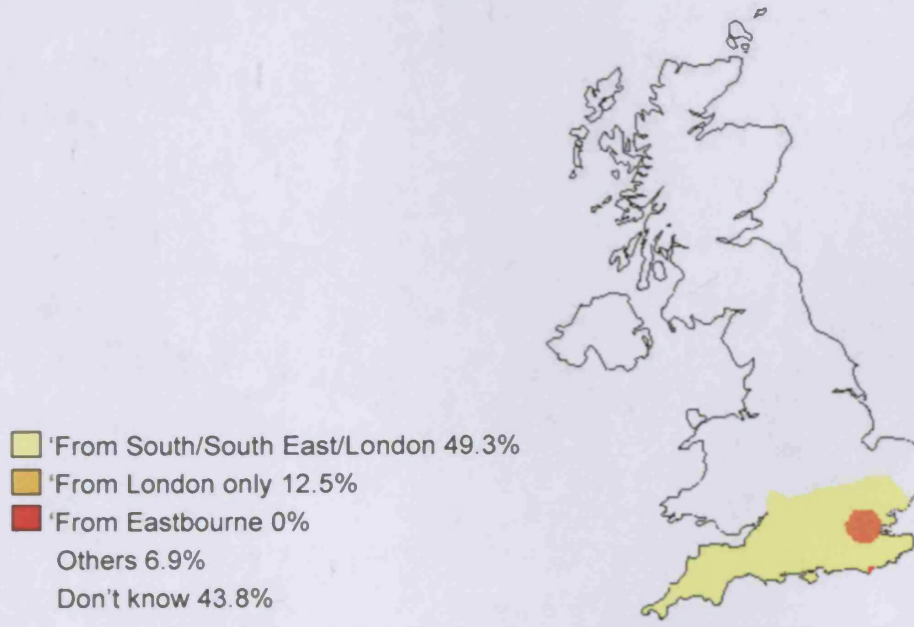
peaker 3: From the South East > Hastings

Figure 4.2.3 Areal representation of accurate identifications of Speaker 3's provenance



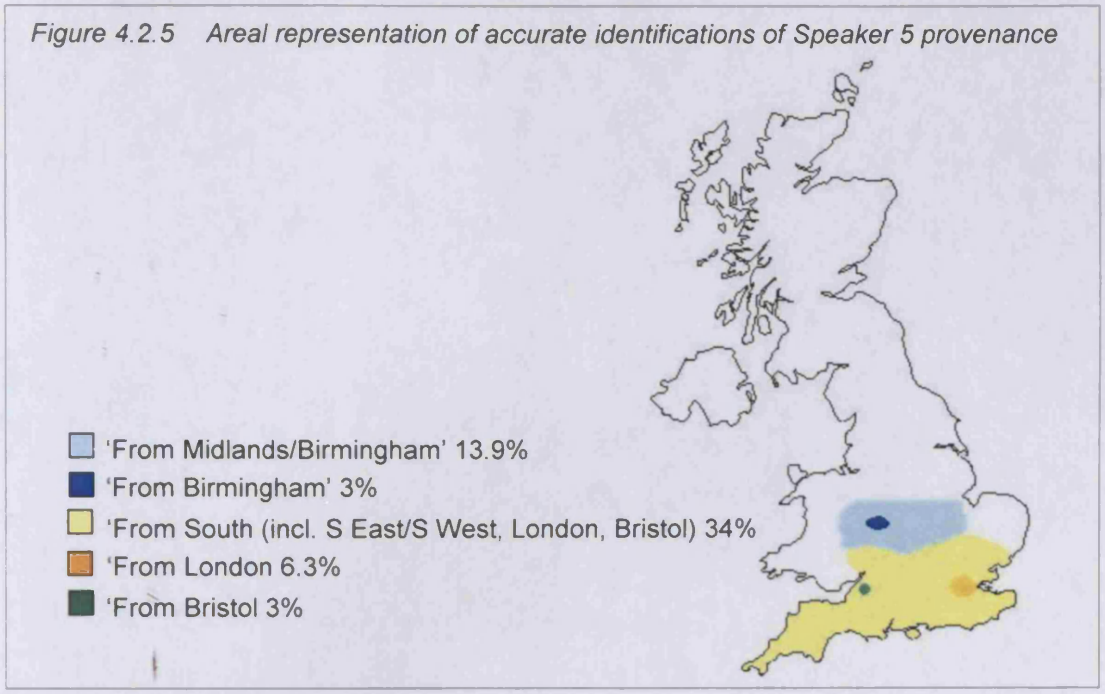
Speaker 4: From the South East > Eastbourne (RP-like)

Figure 4.2.4 Areal representation of 'accurate' identifications of Speaker 4 provenance



Speaker 5: From the Midlands > Birmingham

Figure 4.2.5 Areal representation of accurate identifications of Speaker 5 provenance



The data in the figures above show that Speakers 1, 2 and 3 all received high frequency of regionally accurate estimations of provenance, with Speaker 3's region being most frequently identified (72.2%), followed by Speaker 2=South West (67%) and Speaker 1=Wales (66%). Within these regional estimations, many respondents also gave an urban estimation of provenance, Speaker 3=London (61.8%), Speaker 1=Cardiff (35.5%) and Speaker 2=Bristol (20%). However, the data may point to a difference in the salience of Speaker 3's accent: whilst the urban-level estimations for Speaker 1 and 2 are indeed correct, the urban estimations of Speaker 3 coming from London is incorrect, as the speaker is from the south coast of England, which received no correct estimations. The possible reasons for this urban-level inaccuracy will be considered in the Discussion (Chapter 6).

The data for Speaker 4 are strikingly different from the previous three speakers. Firstly, in relation to Speakers 1, 2 and 3, the number of 'correct' estimations for region of origin drops to 49.3%, a figure that is based on respondents believing the speaker comes from somewhere in the 'South of England', a broad term, potentially encompassing both the South East and the South West regions identified for this study. (Of course, in the real world forensic context, the

questioner could ask the respondent earwitness for further clarification or detail, with the probable result that the geographical spread presented would be smaller.) The number of 'don't know' responses is markedly higher for Speaker 4 than for Speakers 1, 2 and 3, at 43.8%; the number of 'misidentified' responses is smaller (6.9%). The reasons and implications for these estimations of the RP speaker's provenance will be considered in Chapter 6.

Speaker 5 presents a somewhat different set of data relative to the previous four speakers. The most striking points are that the number of accurate regional identifications is dramatically lower (13.9%) and the number of misidentifications much higher (41.7%) than for any of the other speakers. There is also a high number of 'don't know' responses (44.4%). More specifically, hearing Speaker 5's voice leads nearly half of the respondents to misidentify where Speaker 5 is from, while 14% of respondents correctly identify him as being from the Midlands. Finally, we must not neglect the fact that more respondents than for any other speaker suggested that Speaker 5 was from the Midlands (14%) and were 'correct' in doing so. It would be hard to imagine, given the data for the other speakers, that this could occur without there being some genuine aspect of Midlands speech present. We cannot be certain as we have no recordings of the speaker from his pre-18 years living in Birmingham, however, there is strong reason to suspect, given the background information detailed above, that what the respondents are noticing are the remnants of the Birmingham accent that was once a strong feature of the speaker's pronunciation, now diluted and replaced due to professional and social mobility.

4.3 Estimations of speaker origin: are they associated with respondent origin?

In the previous section, all five speakers' voices were examined according to the 144 respondents' estimations of speaker origin. The differences in the patterns of responses suggest that the respondents generally saw the five voices as from geographically or socially distinct backgrounds – Wales, South West, South East (London), South East (RP) and broadly Southern English. In

this section, the data are interrogated to explore the relationship between respondents' estimations of speaker origin and the respondents' own region of origin, and whether respondent origin is associated with the identification of the five voices presented.

To this end, respondents were coded according to the four catchment area regions, Wales, the South West, the South East or the Midlands. As this question is related only to respondents from the four regions within the catchment area, the few respondents from the North of the UK and respondents who did not provide any information on their regional origin were excluded, leaving 136 respondents. It should be remembered that there was a relatively low number of respondents from the Midlands (13 - 9.6% of respondents). Although they are included, data from Midlands' respondents are not treated as statistically robust and these figures should be treated with caution compared to the data from the better represented regions.

Respondents' positive or negative estimations of speaker origin are grouped by region and presented in tables of frequency. In order to interrogate this categorical data, a chi-square test was performed for each speaker and the respondents' origins, investigating any significant associations. The regional estimations tested were:

Speaker 1 - estimated to be from 'Wales'
Speaker 2 - estimated to be from 'South West'
Speaker 3 - estimated to be from 'South East'
Speaker 3 - estimated to be from 'London'
Speaker 4 - estimated to be from 'South/South East/London'
Speaker 4 - estimated to be from 'South East/London'
Speaker 5 - estimated to be from 'Midlands'
Speaker 5 - estimated to be from 'South'

The tables of frequencies containing percentage figures for positive and negative identifications are presented below. In each case, the region that provided the highest percentage frequency of positive identification is listed first,

the descending order of regions indicating decreased percentage frequency. The results of the chi-square tests are presented below each table. The frequencies given in the tables should be taken as illustrative of possible differences in positive identifications according to respondent region, dependent on the chi-square tests demonstrating any significant associations. A commentary on the data, the associations and frequencies follows the tables.

Table 4.3.1 Speaker 1, identified as from 'Wales' by respondents from:

	Respondents guessed 'Wales'		Total	% Correct
Respondent Region	*no*	*yes*		
Wales	2	33	35	94.3
West/South West	18	27	45	60.0
South East	19	24	43	55.8
Midlands	7	6	13	46.2
Total	46	90	136	

Respondent/speaker origin1 - $\chi^2 = 17.513$, df = 3, $\rho = 0.001$

Table 4.3.2 Speaker 2, identified as 'South West' by respondents from:

	Respondents guessed 'South West'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	12	31	43	72.1
West/South West	13	32	45	71.1
Midlands	5	8	13	61.5
Wales	16	19	35	54.3
Total	46	90	136	

Respondent/speaker origin - $\chi^2 = 3.497$, df = 3, $\rho = 0.321$

Table 4.3.3 Speaker 3, identified as 'South East' by respondents from:

	Respondents guessed 'South East'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	7	36	43	83.7
West/South West	12	33	45	73.3
Wales	12	23	35	65.7
Midlands	7	6	13	46.2
Total	38	98	136	

Respondent/speaker origin - $\chi^2 = 7.973$, df = 3, $\rho = 0.046$

Table 4.3.4 Speaker 3, identified as 'South East/London' by respondents from:

	Respondents guessed 'South East/London'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	14	29	43	67.4
West/South West	16	29	45	64.4
Wales	13	22	35	62.9
Midlands	8	5	13	38.5
Total	51	85	136	

Respondent/speaker origin - $\chi^2 = 3.727$, df = 3, $\rho = 0.292$

Table 4.3.5 Speaker 4, identified as 'South/South East/London' by respondents from:

	Respondents guessed 'South/South East/ London'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	12	31	43	72.1
West/South West	23	22	45	48.9
Midlands	7	6	13	46.2
Wales	26	9	35	25.7
Total	68	68	136	

Respondent/speaker origin - $\chi^2 = 16.751$, df = 3, $\rho = 0.001$

Table 4.3.6 Speaker 4, identified as 'South East/London' by respondents from:

	Respondents guessed 'South East/London'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	26	17	43	39.5
West/South West	34	11	45	24.4
Midlands	10	3	13	23.1
Wales	30	5	35	14.3
Total	100	36	136	

Respondent/speaker origin - $\chi^2 = 6.612$, df = 3, $\rho = 0.085$

Table 4.3.7 Speaker 5, identified as 'Midlands' by respondents from:

	Respondents guessed 'Midlands'		Total	% Correct
Respondent Region	*no*	*yes*		
Midlands	9	4	13	30.8
South East	36	7	43	16.3
West/South West	39	6	45	13.3
Wales	33	2	35	5.7
Total	117	19	136	

Respondent/speaker origin - $\chi^2 = 5.243$, df = 3, $\rho = 0.154$

Table 4.3.8 Speaker 5, identified as 'South' by respondents from:

	Respondents guessed 'South'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	26	17	43	39.5
West/South West	36	9	45	20.0
Wales	29	6	35	17.1
Midlands	11	2	13	15.4
Total	102	34	136	

Respondent/speaker origin - χ^2 =7.238, df = 3, ρ = 0. 064

The first point to observe in the results of the chi-square tests is that there are significant associations between the estimated speaker origins and the origins of the respondents in some but not all cases. Significant associations can be observed between respondent origin and estimations that:

Speaker 1 comes from Wales (ρ<0.001)

Speaker 3 comes from the South East (ρ<0.046)

Speaker 4 comes from the South/South East/London (ρ<0.001)

The frequency tables, whilst purely illustrative, do give strong indications as to how the associations relate. For Speaker 1 (Table 4.3.1), as many as 94% of Welsh respondents identify the speaker as Welsh, compared with 60% or below for respondents from English other regions. Such a dramatic drop in frequency is perhaps unsurprising, given the history of social, political, economic and linguistic differences that have existed between England and Wales. Conversely, Speaker 3, from the South East, is most frequently estimated to be from the South East (including London) by respondents from the South East. What is of note here for both recognition of the Welsh speaker and the South East speaker is the asymmetry in perceptual salience: where you are from is related to how salient certain non-community accents are.

An association between respondent and speaker origin is also seen for Speaker 4, a voice with an accent closely associated with RP. Speaker 4 is most frequently identified as from the South/South East/London by respondents from the South East (Table 4.3.5). Respondents from the South West, in Inoue's (1999) terms, part of the Southern dialect region, are also more inclined to

identify Speaker 4 as speaking with a southern variety of English. It is interesting to note that it is the Welsh respondents who return much lower estimation frequencies for an RP-South/South East/London origin than respondents from English regions. This may indicate less familiarity with the hierarchical nature of language varieties in England.

In contrast to the data above, other chi-square tests reveal no significant associations between respondent and speaker origin. These are:

Speaker 2 comes from the South West
Speaker 3 comes from London
Speaker 4 comes from the South East/London
Speaker 5 comes from the Midlands Speaker 5 comes from the South

Table 4.3.2 for the Speaker 2 (South West) estimations shows no significant associations, suggesting that the accent is symmetrically salient. The data for Speaker 3 (Table 4.3.4), show an interesting contrast. There are no significant associations between respondent origin and estimations of Speaker 3 being from London. However, as discussed above, there is a significant association between respondent origins and estimations of the speaker being from the South East (Table 4.3.3). The identification of the speaker with London is one that appears almost symmetrical across all regions, whereas the association with a wider South East is one made most frequently by respondents from the South East. This may reflect a wider understanding of language diversity in the South East region by those who live there.

Likewise, there is a contrast in the data for Speaker 4, the RP speaker. Where there are no significant associations between respondent origin and Speaker 4 being identified with South East/London (Table 4.3.6), there is a significant association between respondents' regional origin and estimations of him being from the South/South East/London (Table 4.3.5). It would appear from the frequencies in Table 4.3.5 that despite assertions in public discourse that RP is 'correct' English pronunciation and not particularly related any one specific region, those non-linguists from the South East of England are highly likely to

identify the speaker as coming from a 'South region (of England) or one of its constituent towns. RP appears to be either more geographically salient or simply more familiar to speakers from the South East. There also appears also to be evidence of a distancing effect with only a quarter of the Welsh respondents being able to regionally locate RP.

Finally, there are no significant associations between speaker and respondent origin with regard to estimations of Speaker 5's origin, be they estimations of Midlands identity or South identity (Tables 4.3.7 and 4.3.8). As seen in the previous section, Speaker 5 served as an alternative to the common approach to language attitudes research, in that there were few positive identifications of his provenance – what many non-linguists might call 'accentless', and yet, few respondents indicated that he spoke with an RP accent or that his origin was 'Oxbridge' or any of the other RP-like labels mentioned for Speaker 4.

4.4 Which accents were most salient to each region?

In the evolution of language attitudes research, there has been an enthusiasm to establish how various regional and social accents are evaluated by the public at large. Can we always assume that within one political nation the various constituent accents are familiar and available (Preston 1996) to all speakers? Is it the case that regional or social distance from another speech variety influences respondent's ability to recognise and label that accent? Or are all respondents equally familiar with the range of speech varieties that exist in one country? There is a paucity of research in the area of mutual recognition of dialects/accents in the UK, and the following section seeks to address that by interrogating the data for examples of frequencies for recognition of speech varieties according to the origin of the respondents. The following questions are asked of the data:

Which of the voices heard were most salient to the respondents according to their own regional provenance? Grouped by their own region of origin, which

voices could respondents best recognise and then correctly identify, utilising relevant city/county/region-related labels?

Table 4.4.1 Salience and availability rates for Welsh respondents

	94.3%	Speaker 1 (from Cardiff, Wales) = Wales
	65.8%	Speaker 3 (from Hastings, South East) = London
	57.2%	Speaker 2 (from Bristol, South West) = West Country
	14.3%	Speaker 4 (RP) = South/South East
	11.4%	Speaker 4 (RP) = London
	5.8%	Speaker 5 (from Birmingham, Midlands) = Midlands
	2.9%	Speaker 3 (from Hastings, South East) = South/South Eas

Figure 4.4.1 Map of salience rates for Welsh respondents

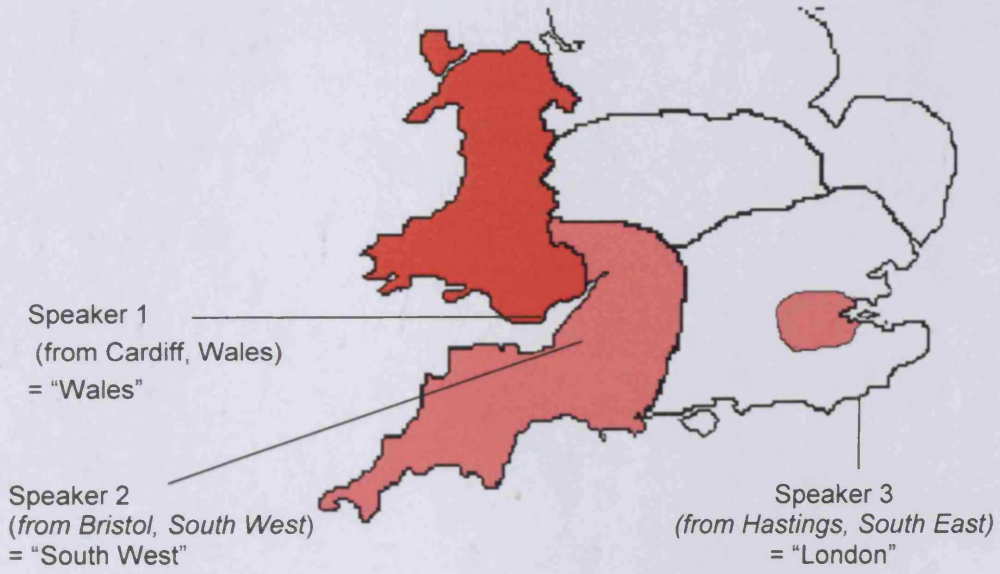


Table 4.4.2 Salience rates for West/South West respondents

	77.8%	Speaker 2 (from Bristol, South West) = West Country
	64.4%	Speaker 3 (from Hastings, South East) = London
	60.0%	Speaker 1 (from Cardiff, Wales) = Wales
	44.4%	Speaker 4 (RP) = South/South East
	13.3%	Speaker 5 (from Birmingham, Midlands) = Midlands
	6.7%	Speaker 4 (RP)= London
	6.6%	Speaker 3 (from Hastings, South East) = South/South Eas

Figure 4.4.2 Map of salience rates for West/South West respondents

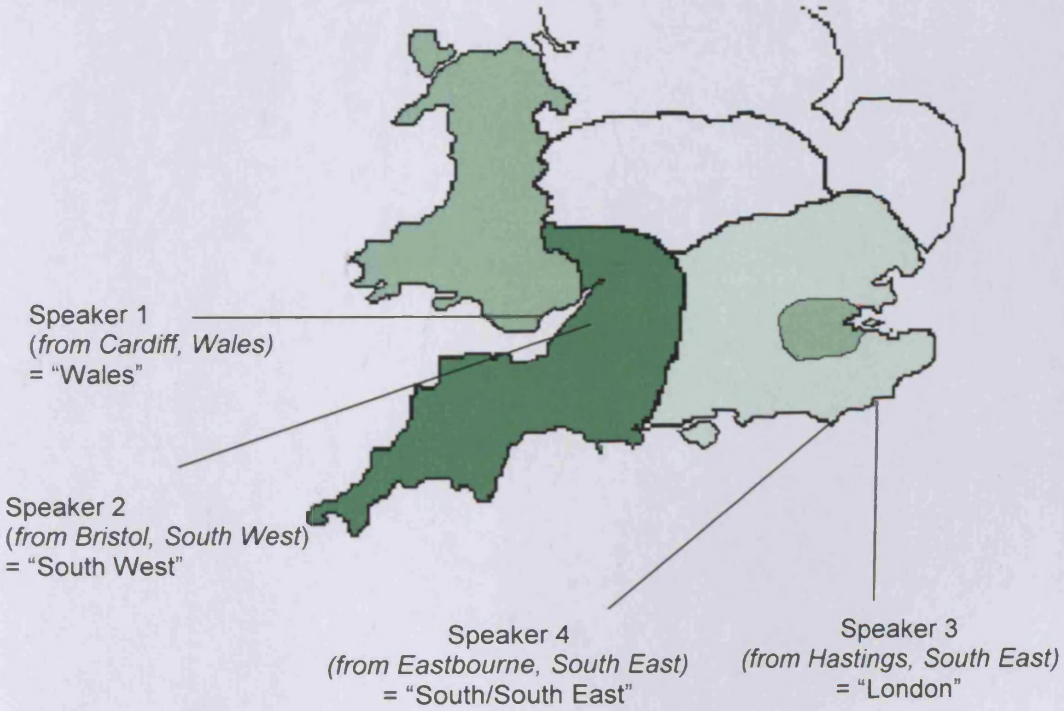


Table 4.4.3 Salience rates for South East respondents

	74.5%	Speaker 2 (from Bristol, South West) = West Country
	67.4%	Speaker 3 (from Hastings, South East) = London
	55.8%	Speaker 1 (from Cardiff, Wales) = Wales
	51.2%	Speaker 4 (RP) = South/South East
	18.6%	Speaker 4 (RP) = London
	16.3%	Speaker 3 (from Hastings, South East) = South/South Eas
	16.3%	Speaker 5 (from Birmingham, Midlands) = Midlands

Figure 4.4.3 Map of salience rates for South East respondents

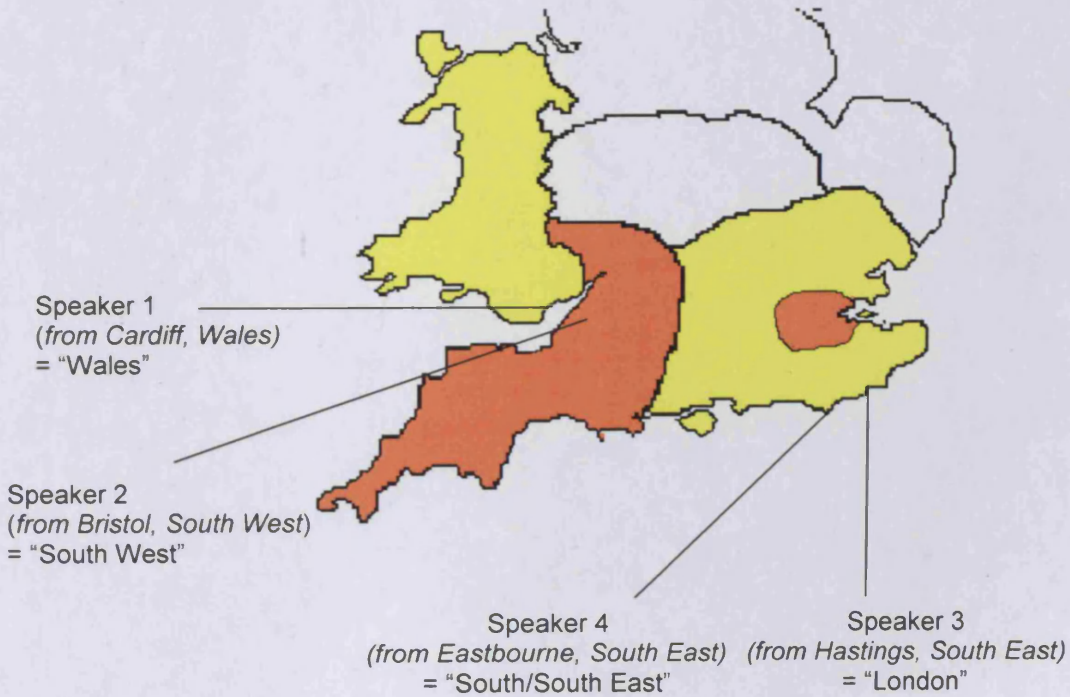
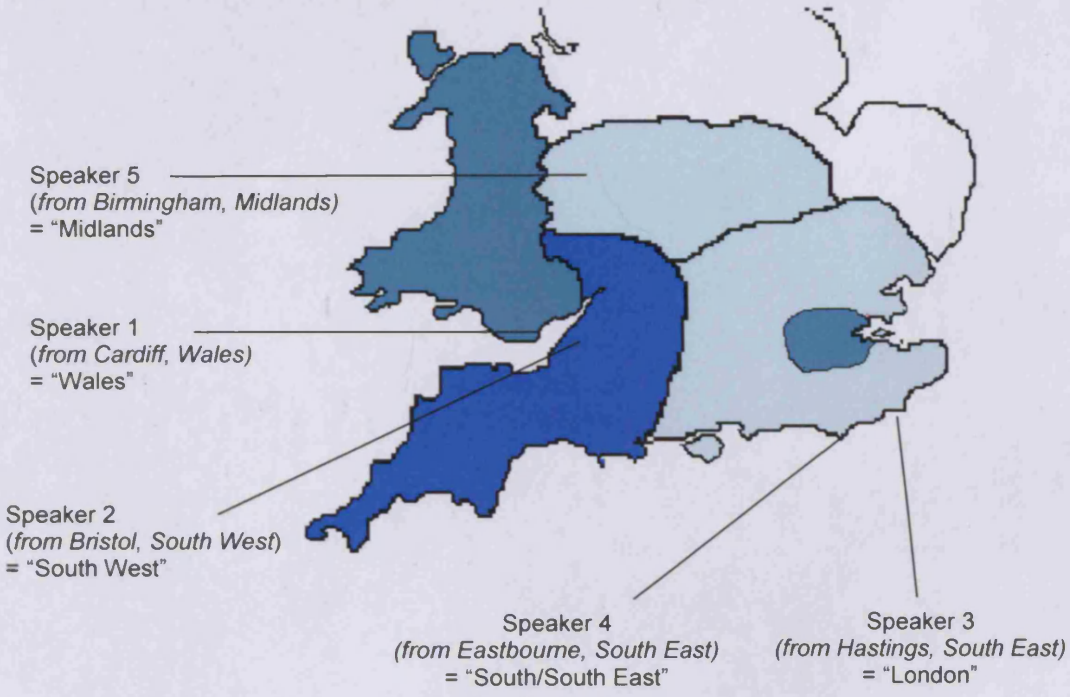


Table 4.4.4 Salience rates for Midlands respondents

	61.6%	Speaker 2 (from Bristol, South West) = West Country
	46.2%	Speaker 1 (from Cardiff, Wales) = Wales
	38.5%	Speaker 3 (from Hastings, South East) = London
	30.8%	Speaker 4 (RP) = South/South East
	30.8%	Speaker 5 (from Birmingham, Midlands)= Midlands
	15.4%	Speaker 4 (RP) = London
	7.7%	Speaker 3 (from Hastings, South East) = South/South Eas

Figure 4.4.4 Map of salience rates for Midlands respondents



For Welsh respondents, the Welsh English speaker was consistently identified as such; the two southern English accents were also well identified; however, the RP and 'accentless' Midlands speaker elicited far fewer accurate identifications. For the respondents from the West/South West, the speaker

from their own region was also the most frequently correctly identified and neighbouring South East and Welsh accents were also strongly identified. In contrast to the Welsh respondents, nearly 50% of respondents from the West/South West placed the RP speaker as coming from the South/South East of England. Similarly high numbers of South East respondents correctly identified the South West and South East speakers, with a strong recognition also for the Welsh speaker. As with the respondents from the South West, the South East respondents also placed the RP speaker within the South/South East. The 'accentless' Midlands speaker was once again the least accurately identified. Due to the low numbers of respondents from the Midlands, these figures are not seen as statistically robust and there is a need for caution in interpreting the findings.

4.5 Estimations of speaker age

Sections 4.1–4.4 dealt with respondents' estimations of speaker region. Section 4.5 deals with a different aspect of respondents' estimation, speaker age. The data below detail respondents' estimations of the five speakers' ages. Each speaker is listed separately, detailing the frequency of attribution by age group, given in table and histogram form.

Figure 4.5.1 Estimations of Speaker 1's age

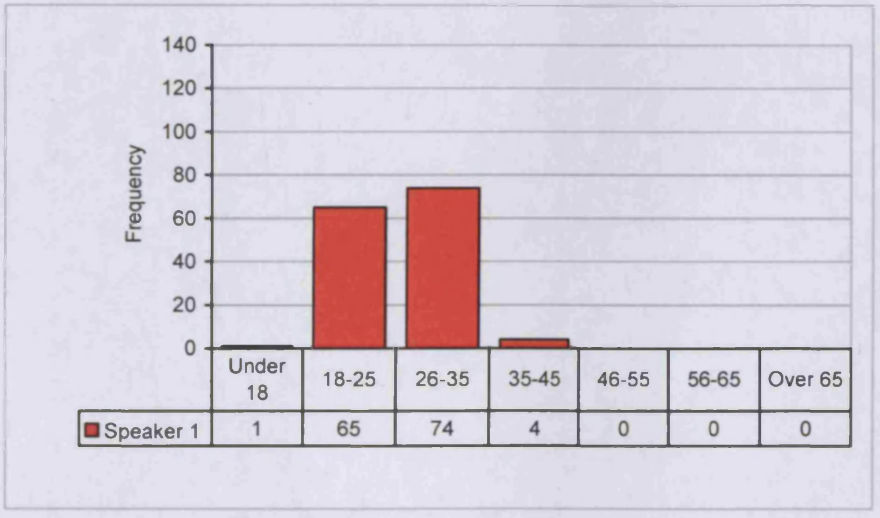


Figure 4.5.2 Estimations of Speaker 2's age

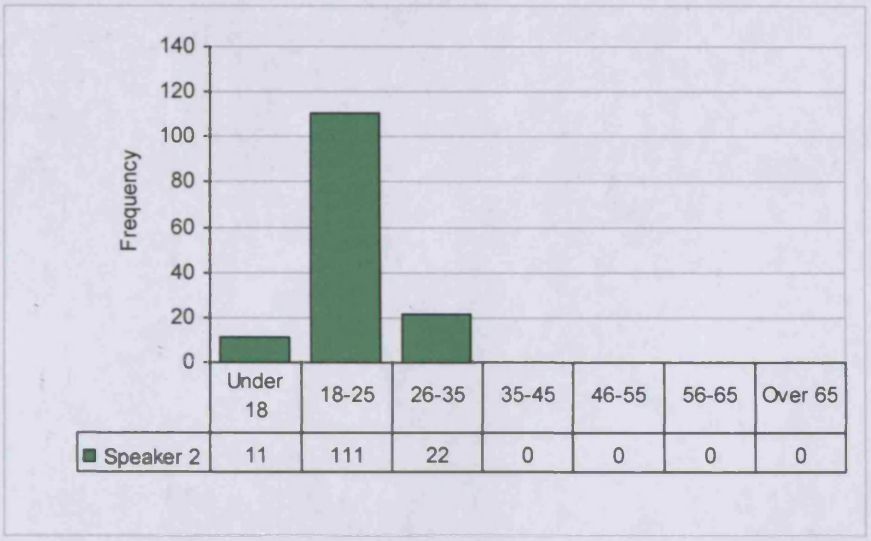


Figure 4.5.3 Estimations of Speaker 3's age

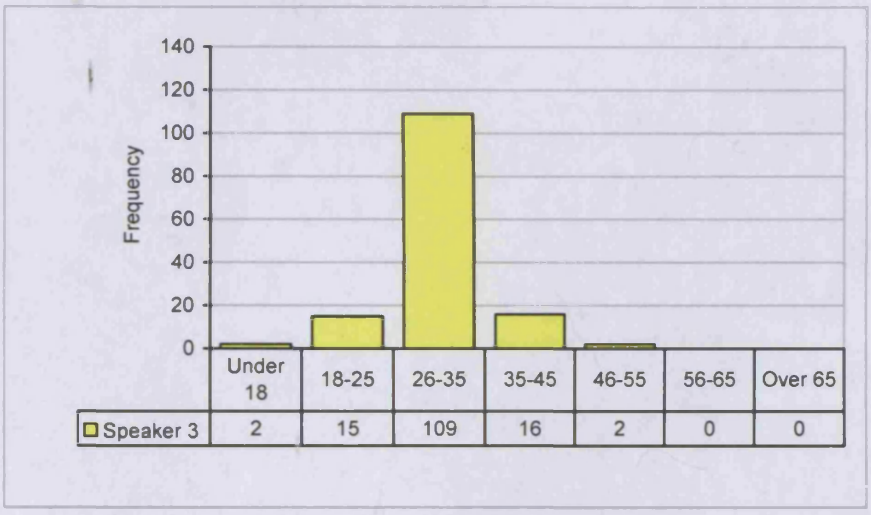


Figure 4.5.4 Estimations of Speaker 4's age - graphic representation

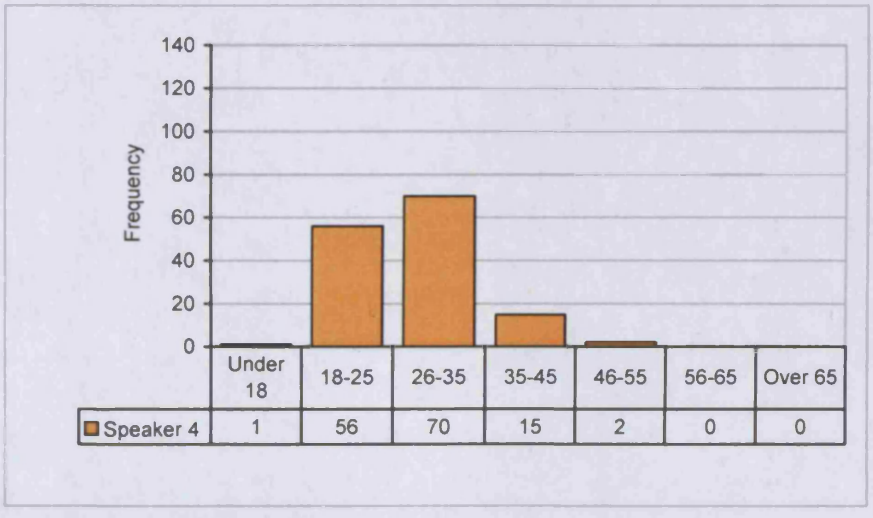
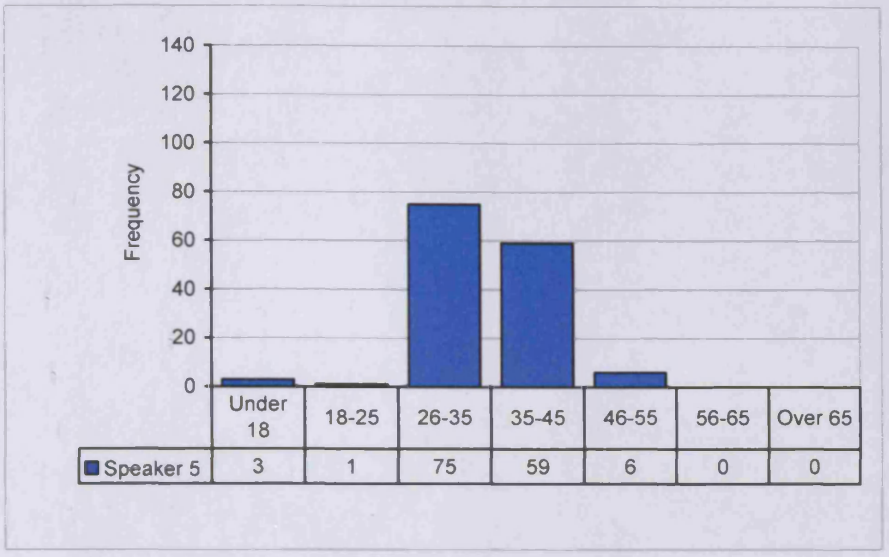


Figure 4.5.5 Estimations of Speaker 5's age - graphic representation



For the age estimations of Speakers 1 and 4, over 95% and 87% respectively of the respondents felt that the speakers' age was between18-35, and there was an almost 50-50 split in opinion as to whether the speaker was in the 18-25 or 26-35 age group. These data are perhaps less surprising for Speaker 1, whose real age is 26, on the cusp of the two age spans, and a little more so for Speaker 4, who was the youngest of the voice donors at 21, meaning that 60% of respondents misjudged the speaker's age. From the estimations that follow in the next section, it is clear respondents believed the Speaker to have a deeper voice, yet this did not prevent the respondents from over-estimating his age, indicating that depth of voice alone cannot predict age estimations.

In contrast to Speakers 1 and 4, the majority of respondents – three quarters of them – estimated Speaker 2's age to be between 18 and 25, indicating a greater degree of certainty, younger than his real age of 27. Only 15% of respondents provided an age estimation of between 26-35. In terms of quantities, Speaker 3 attracts very similar strength of belief regarding his age as Speaker 2; for Speaker 3, however, the majority of respondents estimated the speaker's age to be between 26 and 35, his real age being 30.

Overall, Speaker 5 is estimated to be the oldest speaker present, with 93% of respondents believing the speaker to be between 26 and 45. Although this

means that the 52% of respondents that estimated his age to be between 26 and 35 were right, the 45% that estimated his age to be 35–55 were incorrect.

Having screened the voice samples for factually-neutral content, it is possible that the age impressions are the result of physiology and articulation and we must ask whether widely dispersed and incorrect age estimations are related to the speaker's RP-like accent or perhaps to some other aspect of physiology, such as vocal features, i.e. depth of voice? Alternatively, it is possible that the possible age ranges provided are too narrow for respondents to estimate speaker age through the vocal channel. It may be more reasonable to have wider age ranges, for example, 18–30, or even wider. What we do know is that, overall, respondents did not excel at estimating age within the parameters provided.

4.6 Qualitative descriptions of the speakers' voices

Section 4.2–4.5 considered the more factual aspects of describing a speaker – region and age. Section 4.6 presents data regarding less conventional (certainly in forensic terms) non-linguist descriptions.

In order to explore respondents' perceptions through non-researcher prescribed channels the questionnaire invited respondents to supply their own qualitative comments on the voices encountered. In total, the non-linguists provided a corpus of 1068 voice descriptions, in the form of single word comments or phrases. The contents of these descriptions were analysed for themes, and from this analysis emerged a distinct set of qualitative subcategories invoked by non-linguists to articulate their perceptions regarding the five voices encountered:

Age, e.g. mature, young (comments additional to earlier age estimations)
Voice qualities, e.g. husky, rough
Accent
Description, e.g. regional, rural, posh, no accent
Location e.g. Welsh, West Country
Delivery
Pitch, e.g. deep, like a girl's voice
Speaking tone, e.g. monotone
Force, e.g. loud, soft
Audibility, e.g. clear
Speed, e.g. slow, fast
Fluency, e.g. stutters, hesitant, flowing
Speech impediment, e.g. possible lisp
Education, e.g. well-educated
Type of person, e.g. laddish, bloke, roughish
State of mind/Personality, e.g. relaxed, calm, aggressive
Believability, e.g. genuine, coherent, vague

One option would be to include data on the basis of frequency, excluding those items with low-frequency occurrence. However, the aim here was to collect a full range of comments to be re-presented to a second cohort of respondents at a later point. For this reason the data that appear below is not organised or annotated by frequency but by theme. A full list of all items including frequency is available in the Appendices D. The thematically organised comments appear in Table 4.6.1:

The qualitative subcategories that emerged from the non-linguist data are remarkably similar to those described by the professional forensic linguist Hollien (1990) (see section 1.1.4) when performing forensic speaker analysis:

Figure 4.6.1 - Professional linguist vs. non-linguist comments on voice qualities

Hollien's categorisation	Non-linguist derived categorisation
Articulation	-
Fundamental frequency/heard pitch	Pitch; Tone
General voice quality	Voice quality
Prosody	Speed; fluency; speaking tone
Vocal intensity	Force
Speech characteristics:	
- Dialect	- Dialect
- Unusual linguistic stress/affect	-
- Idiosyncratic language patterns	-
- Speech impediments	- Speech impediment
- Idiosyncratic pronunciations.	-

Placing the non-linguist derived categories alongside those taken from Hollien, there is evidence to suggest that certain of the categories in Hollien's taxonomy, such as 'articulation' or 'idiosyncratic' production, do not appear in non-linguists' categories. This is possibly due to the nature of the vocal percepts simply not displaying those absent features, or an example of non-linguists invoking less scientifically rigorous descriptions, or alternatively such concepts may in a Prestonian sense simply not be available to non-linguists (Preston 1996).

Table 4.6.1 Thematic organisation of respondents' qualitative comments.

	SPEAKER 1	SPEAKER 2	SPEAKER 3	SPEAKER 4	SPEAKER 5
AGE	mature, older	young	young, immature		
	26-35 (real age=27)	18-25 (real age=26)	18-25 (real age=33)	26-35 (real age=20)	18-25 (real age=26)
VOICE QUALITIES	husky	croaky, dry, possible speech impediment		rough, husky	
ACCENT Type	strong regional accent, stresses vowels, Welsh, Cardiff	non-standard, regional, farmer, rural, common, South West, West Country	strong regional accent, no well spoken, cockney, common, rough, pikey	no accent, well-spoken, RP, middle class, posh, public school boy	no accent, Southern, well-spoken, polite, middle class
Locations	Wales 66%, Cardiff 35% (Don't know 14%)	West Country 67%, Bristol 20% (Don't know 19%)	S. East 69%, London 62%, (Don't know 14%)	South 44%, (Don't know 44%)	14% Midlands, (44% Don't know)
DELIVERY Pitch	deep, low, masculine,	medium, high, like a boy's voice like a girl's voice, feminine	high, high-pitched, higher than normal, medium	deep	natural, medium, deep
Speaking tone	monotone, expressionless, rises at end of sentences	monotonous	monotone	monotone, even, expressionless, sombre	
Force	loud	quiet, soft			well-projected
Quality	clear	clear	clear	clear	clear
Speed	slow	slow	fast, quick	slow	
Fluency		pausing, stutters, stumbles, stammers	hesitant, jumpy		expressive, flowing, lively, careful, controlled
Speech impediment		possible lisp?			
EDUCATION				well-educated	
TYPE OF PERSON	laddish, bloke, roughish	farmer	pikey	public school boy	
STATE OF MIND/ PERSONALITY	relaxed, calm, confident, assertive, aggressive	unsure, hesitant, nervous, worried, emotional, looking for assurance	unsure, nervous, worried, edgy, emotional, cocky, confident, expressive	calm, confident, serious	calm, relaxed, confident, self-assured, arrogant, uncertain, hesitant, nervous
BELIEVABILITY	has to think about what he's saying, thinking about answer, pre-planned, trying to sound convincing, rehearsed, detailed, contradicts himself	genuine, innocent voice, most facts well-described, telling the truth, truthful, questions himself	uses lots of fillers, not sure about his story, knows the order of events very well, guilty, innocent	genuine, knows his story, consistent, vague, unsure of the details, trying to come across as unworried, too confident	certain, coherent, confident, honest, not too sure, vague, trying to be unrehearsed

Table 4.6.1 Thematic organisation of respondents' qualitative comments.

	SPEAKER 1	SPEAKER 2	SPEAKER 3	SPEAKER 4	SPEAKER 5
AGE	mature, older	young	young, immature		
	26-35 (real age=27)	18-25 (real age=26)	18-25 (real age=33)	26-35 (real age=20)	18-25 (real age=26)
VOICE QUALITIES	husky	croaky, dry, possible speech impediment		rough, husky	
ACCENT Type	strong regional accent, stresses vowels, Welsh, Cardiff	non-standard, regional, farmer, rural, common, South West, West Country	strong regional accent, no well spoken, cockney, common, rough, pikey	no accent, well-spoken, RP, middle class, posh, public school boy	no accent, Southern, well-spoken, polite, middle class
Locations	Wales 66%, Cardiff 35% (Don't know 14%)	West Country 67%, Bristol 20% (Don't know 19%)	S. East 69%, London 62%, (Don't know 14%)	South 44%, (Don't know 44%)	14% Midlands, (44% Don't know)
DELIVERY Pitch	deep, low, masculine,	medium, high, like a boy's voice like a girl's voice, feminine	high, high-pitched, higher than normal, medium	deep	natural, medium, deep
Speaking tone	monotone, expressionless, rises at end of sentences	monotonous	monotone	monotone, even, expressionless, sombre	
Force	loud	quiet, soft			well-projected
Quality	clear	clear	clear	clear	clear
Speed	slow	slow	fast, quick	slow	
Fluency		pausing, stutters, stumbles, stammers	hesitant, jumpy		expressive, flowing, lively, careful, controlled
Speech impediment		possible lisp?			
EDUCATION				well-educated	
TYPE OF PERSON	laddish, bloke, roughish	farmer	pikey	public school boy	
STATE OF MIND/ PERSONALITY	relaxed, calm, confident, assertive, aggressive	unsure, hesitant, nervous, worried, emotional, looking for assurance	unsure, nervous, worried, edgy, emotional, cocky, confident, expressive	calm, confident, serious	calm, relaxed, confident, self-assured, arrogant, uncertain, hesitant, nervous
BELIEVABILITY	has to think about what he's saying, thinking about answer, pre-planned, trying to sound convincing, rehearsed, detailed, contradicts himself	genuine, innocent voice, most facts well-described, telling the truth, truthful, questions himself	uses lots of fillers, not sure about his story, knows the order of events very well, guilty, innocent	genuine, knows his story, consistent, vague, unsure of the details, trying to come across as unworried, too confident	certain, coherent, confident, honest, not too sure, vague, trying to be unrehearsed

However, perhaps most striking is what the two sets *do* have in common: there is a large overlap between the observations and conceptualisations of the non-linguists and the professional linguist, regarding 'tone', 'quality' 'speed', 'force' and 'dialect'. Moreover, Table 4.6.1 demonstrates that the non-linguists did not stop at voice characteristics, but also supplied qualitative comments regarding age, education, type of person, state of mind/personality and believability, indicating that many non-linguists are ready and equipped to volunteer more than a comment upon a speaker's accent, in contrast to the apparent focus of many forensic enquiries (see Chapter 1 for an illustration of the nature of non-linguist police-earwitness interactions).

A final remark should be made regarding the inclusion of comments on 'believability'. Such comments are a voluntary addition to the elicited responses regarding 'speaker guilt' (see section 4.10 below) as they appear to be in some way indicating a context or rationale for the respondents' decisions, as if certain respondents had been looking for signs of 'leakage' in the management of the speakers' discourse, perhaps exposing examples of duplicity. Given that no real crime was committed by any of the speakers, the respondents could not have been responding to any genuine attempt by the speakers to conceal their guilt. This may be indicative of the many ways in which perceptions of speaker accent as well as voice may influence judgement patterns or magnify prejudicial attitudes. Such comments are well outside the remit of the forensic linguist and would not be considered for inclusion in a typology of forensic voice descriptions.

4.7 Estimations of speaker intelligence

In the first four sections, data were presented regarding the more factual aspects of speaker voice description – accent, speaker region and age – those features being of greatest ecological relevance to current lines of forensic enquiry. Section 4.6 presented an extension of those descriptions in inviting speakers to describe not only accent and age, but also individual voice characteristics using their own non-linguist terms. From Section 4.7 onwards, data are presented from the results of inviting respondents to provide

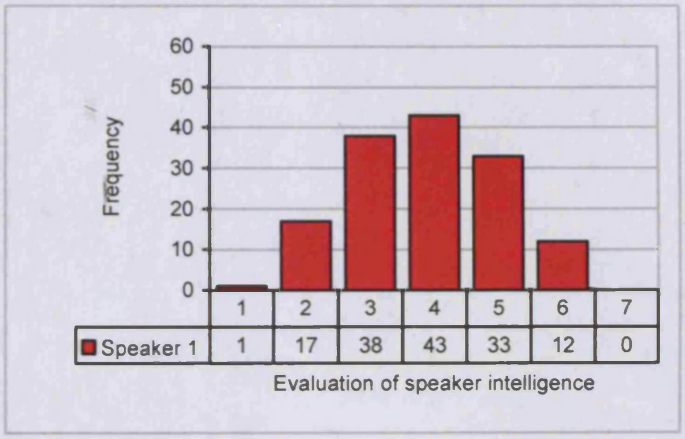
judgements regarding personal attributions of the speakers based only on the vocal channel.

It is important to re-iterate the purpose for such a line of enquiry: The aim in these following sections is not to elicit data that would later form the basis for a typology of speaker descriptions. Rather, it is suggested that whilst eliciting subjective responses from non-linguists regarding how they differentiate and describe voices, there is great value in investigating what other processes may interact with both accent recognition and attributions of speaker guilt (section 4.10, below). As discussed in Chapter 2, a long tradition of language attitudes work suggests that differing accents do provoke varied positive and negative assessments. These will be the focus of sections 4.7–4.10. Section 4.11 will address the question of whether there are any interactions between, for example voice characteristics/speaker accent and speaker guilt, and whether subjective accent assessment may be congruent with high or low assessments of speaker guilt.

For all ratings in the following four sections, respondents were provided with seven-point, bipolar semantic differential scales upon which they were invited to mark their assessments regarding whether they felt the speakers were, for example, more 'intelligent' (up to 7 on the scale) or more 'unintelligent' (down to 1 on the scale). As the medium point on the scale was 4, we are interested to see firstly if the assessments of individual speakers diverge from the middle point, and if so, in which direction. The evaluations for each speaker are given first in the form of bar charts, showing frequency counts for each evaluative point on the scale; this is followed by a summary table of means and finally by the results of ANOVAs. The ANOVAs were conducted to investigate whether any differences that arise in the data were significant or simply a matter of chance. Repeated measures ANOVAs were conducted for each evaluative dimension in order to test for equality and differences between multiple paired data. The ANOVA first checked for significant differences in the data using Mauchly's test of sphericity. If the data did not violate the sphericity assumption, the significance for the factor was taken from the 'sphericity assumed' reading. If the data did violate the sphericity assumption, a Greenhouse Geisser

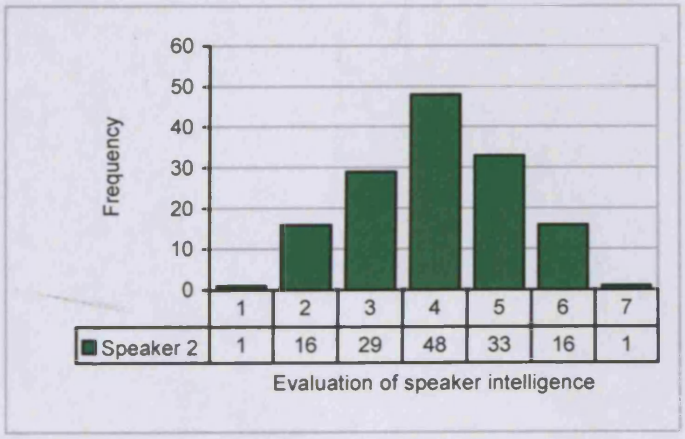
correction factor was applied to the degrees of freedom used to assess the observed F-ratio and a significance reading taken. Results were subject to post-hoc LSD treatment and these workings are not reported in the final results. What is reported here is the final significance readings for the individual rating by all respondents for each speaker with regard to the evaluative dimension under analysis. Where the mean difference is significant, the values are shaded for ease of interpretation.

Figure 4.7.1 Intelligence ratings for Speaker 1 - Wales



n=144
Mean =3.9
Std. Dev.=1.16

Figure 4.7.2 Intelligence ratings for Speaker 2 - South West



n=144
Mean =4.0
Std. Dev.=1.21

Figure 4.7.3 Intelligence ratings for Speaker 3 - South East

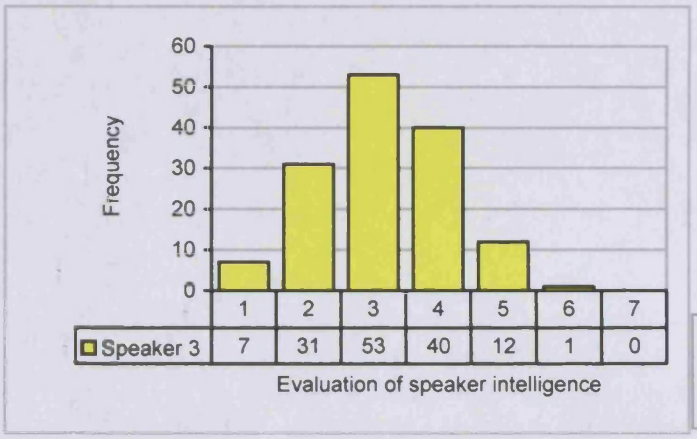


Figure 4.7.4 Intelligence ratings for Speaker 4 - South East (RP-like)

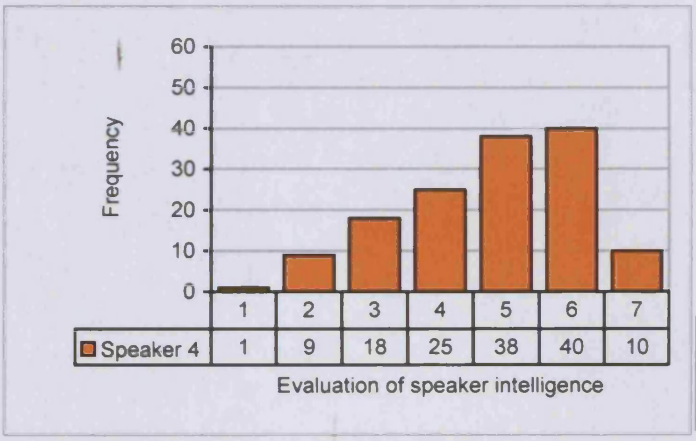


Figure 4.7.5 Intelligence ratings for Speaker 5 - Midlands

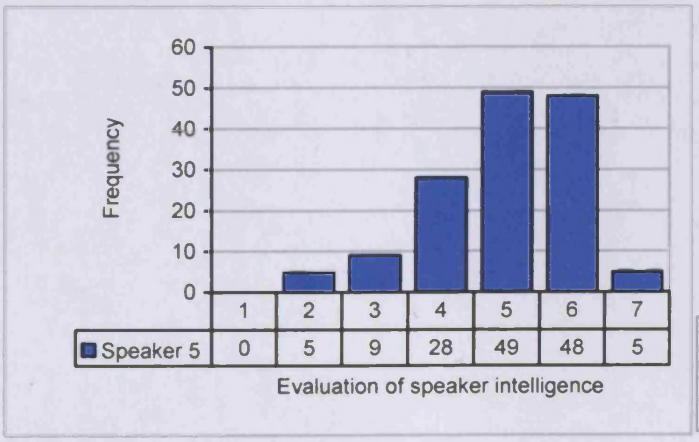
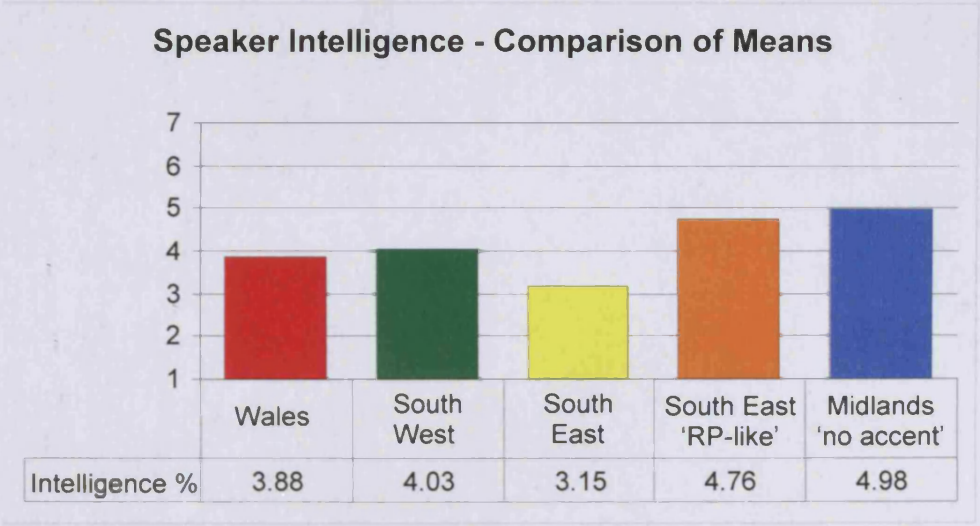


Figure 4.7.6 Speaker intelligence: table of means for each speaker



	Wales	South West	South East	South East 'RP-like'	Midlands 'no accent'
Intelligence %	3.88	4.03	3.15	4.76	4.98

Figure 4.7.7 Repeated measures ANOVA for Speaker Intelligence ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 Intelligence	0.267
Speaker 1 - Speaker 3 Intelligence	0.001
Speaker 1 - Speaker 4 Intelligence	0.001
Speaker 1 - Speaker 5 Intelligence	0.001
Speaker 2 - Speaker 3 Intelligence	0.001
Speaker 2 - Speaker 4 Intelligence	0.001
Speaker 2 - Speaker 5 Intelligence	0.001
Speaker 3 - Speaker 4 Intelligence	0.001
Speaker 3 - Speaker 5 Intelligence	0.001
Speaker 4 - Speaker 5 Intelligence	0.126

Adjustment for multiple comparisons: Least significant difference

The tables above do suggest variation between the assessments of each speaker's intelligence. Both Speaker 1 (Cardiff, Wales) and Speaker 2 (Bristol, South West) elicit responses around the middle point of 4 indicating that overall the speakers were not regarded as either remarkably intelligent or unintelligent. However, Speaker 3 (Hastings, South East) appears to provoke rather more negative assessments. In stark contrast, both Speakers 4 (South East, RP-like) and 5 (Midlands, 'no accent') receive assessments much above the medium 4 score. There would appear to be a division in the evaluation of accents, therefore, with those voices displaying more overt regional features being

evaluated more negatively in terms of intelligence, compared to the RP or 'regionally neutral' voices which are seen as largely more intelligent.

For all tests, with the exception of Speaker 1–Speaker 2 intelligence, and Speaker 4–5 intelligence, the low p-values, between the other speakers, for example Speaker 1–Speaker 3, $\rho<0.00$, give us strong reasons to suspect that the treatment, i.e. the use of different voices as stimuli, has had an effect, and that the differences observed between the means were not a coincidence of random sampling.

On the other hand, the larger ρ-values observed between Speakers 1–2, and Speakers 4–5 do not give us any evidence to conclude that Speakers 1 and 2 and Speakers 4 and 5 are seen differently by the respondents. This is, of course, not the same as saying that the treatment had no effect. However, this does suggest that there may be three distinct observable rating levels for the five speakers; one for Speakers 1 and 2, one for Speaker 3 and one for Speakers 4 and 5.

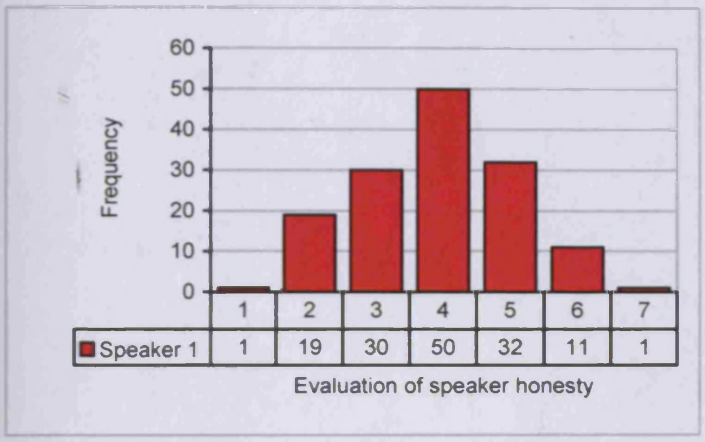
In terms of social stereotypes, one might argue that the respondents grouped and rated the speakers in rather predictable ways: The 'estuary' voice coming lowest, the two rural and Welsh voices being in the middle, and RP and the 'accentless' speaker being ranked highest.

We must remember that the judges were responding only to one token of each speech variety and that there could be other features at work, such as vocal characteristics and stylistic features of delivery. The following two sections will investigate whether similar patterns occur across the other dimensions for 'honesty' and 'aggression', and whether these patterns group speakers together in a similar way in order to expand our view of the responses.

4.8 Estimations of speaker honesty

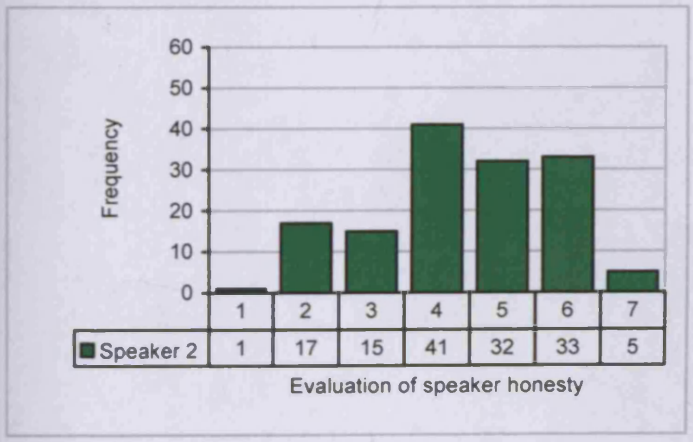
Using the same format as for assessments of speaker 'intelligence', respondents were invited to mark their feelings with regards to speaker 'honesty', being presented with a seven-point bipolar semantic differential scale with 'honest' (up to 7 on the scale) and 'dishonest' (down to 1 on the scale) placed at either end, a score of 4 being the middle point. Below are histograms representing the respondents' assessments, speaker by speaker:

Figure 4.8.1 Honesty ratings for Speaker 1 - Wales



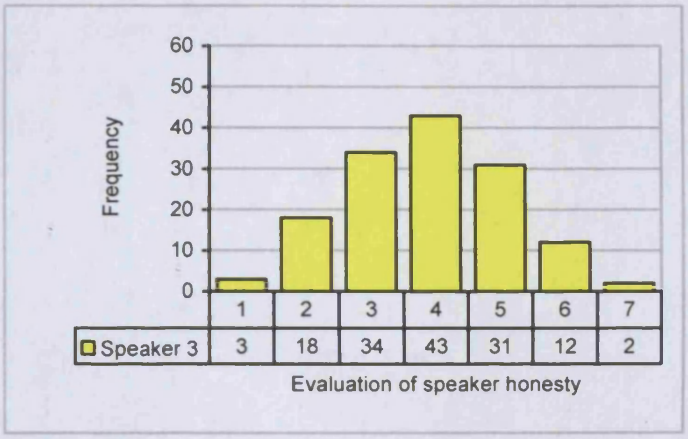
n=144
Mean =3.9
Std. Dev.=1.18

Figure 4.8.2 Honesty ratings for Speaker 2 - South West



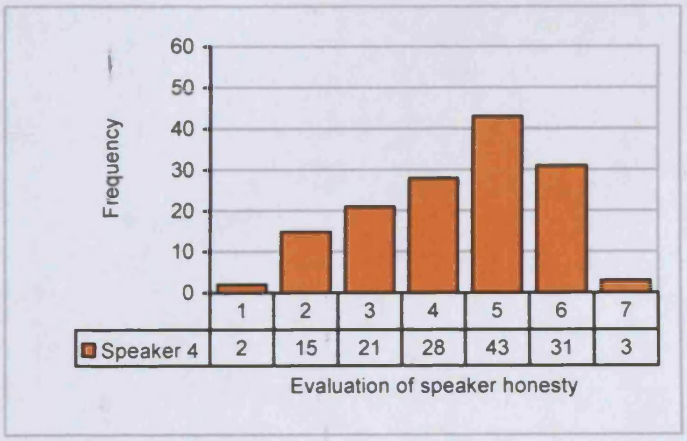
n=144
Mean =4.4
Std. Dev.=1.39

Figure 4.8.3 Honesty ratings for Speaker 3 - South East



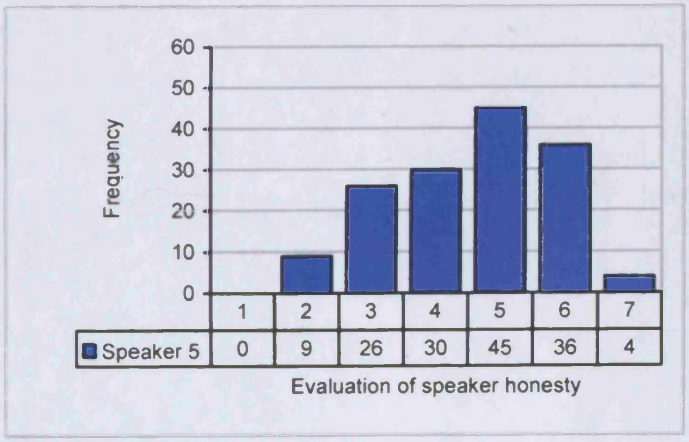
n=144
Mean =3.8
Std. Dev.=1.30

Figure 4.8.4 Honesty ratings for Speaker 4 - South East, 'RP-like'



n=142
Mean =4.4
Std. Dev.=1.42

Figure 4.8.5 Honesty ratings for Speaker 5 - Midlands 'no accent'



n=144
Mean =4.6
Std. Dev.=1.25

The histograms above once again reveal some differences between the assessments for each speaker. Speaker 1 and Speaker 3 both receive

assessments almost exactly on the middle score of 4 (3.9 and 3.8) with an almost symmetrical bell curve. The picture for Speaker 2 is more complex with a higher mean of 4.4 but with a disrupted bell curve, indicating a certain degree of polarisation of views. Nevertheless, more than two thirds of respondents rated the speaker to be 4+ on the scale, indicating that in general there is a positive consensus. The data for Speakers 4 and 5, (RP-like and 'regionally neutral') yield noticeably positive evaluations, both with high means than (4.4 and 4.6) and with slight skewedness in the distribution.

Figure 4.8.6 Speaker honesty: comparison of means for each speaker

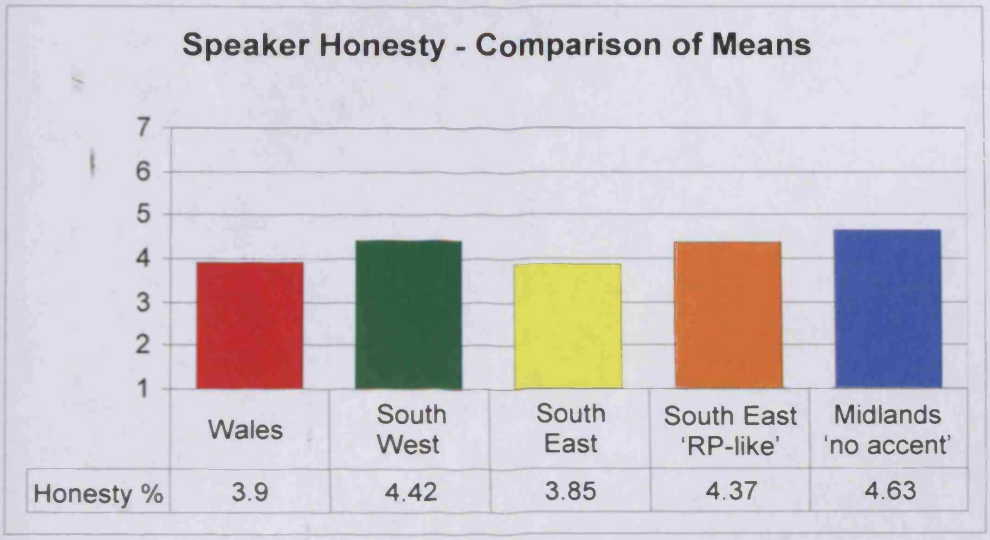


Figure 4.8.7 Repeated measures ANOVA for speaker honesty ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 Honesty	0.001
Speaker 1 - Speaker 3 Honesty	0.676
Speaker 1 - Speaker 4 Honesty	0.001
Speaker 1 - Speaker 5 Honesty	0.001
Speaker 2 - Speaker 3 Honesty	0.001
Speaker 2 - Speaker 4 Honesty	0.731
Speaker 2 - Speaker 5 Honesty	0.167
Speaker 3 - Speaker 4 Honesty	0.001
Speaker 3 - Speaker 5 Honesty	0.001
Speaker 4 - Speaker 5 Honesty	0.071

Adjustment for multiple comparisons: Least significant difference

Respondents appear to have evaluated the speakers differently along the lines of honesty. Speakers 2 (Bristol), 4 (RP-like) and 5 ('regionally neutral') appear

to 'cluster' in that there is no evidence to suggest that they are viewed significantly differently from each other in terms of their honesty. Likewise, Speakers 1 and 3 are not seen as significantly different in terms of their honesty. However, although the means are similar, we can observe two possible distinct groups which are consistently and significantly viewed as different in terms of their honesty: Speakers 1 (Wales) and 3 (South East) versus Speakers 2 (South West), 4 (South East, RP-like) and 5 (Midlands).

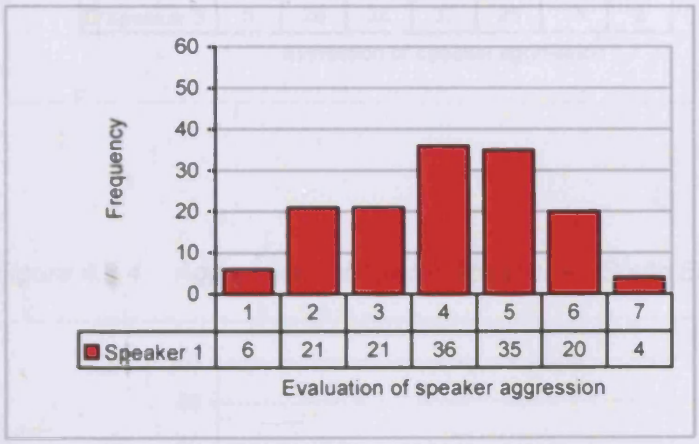
The motivation for these responses is a matter of speculation. However, we can see a number of familiar patterns: firstly, for 'speaker intelligence' and 'speaker honesty', both Speakers 4 (RP-like) and 5 ('regionally neutral') have been seen in positive terms by the majority of respondents. Secondly, and in contrast, evaluations for Speakers 1 (Cardiff) and 3 (Hastings) have been below 4 for ratings of ¡intelligence' and 'honesty'. These data would be consistent with previous research utilising a conceptual approach which has seen RP and 'accentless' speech appraised highly, and urban dialects frequently downgraded, receiving the lowest evaluations from the non-linguist respondents. (See Giles (1970) and Bishop et al. (2005) for examples of ratings of accents in the UK). The relationship between the concepts used in this study and those used by other studies ('social attractiveness', 'prestige', etc) will be dealt with in the Chapter 6, Discussion.

Thirdly, and of particular interest in itself, is the somewhat anomalous evaluation of Speaker 3 (Bristol): For evaluations of 'speaker intelligence', Speaker 3 received a mean of almost exactly 4, whereas for 'speaker honesty', the ratings are more positive at 4.42, the second highest evaluation overall for speaker 'honesty'. Similar patterns were observed in Giles (1970) and Bishop et al. (2005), in that speakers are rated differently depending on the evaluative dimension. Alternatively, there may be something in the speaker's vocal performance, aside from the accent, that the judges are responding to: something in the tone, pitch, speed, etc. This will be returned to in the Discussion.

4.9 Estimations of speaker aggression

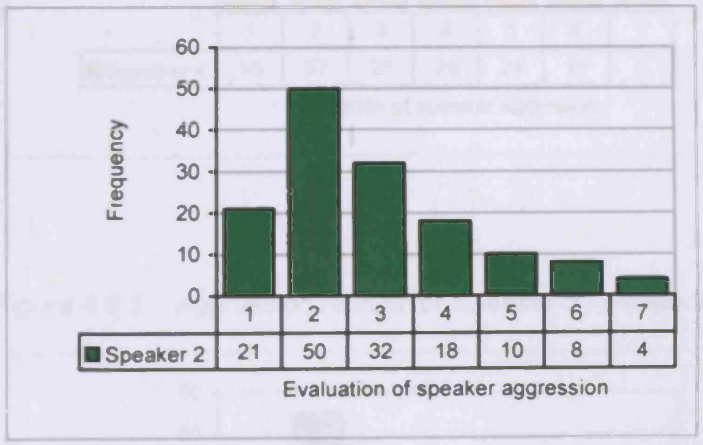
As for the previous two sections, 'speaker aggression' was evaluated by the respondents on a 7-point bi-polar semantic differential scale 'aggressive' 7 and 'unaggressive' 1. The scores for the five speakers a given below in bar charts.

Figure 4.9.1 Aggression ratings for Speaker 1 - Wales



n=144
Mean =4.0
Std. Dev.=1.51

Figure 4.9.2 Aggression ratings for Speaker 2 - South West



n=144
Mean =2.9
Std. Dev.=1.53

Figure 4.9.3 Aggression ratings for Speaker 3 - South East

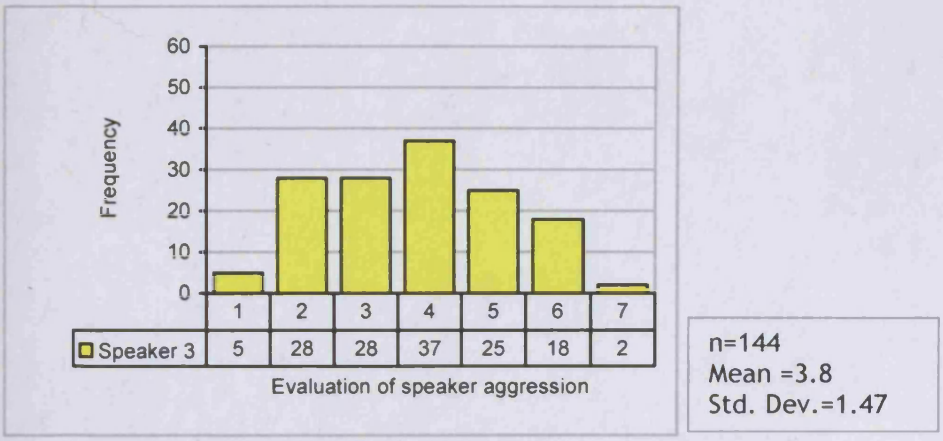


Figure 4.9.4 Aggression ratings for Speaker 4 - South East, RP-like

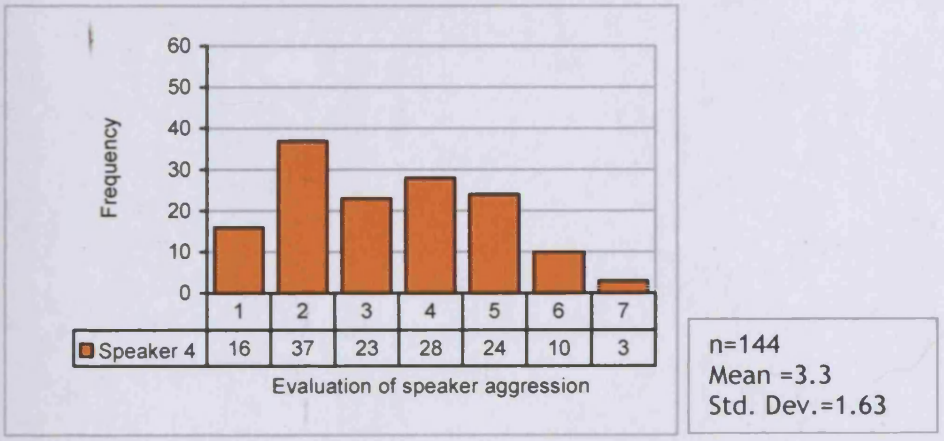


Figure 4.9.5 Aggression ratings for Speaker 5 - Midlands, 'no accent'

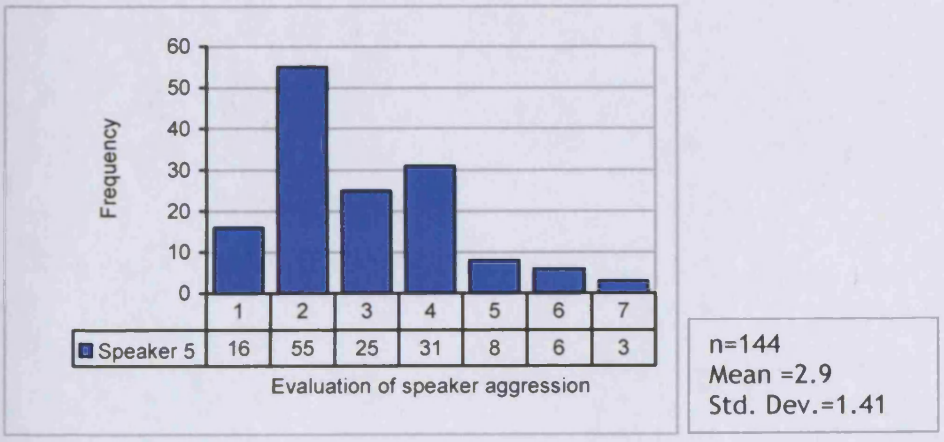


Figure 4.9.6 Speaker aggression: comparison of means for each speaker

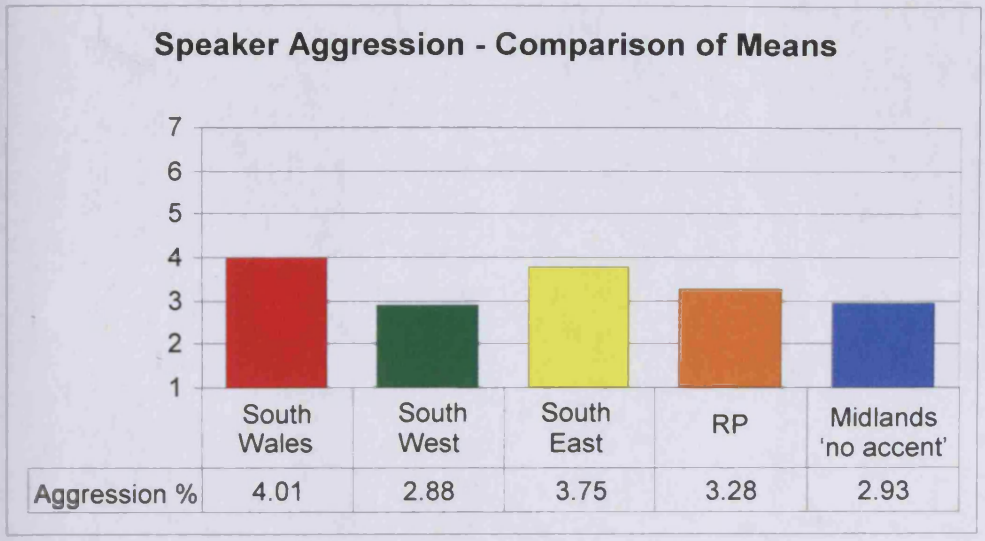


Figure 4.9.7 Repeated measures ANOVA for speaker aggression ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 Aggression	0.001
Speaker 1 - Speaker 3 Aggression	0.107
Speaker 1 - Speaker 4 Aggression	0.001
Speaker 1 - Speaker 5 Aggression	0.001
Speaker 2 - Speaker 3 Aggression	0.001
Speaker 2 - Speaker 4 Aggression	0.010
Speaker 2 - Speaker 5 Aggression	0.710
Speaker 3 - Speaker 4 Aggression	0.003
Speaker 3 - Speaker 5 Aggression	0.001
Speaker 4 - Speaker 5 Aggression	0.017

Adjustment for multiple comparisons: Least significant difference

We can note from the histograms that there is variance in the extent to which respondents shared the same views of speaker aggression. Speaker 5 (Midlands, 'no accent'), for example, has a low mean and the lowest standard deviation, indicating a large amount of consensus that the speaker is not particularly aggressive; Speaker 4 (South East, 'RP-like'), by contrast, has a standard deviation of 1.63, indicating that there was there was a fair amount of disagreement between the speakers.

Speakers 1 and 3 are not viewed significantly different from each other, and despite not being evaluated as either highly aggressive or lacking aggression, they are viewed as significantly more aggressive than all other speakers. We can also see how Speakers 2 and 5, who have been evaluated positively across the previous dimensions, continue to enjoy significantly more positive evaluations than the other speakers with lower evaluations of aggression. Across all of the dimensions (intelligence, honesty and aggression) these are by far the lowest means scores that the respondents have provided.

Based on the points discussed regarding Speaker 2 in the last section on 'speaker honesty', it is of note that Speaker 2's evaluations are more inclined towards the pattern for Speakers 4 and 5's evaluations rather than those for Speaker 1 and 3. This is unsurprising given the findings of both Giles (1970) and Bishop et al. (2005) which suggest that on both the dimensions of social attractiveness and prestige, the West Country accent is more positively evaluated than a speaker from Wales or London, although further discussion will follow in Chapter 6 as to the nature of the similarity between the concepts utilised by Giles and Bishop et al. and this study.

4.10 Attributions of guilt

4.10.1 Crime type

In Study 1, respondents were told that the five voices that they had heard were those of five people who had been detained by the police after a crime had been committed. Respondents were told that one or none of the men may have committed the crime but that there was the chance that all of them could be innocent. They were then asked the following question:

"Based on the voices that they have heard, what is the likelihood that any of the speakers may have committed a (suggested on the questionnaire) crime?"

Following Seggie (1983), where respondents were seen to vary in their attributions of guilt according to the crime described to them, two crimes were

suggested: 52% of the respondents were told that the crime that had been committed was that of assault and robbery and the remaining 48% of respondents were told that the crime was one of sexual assault. The attributions of guilt were subsequently analysed, comparing the means for each crime, (Figures 4.10.1.1; 4.10.1.3). A repeated measures ANOVA was used to assess the equality of means and identify any significant differences between the evaluations (Figures 4.10.2; 4.10.4).

Figure 4.10.1.1 Attributions of speaker guilt: means for sexual assault

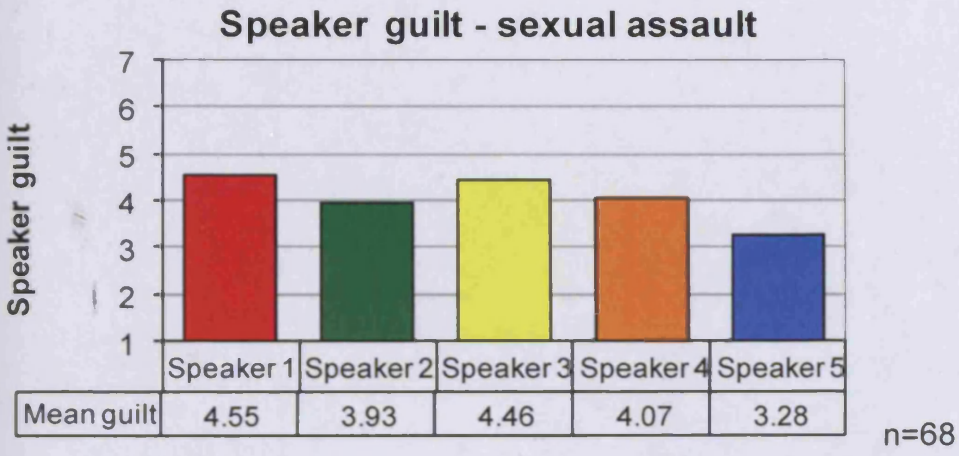


Figure 4.10.1.2 Repeated measures ANOVA for guilt - sexual assault ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 Guilt - sexual assault	0.009
Speaker 1 - Speaker 3 Guilt - sexual assault	0.690
Speaker 1 - Speaker 4 Guilt - sexual assault	0.068
Speaker 1 - Speaker 5 Guilt - sexual assault	0.001
Speaker 2 - Speaker 3 Guilt - sexual assault	0.015
Speaker 2 - Speaker 4 Guilt - sexual assault	0.634
Speaker 2 - Speaker 5 Guilt - sexual assault	0.004
Speaker 3 - Speaker 4 Guilt - sexual assault	0.172
Speaker 3 - Speaker 5 Guilt - sexual assault	0.001
Speaker 4 - Speaker 5 Guilt - sexual assault	0.001

Adjustment for multiple comparisons: Least significant difference

The data in Figures 4.10.1 and 4.10.2 show us that there are significant differences between the attributions of guilt for each speaker. Speakers 1 and 3 are seen as significantly more likely to have committed the crime than Speakers 2 and 5. Speaker 2 is seen as significantly less likely to have committed the crime than Speakers 1 and 3, yet more likely than Speaker 5. Speakers 2 and 4

are not seen as significantly different. Speaker 5 if seen as significantly less likely to have committed the crime than all other speakers. The data suggest that for the crime of sexual assault, those speakers with the urban accents are evaluated more negatively than other speakers; the stereotypical West Country speaker and RP speakers are evaluated as less likely to have committed the crime, but the speaker with the least regionally and socially marked accent, Speaker 5, is evaluated significantly more positively than the other speakers present.

Moving to the second crime scenario, we can again view the mean evaluations for each speaker, followed by repeated measures ANOVA, to look for evidence of significant differences between the evaluations.

Figure 4.10.1.3 Attributions of speaker guilt: means for assault and robbery

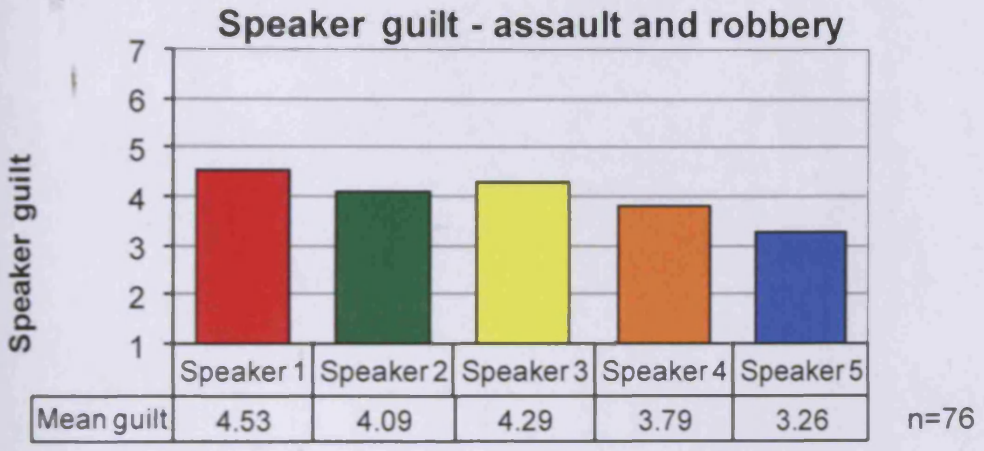


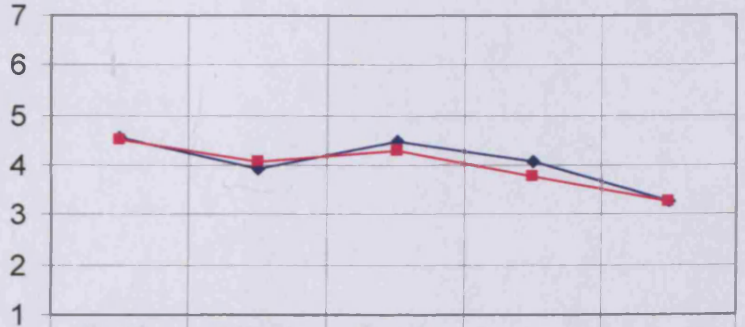
Figure 4.10.1.4 Repeated measures ANOVA for guilt - assault and robbery ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 Guilt - assault and robbery	0.090
Speaker 1 - Speaker 3 Guilt - assault and robbery	0.307
Speaker 1 - Speaker 4 Guilt - assault and robbery	0.005
Speaker 1 - Speaker 5 Guilt - assault and robbery	0.001
Speaker 2 - Speaker 3 Guilt - assault and robbery	0.421
Speaker 2 - Speaker 4 Guilt - assault and robbery	0.329
Speaker 2 - Speaker 5 Guilt - assault and robbery	0.003
Speaker 3 - Speaker 4 Guilt - assault and robbery	0.090
Speaker 3 - Speaker 5 Guilt - assault and robbery	0.001
Speaker 4 - Speaker 5 Guilt - assault and robbery	0.066

Adjustment for multiple comparisons: Least significant difference

The evaluation means of speaker guilt for the crime of assault and robbery show that Speaker 1 is seen as significantly more likely to have committed the crime than Speakers 4 and 5. Speakers 1, 2 and 3 are not seen as significantly different from each other. Speakers 2, 3 and 4 are not seen as significantly different from each other. Speaker 5 is once again seen as significantly different from Speaker 1, 2 and 3. Only Speaker 4 is not seen as significantly different from Speaker 5. What we see in these data are very similar patterns to those displayed for the crime of sexual assault. Speaker 1 is consistently viewed as more likely to have committed either crime, and likewise, Speaker 5 is consistently viewed as the least likely to have committed either crime. The means for Speaker 2 indicated that the speaker is consistently evaluated the second most likely to have committed the crime, although in neither case are the means significantly different from those for Speaker 1. Further, in neither case do the means for Speakers 2 or 4 indicate that the speakers are evaluated significantly differently. If we compare the ratings for the two groups evaluating two different crimes, what is striking about the data is the similarity between the mean evaluations for each crime. Plotting them onto the same area graph illustrates this point.

Figure 4.10.1.5 Attributions of guilt: sexual assault versus assault and robbery



	South Wales	West Country	South East	RP	Midlands
Sexual Assault	4.55	3.93	4.46	4.07	3.28
Assault and Robbery	4.53	4.09	4.29	3.79	3.26

n=144

The similarity between the means for each speaker in each crime scenario suggests that if crime type is interacting with speaker accent/voice, then each

crime is interacting in a remarkably similar way. A speculation might be that there is in fact some meta-category invoked by the suggested scenarios; in this case, 'violent personal attack'. As the respondents for each crime were not exposed to both conditions, we are unable to statistically confirm this speculation, but it should form the basis of further research. However, and as a point of conjecture based on the very close evaluations of the speakers, we could assume that the two crime scenarios are viewed identically. If so, we can run a tentative statistical analysis for the meta-crime of 'violent personal attack', synthesising the two very similar sets of results into one (Figures 4.10.6 to 4.10.12), to give us a general sense of how the speakers were viewed overall for the likelihood of committing a violent personal attack. Being derived from data for two separate groups, these results are indicative rather than statistically robust.

Figure 4.10.1.6 Combined guilt ratings for Speaker 1 - Wales

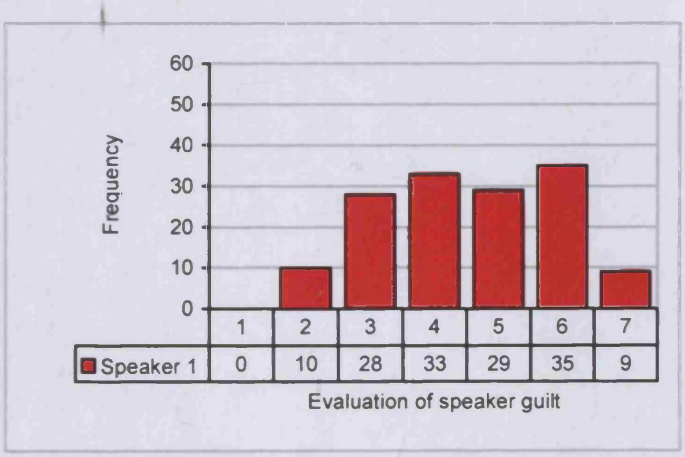


Figure 4.10.1.7 Combined guilt ratings for Speaker 2 - South West

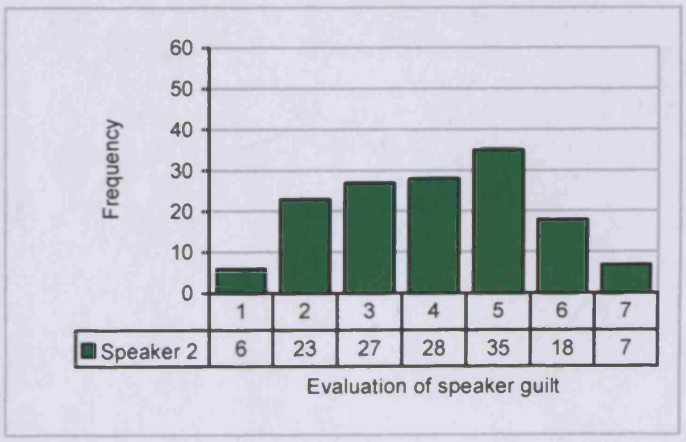


Figure 4.10.1.8 Combined guilt ratings for Speaker 3 - South East

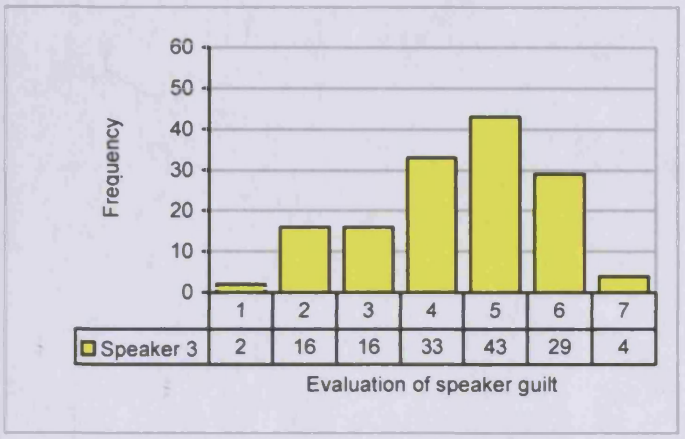


Figure 4.10.1.9 Combined guilt ratings for Speaker 4 – Eastbourne - RP-like

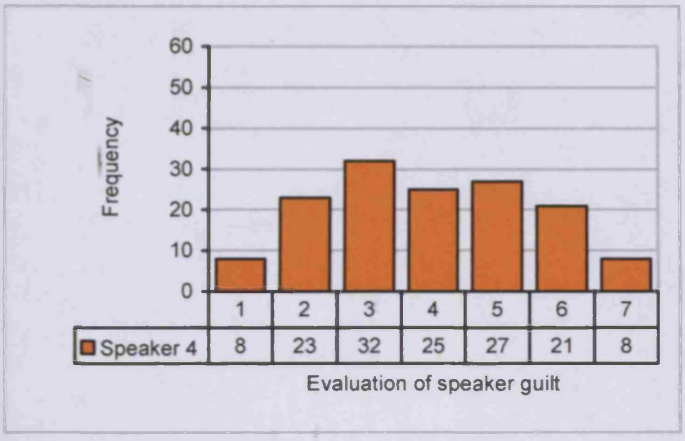


Figure 4.10.1.10 Combined guilt ratings for Speaker 5 - Birmingham

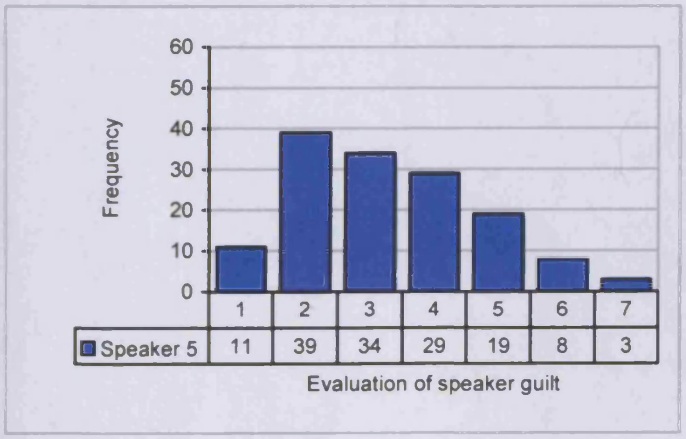
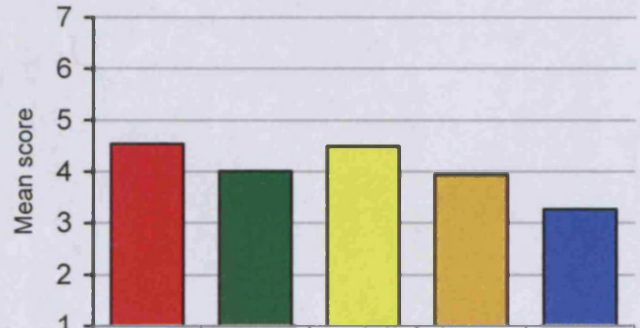


Figure 4.10.1.11 Speaker guilt: comparison of means for each speaker in combined crimes



	Wales	South West	South East	South East (RP-	Midlands
Speaker guilt	4.54	4.01	4.49	3.94	3.27

Figure 4.10.1.12 Repeated measures ANOVA for speaker guilt ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 Guilt	0.002
Speaker 1 - Speaker 3 Guilt	0.326
Speaker 1 - Speaker 4 Guilt	0.001
Speaker 1 - Speaker 5 Guilt	0.000
Speaker 2 - Speaker 3 Guilt	0.024
Speaker 2 - Speaker 4 Guilt	0.732
Speaker 2 - Speaker 5 Guilt	0.000
Speaker 3 - Speaker 4 Guilt	0.031
Speaker 3 - Speaker 5 Guilt	0.000
Speaker 4 - Speaker 5 Guilt	0.000

Adjustment for multiple comparisons: Least significant difference

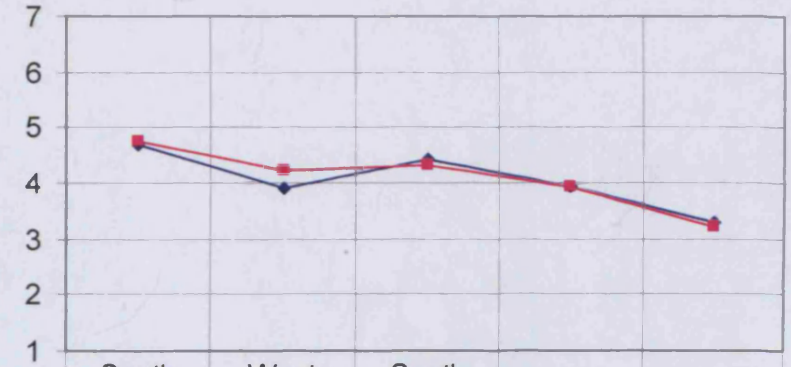
The histograms give us not only the means but also a view of the distributions of the attributions of guilt. There are three speakers (1, 2 and 4) for whom there are wide distributions of evaluations across the scale, indicating some discord in the respondents' feelings towards the speakers in terms of them being potential criminals. The evaluations for Speaker 3 however, have a narrower distribution and a lot more respondent consensus with nearly 70% placing the speaker between points 4 and 6 on the scale. This may indicate some common feeling amongst the respondents that there is belief that the speaker may be guilty. Speaker 5, yet again receives the most positive evaluation with a lower mean than the other speakers and a fairly narrow distribution.

We see that overall, respondents did not make a significant difference between Speakers 1 and 3. Nor did they significantly differentiate between Speakers 2 and 4. As with other sections, we see a 'clustering' of significant differentiations, with Speakers 1 and 3 being rated differently from Speaker 5 for likelihood of guilt, who was differentiated from Speakers 2 and 4. We see that the lowest evaluation of likelihood of committing the meta-crime of 'violent personal attack' is enjoyed by Speaker 5 from the Midlands.

4.10.2 Question order

In the administration of the research, two questionnaires were used, involving identical questions, but in two different sequences - questions on speaker crime first or questions on speaker region first. 66% of respondents were asked to evaluate the speakers' guilt at the end of the questionnaire, having completed all of the tasks. 33% were asked to assess the speakers' guilt at the beginning of the questionnaire after hearing the speakers for the first time and before being asked about such points as the origin of the speakers and the 'intelligence', 'honesty', 'aggressiveness' of the speakers. The mean evaluations of the speakers according to question sequence are presented in 4.10.2.1 (the evaluations from the two different crime types are treated as one group).

Figure 4.10.2.1 Attributions of guilt: crime first versus region first



	South Wales	West Country	South East	RP	Midlands
Crime first	4.67	3.89	4.41	3.94	3.29
Region first	4.74	4.22	4.33	3.94	3.22

n=144

There is a slight difference in the means for Speakers 2 and 3 between the crime first-region first responses. The question arises whether this variation had any significant effect on the responses, for which a similar test, using a repeated measures ANOVA, can be conducted on the responses given by both groups.

Figure 4.10.2.2 Attributions of guilt - crime or region 1st: Tests of within-subjects effects

Source		Sig.
Crime first * Region first	Greenhouse-Geisser	.445

The output tells us that there was no significant difference between the evaluations given by either group, and so the question order had no significant effect. The two groups may be treated as one.

4.11 Principal component analysis and correlations

Having detailed in the previous sections how respondents had rated the speakers' intelligence, honesty, aggression and guilt, and how certain speakers were seen as distinct from each other across these dimensions, it follows that that we should also examine how these dimensions relate to each other for each speaker. This section relates to the research questions in section 2.14.2, specifically, to what extent do the evaluative dimensions interact and what this can tell us about non-linguist language ideology.

When considering the variation present, we need to ask to whether the four variables combine to account for a majority of the variation. If so, in what direction did they vary? For example, did each speaker attract consistently high, medium or low ratings, or were some features more prominent than others? And to what extent do the variables for each speaker vary together, either positively or negatively? Zahn and Hopper (1985), when designing their Speech Evaluation Instrument, recommended that evaluative dimensions be subject to analysis, and accordingly, principal component analyses were performed on the four variables, speaker intelligence, honesty, aggression and guilt.

4.11.1 Principal component analysis & correlations: Speaker 1

The table below shows the factorial loading for the four variables. The first points to notice are that one of the factors has an Eigenvalue greater than 1, that it accounts for over 50% of the variance and that the first three factors account for more than 80% of the variance accepted levels of significance.

Table 4.11.1.1 Principal component analysis for Speaker 1

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative
1	2.081	52.037	52.037
2	0.778	19.448	71.485
3	0.641	16.013	87.498
4	0.500	12.502	100

Notes: The Eigenvalue is above 1. It accounts for more than 50% of variance

Given this, we investigate the components of the first factorial loading in a component matrix:

Table 4.11.1.2 Component matrix for Speaker 1

Component Matrix(a)	Component 1
Speaker 1 Honesty	-0.802
Speaker 1 Guilt	0.755
Speaker 1 Aggression	0.712
Speaker 1 Intelligence	-0.601

Extraction Method: Principal Component Analysis
a - 1 components extracted.

Notes: All are part of one factor, scoring above 0.5

As we can see, all four factorial loadings for each of the four variables score above 0.5, meaning that all can be considered significant and that their contributions to the factor are roughly equal. As only one factor is present, no rotation is necessary.

What we can see from these data is that both 'honesty' and 'intelligence' have an inverse relationship with 'guilt' and 'aggression'. This is in line with what our common sense expectations might tell us: We should expect that 'guilt' and 'honesty' would have an inverse relationship, as our semantic interpretation of

the factor should lead us to expect the more guilty speakers to be seen as less honest and vice versa. Of note here is that 'aggression' takes the same direction as 'guilt', and again, given the nature of the criminal context – committing assault and/or robbery/sexual assault – we might expect attributions of higher guilt to be in-line with attributions of 'aggression'. Also worth noting is the positive relationship between 'honesty' and 'intelligence' and the inverse relationship between 'intelligence' and 'honesty'/'guilt'. To investigate the co-variation between these factorial variables, a Pearson *r* correlation was performed. The following tables throughout the rest of the chapter use the term 'substantial' relationship and 'definite' relationship in quantifying the strength of correlations taken from (Guilford 1956) as a rough guide to interpreting correlatory relationships.

Figure 4.11.1.1 Pearson r correlation coefficient (2 tailed) for Speaker 1

'Substantial' relationship (Guilford 1956)

Guilt – Honesty	r = -.485,	p<0.01	
Honesty – Aggression	r = -.432,	p<0.01	

'Definite' relationship (Guilford 1956)

Guilt – Aggression	r = .363,	p<0.01	
Intelligence – Honesty	r = .322,	p<0.01	
Guilt – Intelligence	r = -.282,	p<0.01	
Intelligence – Aggression	r = -.250,	p<0.01	df=143

The correlational analysis appears to replicate the direction of the relationships present in the factorial loadings. However, from the forensic point of view, it is the relationship between the four factorial loadings which may be of more interest. In earlier sections, we saw how the ANOVA established independence of certain variables from others, for example, how the scores for Speaker 1 and 3 'guilt' were seen differently from those for Speakers 2, 4 and 5 and how Speaker 2 and 4 'guilt' were seen differently from Speaker 1, 3 and 5. What is particularly prominent here is the extent to which 'honesty' plays a role in these relationships. What we can observe above, is the way in which increased attributions of honesty have a positive relationship with 'unaggressiveness', 'intelligence' and as we might expect, 'guilt'. That is to say, the correlational relationships above indicate that for Speaker 1, positive attributions of 'honesty'

and 'innocence' appear to be related to impressions of the speaker being 'unaggressive' and 'intelligent'. And where the speaker is viewed as more 'unintelligent' or 'aggressive', significantly inverted assessments for honesty and guilt co-occur.

4.11.2 Principal component analysis & correlations: Speaker 2

For Speaker 2, similar patterns are repeated as for Speaker 1: there is one factorial loading with an Eigenvalue above 1, accounting for nearly 50% of the variance:

Table 4.11.2.1 Principal component analysis for Speaker 2

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative
1	1.788	44.703	44.703
2	0.976	24.397	69.1
3	0.802	20.060	89.16
4	0.434	10.840	100

Extraction Method: Principal Component Analysis.

The one factor accounts for 44% of the variance. However, of the factorial variables therein, only 'honesty', 'guilt' and 'aggression' have a loading of greater than 0.5, suggesting that 'intelligence' may not have been a significant part of the single significantly loaded factor (again, as only one factor emerged, no rotation was necessary):

Table 4.11.2.2 Component matrix for Speaker 2

Component Matrix(a)	Component 1
Speaker 2 Honesty	0.855
Speaker 2 Guilt	-0.670
Speaker 2 Aggression	-0.615
Speaker 2 Intelligence	0.480

Extraction Method: Principal Component Analysis.
a - 1 components extracted.

There are three components to the factor. Intelligence does not appear to play a major part. The relationships between the variables can be further investigated through Pearson r rotation:

Figure 4.11.2.1 Pearson r correlation coefficient (2 tailed) for Speaker 2

'Substantial' relationship (Guilford 1956)

Guilt – Honesty	r = -.460	p<0.01

'Definite' relationship (Guilford 1956)

Honesty – Aggression	r = -.339,	p<0.01	
Intelligence – Honesty	r = .327,	p<0.01	df=143

The correlation data suggest that 'aggression' is positively correlated to 'honesty' determination of 'honesty', and there is also a positive correlation between intelligence and honesty.

In relation to the ratings for Speaker 1, we can see that both 'aggression' and 'intelligence' share a slightly less significant relationship with honesty. Yet of note is the direction of the significant relationships between the variables. For example, increased perceptions of 'aggression' correlate with decreased perceptions of 'honesty'. Decreased perceptions of 'honesty' correlate with increased perceptions of guilt. The direction of these relationships is the same as for Speaker 1.

4.11.3 Principal component analysis & correlations: Speaker 3

For Speaker 3, the principal component analysis indicates one factor with an Eigenvalue of more than 1 (again, as with the analysis for the previous two speakers, only one factor emerged, indicating that no rotation was required):

Table 4.11.3.1 Principal component analysis for Speaker 3

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative
1	1.781	44.514	44.514
2	0.907	22.671	67.185
3	0.746	18.651	85.863
4	0.567	14.164	100

Extraction Method: Principal Component Analysis

The loadings for this factor indicate the relevance of all four variables, each having a score of over 0.5. Of note is that once more we see the emergence of the same direction of relationship between the four loadings:

Table 4.11.3.2 Component matrix for Speaker 3

Component Matrix(a)	Component 1
Speaker 3 Honesty	-0.775
Speaker 3 Aggression	0.694
Speaker 3 Guilt	0.657
Speaker 3 Intelligence	-0.515

Extraction Method: Principal Component Analysis.
a - 1 components extracted.

Performing a Pearson *r* correlation, we can see that the strongest relationship that emerges is that between 'honesty' and 'aggression' and a weaker relationship emerges between 'intelligence' and the other variables:

Figure 4.11.3.1 Pearson r correlation coefficient (2 tailed) for Speaker 3

'Substantial' relationship (Guilford 1956)

Aggression – Honesty	r = -.416,	p<0.01	

'Definite' relationship (Guilford 1956)

Guilt – Honesty	r = -.305,	p<0.01	
Guilt – Aggression	r = -.238,	p<0.01	
Intelligence – Honesty	r = .225,	p<0.01	
Guilt – Intelligence	r = -.217,	p<0.01	df=143

This would suggest that although variables such as 'intelligence' may be associated with the 'cockney' voice in traditional language attitudes research, in this case, where attributions of 'guilt' and 'honesty' have been made, one of the strongest relationships shared by 'honesty' and 'guilt' is with 'aggression', rather than 'intelligence'.

Is this association one that is generally associated with the 'estuary' speaker? Or is this association related to the specific 'performance' of the speaker? These questions will be explored in the Discussion chapter.

4.11.4 Principal component analysis & correlations: Speaker 4

A principal component analysis of the evaluations for Speaker 4 reveals once again one significantly loaded factor, with an Eigenvalue above 1, accounting for almost 50% of the variance:

Table 4.11.4.1 Principal component analysis for Speaker 4

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative
1	1.966	49.141	49.141
2	0.812	20.308	69.449
3	0.672	16.811	86.26
4	0.550	13.740	100

Extraction Method: Principal Component Analysis.

This factor comprises all four variables, each with a loading greater than 0.5. Again, we see the same direction in the relationships between the four loadings:

Table 4.11.4.2 Component matrix analysis for Speaker 4

Component Matrix(a)	Component 1
Speaker 4 Honesty	0.790
Speaker 4 Aggression	-0.720
Speaker 4 Guilt	-0.663
Speaker 4 Intelligence	0.619

Extraction Method: Principal Component Analysis.
a - 1 component extracted.

Performing a Pearson *r* correlation, we can see that the strongest relationship that emerges is that between 'honesty' and 'aggression'.

Figure 4.11.4.1 Pearson r correlation coefficient (2 tailed) for Speaker 4

'Substantial' relationship (Guilford 1956)

Aggression – Honesty	r = -.410,	p<0.01

'Definite' relationship (Guilford 1956)

Guilt – Honesty	r = -.383,	p<0.01	
Intelligence – Honesty	r = .347,	p<0.01	
Guilt – Aggression	r = .305,	p<0.01	
Aggression – Intelligence	r = -.274,	p<0.01	df=143

We can also see a slightly higher correlation between 'honesty' and 'intelligence' in the variables for Speaker 4, higher than any of those for the Speakers 1-3. Nevertheless, it is the perception of 'aggression' that correlates most strongly with 'honesty'. This will need examining in the context of previous language attitude research. Have RP-related accents regularly been characterised by an association with 'aggression'? Or is this association located in the context of the speaker's 'performance'?

4.11.5 Principal component analysis & correlations: Speaker 5

The principal component analysis for Speaker 5 reveals 1 factor with an Eigenvalue above 1, accounting for 45% of the variance:

Table 4.11.5.1 Principal component analysis for Speaker 5

Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative
1	1.815	45.387	45.387
2	0.913	22.817	68.204
3	0.838	20.961	89.165
4	0.433	10.835	100

Extraction Method: Principal Component Analysis.

This factor has three loadings above 0.5, and we see as with Speaker 2, the loading for 'intelligence' falls just below 0.5, suggesting that intelligence was not a significant factor in the relationship between the four variables:

Table 4.11.5.2 Component matrix for Speaker 5

Component Matrix(a)	Component 1
Speaker 5 Honesty	0.861
Speaker 5 Guilt	-0.647
Speaker 5 Aggression	-0.642
Speaker 5 Intelligence	0.494

Extraction Method: Principal Component Analysis.
a - 1 component extracted.

Performing a Pearson *r* correlation, we can see that honesty is most strongly related to the factor, with the strongest relationships that emerge being those between 'honesty' and 'aggression' and 'honesty' and 'guilt':

Figure 4.11.5.1 Pearson r correlation coefficient (2 tailed) for Speaker 5

'Substantial' relationship (Guilford 1956)

Aggression – Honesty	r = -.440,	p<0.01
Guilt – Honesty	r = -.423,	p<0.01

'Definite' relationship (Guilford 1956)

Intelligence – Honesty	r = .296,	p<0.01	df=143

It appears that the attributions of 'guilt' and 'honesty' are somewhat clearer cut for Speaker 5 than with other speakers, being associated most significantly with 'aggression'. The data suggests a weaker relationship between 'honesty', 'guilt' and 'intelligence', and that for those individuals who did give higher attributions of 'honesty' and 'guilt' there is an association with 'aggression'.

4.12 Summary

The results from Stage 1 have addressed the research questions in 2.14.2 and indicated the following:

1/. There is variation in the recognition rates for each speaker accent and this variation is mediated by the regionality of the speaker's accent, and the social connotations of the speaker's accent and the respondents' region of origin.

2/. Respondents were broadly accurate in their estimations of speaker age, but factors such as social stereotypes and the use of age categories may strengthen or weaken the accuracy of these estimations.

3/. Evaluations of speaker accent in this research follow similar patterns to those in previous studies. There is a complex relationship between estimations of speaker origin, general evaluations of the speaker and attributions of guilt.

4/. Non-linguists are able to provide rich and detailed descriptions of speaker voice characteristics which frequently align with those categories used by professional linguist for the voice description tasks. The detail of these findings and their significance will be discussed in Chapter 6.

5 - RESULTS: Stage 2

5.1 Introduction

Chapter 5 concerns itself with the results of Stage 2. Administered to a second cohort of respondents, Stage 2 served a number of follow up and innovative roles. Before detailing the results for Stage 2, below is a summary of the main reasons for conducting a second Stage of the research:

i/. It allowed for a repeat of the regional estimation experiment, but with an enhanced elicitation method, namely the use of guided perceptual mapping.

ii/. Based on some of the ambiguities elicited from the first Stage regarding the provenance of Speakers 4 and 5 (London/South/South East), Stage 2 saw the introduction to the composite 'strength of accent' scale on which respondents could mark their perceptions regarding a speaker's 'regional/no accent/well spoken' status.

iii/. Having analysed the qualitative voice descriptions from the respondents in Stage 1 and reduced the data to audiofit themes of bi-polar and non-polar descriptions, Stage 2 was an opportunity to administer the audiofit to a new group in order to observe how and whether non-linguists would use them for describing the speakers.

iv/. Having heard informal comments made by respondents during Stage 1, referring to who the speakers sounded like, Stage 2 represented an opportunity to elicit these opinions in a more formalised way, by presenting it as one of the questions in the research instrument.

v/. Having elicited attributions of guilt for the two suggested crimes in Stage 1, Stage 2 presented an opportunity to change the crime, presenting respondents with a control, a non-criminal scenario, based on a romantic liaison.

Chapter 5 begins with the data from respondent estimations of speaker provenance (Section 5.2), followed by tests of association between respondent and speaker origin (5.3) and data indicating which accents were most salient for each region (5.4). Sections 5.5- 5.8 investigate aspects of the audiofit in a way that departs from the methodology used in Stage 1. Section 5.5 presents data on evaluations of strength of accent, 5.6 presents evaluations of voice features via bipolar semantic scales and 5.7 presents non-polar voice features that did not lend themselves to scale measurement. Section 5.8 presents results of asking respondents who they think the speaker might sound like. Section 5.9 returns to the issue of estimations of guilt which is contrasted in 5.10 with estimations of success in a romantic liaison. The chapter finishes at 5.11 with a short summary.

5.2 Where did respondents believe speakers are from?

In this section, the question addressed is where the respondents believe the speakers are from. The respondents were asked first to give details of their own origin, in either town/city/county/region/country form, an activity which served as a 'warm-up exercise', encouraging cognitive elaboration, for the task of estimating speakers' of origin, which follows. Respondents were then presented with the five speaker's voices and asked to estimate their origin.

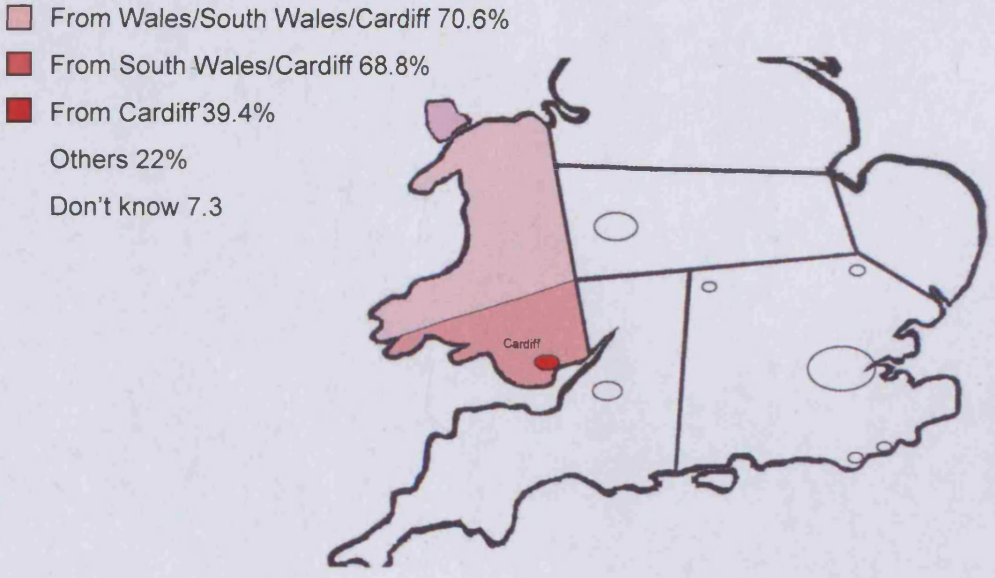
The major difference between this exercise in Study 2 and the estimation of speaker provenance in Study 1 is that respondents in Study 2 were also provided with the idealised perceptual map of the UK and list of regions and cities, as detailed in Figures 3.2.5.6 and 3.2.5.7. These they used first to plot their own origin, activating their areal taxonomies of the UK, and then, having familiarised themselves with this conceptual map, they used the map to plot the

speakers' origins. Respondents used either the map, the list or both as they saw fit.

Figures 5.2.1 – 5.2.5 are breakdowns, voice by voice, of where the respondents believed the speakers to hail from. For statistical efficiency, all smaller towns suggested were recoded as belonging to the relevant larger region in which they are located, aside from major urban conurbations within each region (Cardiff, Bristol, London, Birmingham). For each speaker, there is a tabular breakdown, and a graphic, a chi-square test and a spatial representation.
In order to explore how both specific and accurate respondents' correct estimations for speaker origin were, they are represented with coloured shading the intensity of colour/darker shade representing the greater specifity with which the respondents identified the Speaker. The circles represent the principal cities in the relevant region.

Speaker 1: From Wales > South Wales > Cardiff

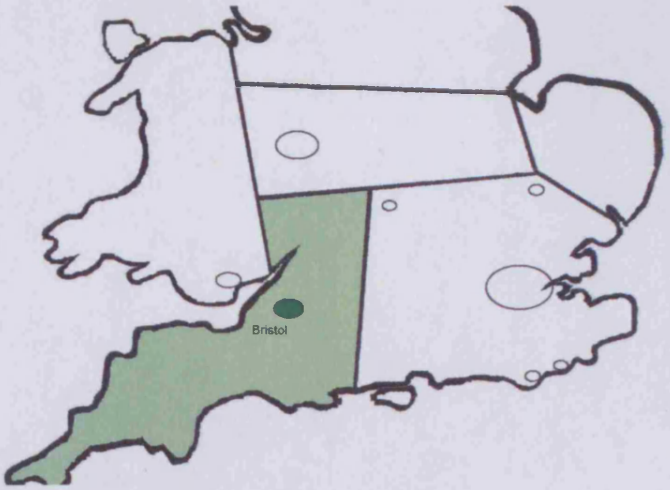
Figure 5.2.1 Areal representation of respondents' identifications of Speaker 1



Speaker 2: From South West > Bristol

Figure 5.2.2 Areal representation of respondents' identifications of Speaker 2

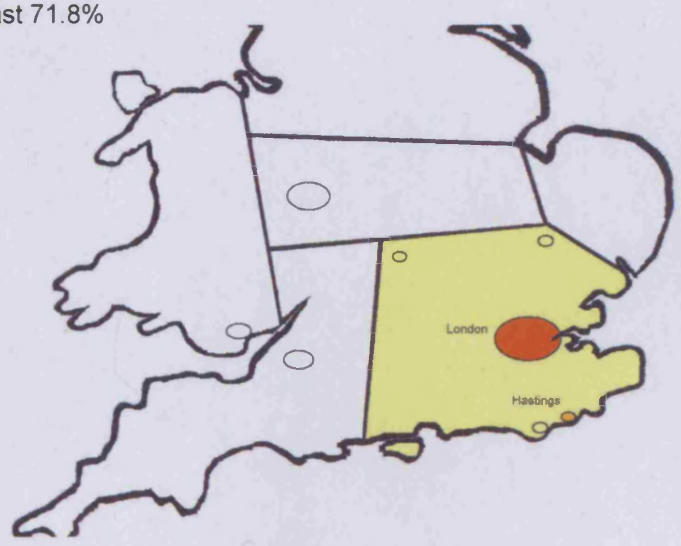
From South West England 70.4%
From Bristol 42.6%
Others 20.4%
Don't know 9.3%



Speaker 3: From the South East > Hastings

Figure 5.2.3 Areal representation of respondents' identifications of Speaker 3

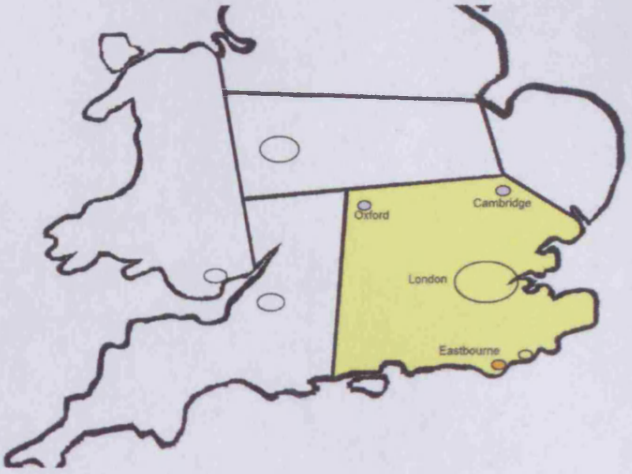
From London/South East 71.8%
From London 51.8%
From Hastings 0%
Others 20%



Speaker 4: From the South East > Eastbourne (RP-like)

Figure 5.2.4 Areal representation of respondents' identifications of Speaker 4

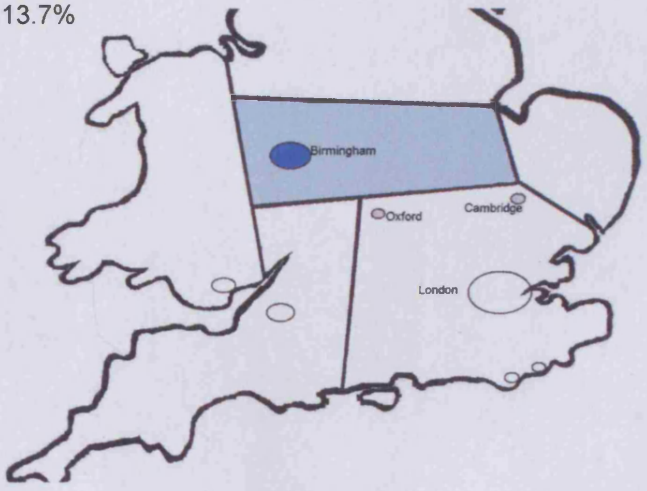
From Oxford/Cambridge 39.5%
From London/South East /South 20.2%
From Eastbourne 0%
Others12.5%
Don't know 24%



Speaker 5: From the Midlands > Birmingham

Figure 5.2.5 Areal representation of respondents' identifications of Speaker 5

From Midlands 17.6%
From Birmingham 1%
From Oxford/Cambridge 23.5%
London/South/South East 13.7%
Others 23.5%
Don't know 20.6%



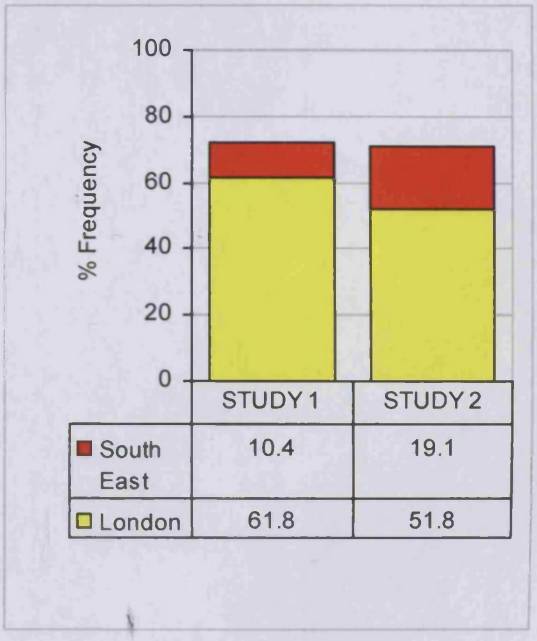
Observing the data in the figures above, we see that results for Speakers 1 and 2 are very similar not only to each other, but also to those observed in Stage 1.

71% of respondents recognised Speaker 1's accent as being from Wales (Figure 5.2.1) - a small rise on the figure for Stage 1 (66%), and 70% (Figure 5.2.2) estimated Speaker 2's origin as being the West Country, up from 67% in Stage 1. The number of correct identifications of the speakers' cities of origin also increased between Stage 1 to Stage 2, from 35% to 39% for Speaker 1 being from Cardiff, and from 20% to 43% to for Speaker 2 being from Bristol. There is also a decrease in the number of 'don't know' responses for both speakers, down from 13% to 7% for Speaker 1 and from 19% to 9% for Speaker 2. There was a small increase in the number of incorrect answers, up from 20% to 22% for Speaker 1 and from 14% to 20% for Speaker 2. For all of these differences there are two possible explanations. Firstly, the difference from Stage 1 to Stage 2 could simply be down to the use of a new set of respondents. Alternatively, the manner of elicitation – the use of guided perceptual mapping could be responsible. Given that all the trends have followed the same directions for Speaker 1 and 2, showing increases and decreases on the same variables, it appears that there is a strong likelihood that the questioning technique has had some effect on the responses elicited. If this is the case, it appears that the effect of guided perceptual mapping is to increase the likelihood of respondents making a selection of speaker origin, rather than saying 'don't know', increasing the number of correct, but also incorrect responses.

Responses to Speaker 3 demonstrate a similar majority belief to that shown for Speakers 1 and 2, with 70% of respondents believing the speaker to be from the South East. Unlike for Speakers 1 and 2, the 70% does not represent an increase in frequency compared with Study 1 which was also around 70%. However, what is different is the composition of the 70%. There is still a large amount of technical inaccuracy, with the majority of respondents believing the speaker to be from London itself; but this figure is noticeably lower in Study 2, compared with Study 1, 51.8% compared 61.8%. Instead, there has been an increase in the number of respondents believing the speaker to be from the wider South East rather than London, from 10% to 19%. This may be related to the skills and knowledge of the cohort of respondents, however, we must not exclude the possibility that the perceptual map, offering a South East region

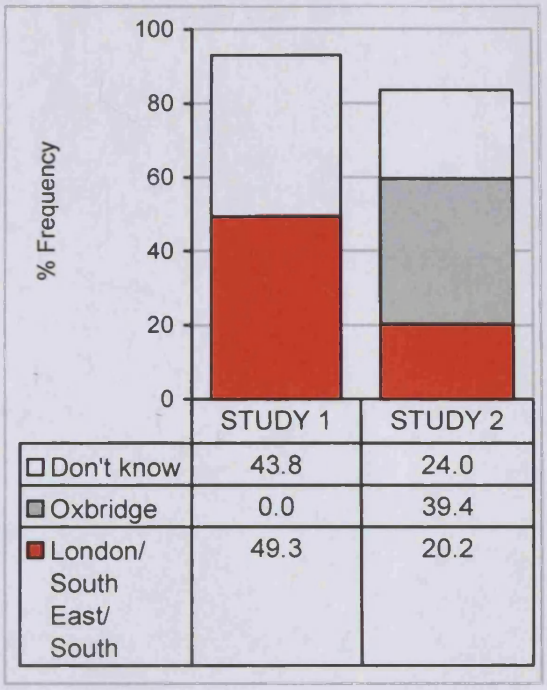
alongside London, may be offering a more salient alternative to a sizeable minority of respondents.

Figure 5.2.6 Speaker 3 - % Identifications of London/South East, Study 1 vs. Study 2



It is with Speaker 4 that we see the greatest difference between the data of Stage 1 and Stage 2. We saw in the data for Stage 1 that identifying the regional origin of the RP-accented Speaker 4 presented a problem for many respondents. To address this difficulty, it was decided to include on the perceptual map locations that non-linguists may associate with RP speakers. There is a widely-held non-linguist misconception (following such common phrases as 'Oxford English') that 'proper' or 'correct' English pronunciation exists, and that it is derived from those varieties spoken in the most famous university cities in England, Oxford and Cambridge (commonly referred to in the United Kingdom by the portmanteau 'Oxbridge'). It was speculated in the research design for Stage 2 that by offering respondents the Oxford and Cambridge locations, respondents may be able to articulate any sense of RP-ness they heard, utilising status-loaded labels to answer a geographical question. A second speculation was that any use of the Oxbridge cities would only be to articulate RP-ness, rather than to describe the urban accents of the respective cities.

Figure 5.2.7 Speaker 4 - % Identifications for Oxbridge, Study 1 vs. Study2



	STUDY 1	STUDY 2
Don't know	43.8	24.0
Oxbridge	0.0	39.4
London/ South East/ South	49.3	20.2

The results show some remarkable outcomes. Firstly, Speaker 3, whose accent is closer to the estuary English accent of Cambridgeshire and Speaker 2, whose accent is closer to the rhotic accent of rural Oxfordshire, were at no point associated by any respondent with either Oxford or Cambridge. This would seem to suggest that for the respondents in Stage 2, the only clear salient variety connoted by the regional labels 'Oxford' and 'Cambridge' is an RP-like one. Moreover, the data show an extremely high uptake of the use of Oxbridge as speaker locations. Where in Stage 1, respondents had apparently articulated their impressions of Speaker 4's RP accent by describing him as from the South, South East or London, in Stage 2, nearly 40% of respondents articulated their sense of RP-ness by utilising the Oxbridge labels which they chose *in preference* to the other marked cities and regions. Also, we can see in Figure 5.2.7 that with the provision of the labels, 'Oxford' and 'Cambridge', some 70% of respondents in Stage 2 were able mobilise a geographical label to articulate their perceptions compared with the data from Stage 1, in which only 49.3% of respondents suggested a label that correctly related to the speaker's origin. Furthermore, the number of 'don't knows' has nearly halved from 43.8% in Stage 1 to 24% in Stage 2. The data from Stage 2 using a map with Oxford and

Cambridge marked on them suggests that there is a way of assisting non-linguists to represent essentially social judgement in a geographical format.

Finally, we can also observe that not one respondent suggested that the speaker was from Eastbourne, his home town. Although it is acknowledged, of course, that the estimation of 'Eastbourne', as with Speaker 3 and 'Hastings', would be a more specific estimation than 'Cardiff', 'Bristol' or 'London'.

With Speaker 5 there is a slight increase in the identification of the speaker being from the Midlands, up from 13.9% in Stage 1 to 18.6% in Stage 2. But most noticeably, a quarter of all respondents took advantage of the provision of 'Oxford/Cambridge' as a label. This suggests that as with Stage 1, there was a definite minority of respondents that believed Speaker 5 to be 'non-regional'. Furthermore, less than half the number of respondents as Stage 1 indicated in Stage 2 that they did not know where the speaker was from, down from 44.4% to 20.6%. Assuming the label of 'Oxbridge' to be in some way meaningful to the respondents, the number of general incorrect attributions also fell from 41.7% in Stage 1, to 34.3% in Stage 2.

5.3 Estimations of speaker origin: are they associated with respondent origin?

As with Stage 1, regional estimations of provenance were also collated according to the respective origins of the respondents – i.e. organised in terms of the region of origin. Important to note here is that there can be no ambiguity as to the respondents' origins as they were provided with the regional categorical lists and a conceptual map on which they could mark their specific region. However, throughout this Stage it has been important to remember that there are no strict isoglosses in the United Kingdom, demarcating a sudden switch from one regional speech variety to another, and that it is general tendencies that we are seeking to observe. Furthermore, in Stage 2 as in Stage 1, the number of respondents, especially from the Midlands, was lower than that

from the other three regions, necessitating a loose interpretation of the data from those speakers when considered in comparison with other regions.

In the previous section, all five speakers' voices were examined according to the 112 respondents' estimations of speaker origin. As with the same exercise in Study 1 (Chapter 4), we saw differences in the patterns of responses that suggest that the respondents generally saw the five voices as from geographically or socially distinct backgrounds – Wales, South West, South East (London), South East (RP) and broadly Southern English. In this section, the data are interrogated to explore the relationship between respondents' estimations of speaker origin and the respondents' own region of origin and whether respondent origin is a factor in the identification of the five voices presented. The coding for respondent origin is the same as for the equivalent exercise in Chapter 4, Wales, the South West, the South East or the Midlands.

As this question is related only to respondents from the four regions within the catchment area, the few respondents who did not provide any information on their regional origin were excluded, leaving 109 respondents. Once again, there was a relatively low number of respondents from the Midlands (13) and, although they are included, data from Midlands' respondents are not treated as statistically robust and these figures should be treated with caution compared to the data from the better represented regions.

The question directed at these data is related to the research question in Section 2.14: 'What factors may influence respondents' estimations'. Respondents' positive or negative estimations of speaker origin were grouped by region in tables of frequency. In order to interrogate this categorical data, a chi-square test was performed for each speaker and regional estimation, investigating any significant associations. The regional estimations tested were:

Speaker 1 - estimated to be from 'Wales'
Speaker 1 - estimated to be from 'South Wales'
Speaker 1 - estimated to be from 'Cardiff'
Speaker 2 - estimated to be from 'South West'
Speaker 2 - estimated to be from 'Bristol'
Speaker 3 - estimated to be from 'South East'
Speaker 3 - estimated to be from 'London'
Speaker 4 - estimated to be from 'South East/London'
Speaker 4 - estimated to be from 'London'
Speaker 4 - estimated to be from 'Oxbridge'
Speaker 5 - estimated to be from 'Midlands'
Speaker 5 - estimated to be from 'South/South East'
Speaker 5 - estimated to be from 'Oxbridge'

The tables of frequencies containing percentages figures for positive identifications are presented below. The frequencies given in the tables should be taken as illustrative of possible differences in positive identifications according to respondent region. A commentary on the data, the associations and frequencies follows the tables. In each case, the region that provided the highest percentage frequency of positive identification is listed first, the descending order of regions indicating decreased percentage frequency. The results of the chi-square tests are presented below each table.

Table 5.3.1 Speaker 1, identified as from 'Wales' by respondents from:

	Respondents guessed 'Wales'		**Total**	**% Correct**
Respondent Region	*no*	*yes*		
Wales	7	30	37	81.1
Midlands	3	11	14	78.6
West/South West	15	23	38	60.5
South East	9	11	20	55.0
Total	34	75	109	

Respondent/speaker origin1 - $\chi^2 = 6.209$, df = 3, $\rho = 0.102$

Table 5.3.2 Speaker 1, identified as from 'South Wales' by respondents from:

	Respondents guessed 'South Wales'		Total	% Correct
Respondent Region	*no*	*yes*		
Wales	7	30	37	81.1
Midlands	4	10	14	71.4
West/South West	16	22	38	57.9
South East	9	11	20	55.0
Total	36	73	109	

Respondent/speaker origin1 - $\chi^2 = 6.168$, df = 3, $\rho = 0.104$

Table 5.3.3 Speaker 1, identified as from 'Cardiff' by respondents from:

	Respondents guessed 'Cardiff'		Total	% Correct
Respondent Region	*no*	*yes*		
Wales	21	16	37	43.2
West/South West	23	15	38	39.5
Midlands	9	5	14	35.7
South East	14	6	20	30.0
Total	67	42	109	

Respondent/speaker origin1 - $\chi^2 = 6.209$, df = 3, $\rho = 0.102$

Table 5.3.4 Speaker 2, identified as 'South West' by respondents from:

	Respondents guessed 'South West'		Total	% Correct
Respondent Region	*no*	*yes*		
West/South West	6	32	38	84.2
Midlands	4	10	14	71.4
Wales	14	23	37	62.2
South East	8	12	20	60.0
Total	32	77	109	

Respondent/speaker origin - $\chi^2 = 5.753$, df = 3, $\rho = 0.124$

Table 5.3.5 Speaker 2, identified as from 'Bristol' by respondents from:

	Respondents guessed 'Bristol'		Total	% Correct
Respondent Region	*no*	*yes*		
West/South West	17	21	38	55.3
South East	11	9	20	45.0
Wales	25	12	37	32.4
Midlands	10	4	14	28.6
Total	63	46	109	

Respondent/speaker origin1 - $\chi^2 = 5.236$, df = 3, $\rho = 0.155$

Table 5.3.6 Speaker 3, identified as from 'South East/London by respondents from:

	Respondents guessed 'South East/London'		Total	% Correct
Respondent Region	*no*	*yes*		
Midlands	1	13	14	92.9
South East	3	17	20	85.0
West/South West	11	27	38	71.1
Wales	17	20	37	54.1
Total	32	77	109	

Respondent/speaker origin1 - $\chi^2 = 10.231$, df = 3, $\rho = 0.016$

Table 5.3.7 Speaker 3, identified as from 'London' by respondents from:

	Respondents guessed 'London'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	6	14	20	70.0
Wales	15	22	37	59.5
West/South West	18	20	38	52.6
Midlands	7	7	14	50.0
Total	46	63	109	

Respondent/speaker origin1 - $\chi^2 = 4.550$, df = 3, $\rho = 0.208$

Table 5.3.8 Speaker 4, identified as from 'South East/London' by respondents from:

	Respondents guessed 'South East/London'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	15	5	20	25.0
Midlands	12	2	14	14.3
West/South West	34	4	38	10.5
Wales	37	0	37	0.0
Total	98	11	109	

Respondent/speaker origin1 - $\chi^2 = 17.513$, df = 3, $\rho = 0.001$

Table 5.3.9 Speaker 4, identified as from 'South/S.East/London by respondents from:

	Respondents guessed 'South/S. East/London'		Total	% Correct
Respondent Region	*no*	*yes*		
South East	14	6	20	30.0
West/South West	30	8	38	21.1
Midlands	12	2	14	14.3
Wales	32	5	37	13.5
Total	88	21	109	

Respondent/speaker origin1 - $\chi^2 = 2.570$, df = 3, $\rho = 0.462$

Table 5.3.10 Speaker 4, identified as from 'Oxbridge' by respondents from:

	Respondents guessed 'Oxbridge'		Total	% Correct
Respondent Region	*no*	*yes*		
Midlands	6	8	14	57.1
South East	11	9	20	45.0
West/South West	25	13	38	34.2
Wales	27	10	37	27.0
Total	69	40	109	

Respondent/speaker origin1 - $\chi^2 = 4.703$, df = 3, $\rho = 0.195$

Table 5.3.11 Speaker 5, identified as from 'Midlands' by respondents from:

	Respondents guessed 'Midlands'		Total	% Correct
Respondent Region	*no*	*yes*		
Respondent Region	*no*	*yes*		
Wales	29	8	37	21.6
West/South West	32	6	38	15.8
South East	17	3	20	15.0
Midlands	13	1	14	7.1
Total	91	18	109	

Respondent/speaker origin1 - $\chi^2 = 0.164$, df = 3, $\rho = 0.650$

Table 5.3.12 Speaker 5, identified as from 'South/South East' by respondents from:

	Respondents guessed 'South/South East		Total	% Correct
Respondent Region	*no*	*yes*		
Wales	33	4	37	10.8
West/South West	34	4	38	10.5
South East	18	2	20	10.0
Midlands	13	1	14	7.1
Total	98	11	109	

Respondent/speaker origin1 - $\chi^2 = 0.163$, df = 3, $\rho = 0.983$

Table 5.3.13 Speaker 5, identified as from 'Oxbridge' by respondents from:

	Respondents guessed 'Oxbridge'		Total	% Correct
Respondent Region	*no*	*yes*		
Wales	28	9	37	24.3
West/South West	32	6	38	15.8
South East	15	5	20	25.0
Midlands	10	4	14	28.6
Total	85	24	109	

Respondent/speaker origin1 - $\chi^2 = 1.427$, df = 3, $\rho = 0.699$

Taking first of all the frequencies in the tables above, we can see some patterns that we might expect: the Welsh speaker is identified as being from Wales, South Wales and Cardiff most frequently by respondents from Wales with respondents from the region furthest from Wales, the South East, returning the lowest frequency. In fact, we see yet again that those respondents with the lowest *mutual* recognition rates are from the South East of England and Wales, suggesting that regardless of direct mileage, it is the accents of these two regions that remain the most psycholinguistically or perceptually distanced from each other. This may be because they sound the same or that conceptually, London and the South East *are* the same, and it may be only those that live nearer the region that make the distinction between the two.

As with Study 1, we can also observe that the speaker from the South West is most frequently identified as being from the South West and Bristol by respondents from the South West. However, whilst all of these frequencies are indicative of potential patterns, we are unable to draw any firm conclusions on these points due the results of the chi-square test, which indicate no significant associations between speaker origin for Speakers 1 and 2 and respondents' origin. In fact, where in Study 1, we saw three significant associations between speaker identification and respondent region, Speaker 1 being from Wales, Speaker 3 from the South East and Speaker 4 from the South/South East/London, in Study 2, we see only two associations, an association for Speaker 1 being absent:

Significant associations found ($\rho<0.05$):
Speaker 3 comes from the South East/London ($\rho<0.016$)
Speaker 4 comes from the South East/London ($\rho<0.001$)

We do, however, see a significant association for Speaker 3, where the two regions that most frequently identified Speaker 3 as being from the South East are the South East and perhaps surprisingly, the Midlands (perhaps due to the central position adjacent to the South East or simply as an artefact of a lower number of respondents representing that region). This is not what we might

have expected. The estimations for Speakers 4 and 5 show no consistent patterns.

We can posit two possible explanations for this outcome. Firstly, the group of respondents in Study 2, despite being of a similar age and educational profile to the respondents in Study 1, and despite being exposed to the same five voices, nevertheless were different in their abilities to identify speaker origins and or a better geo-linguistic knowledge of the variation in the UK. For example, we saw in the previous section, that the positive identification rates for Speakers 1 and 2 were higher than in Study 1. However, the alternative explanation is that the manner of elicitation may have motivated different identifications from the respondents. Respondents in Study 2 were provided with a tabulated list and perceptual maps through which to articulate their perceptions, and this may be a motivator for differing results.

5.4 Which accents were most salient to each region?

Which of the voices heard were most salient to the regional respondents? For each region, which voices could they best recognise and then correctly identify, utilising relevant city/county/region-related labels?

Table 5.4.1 Salience rates for Welsh respondents

81.1%	Speaker 1 (*from Cardiff, Wales*) = Wales
59.5%	Speaker 2 (*from Bristol, South West*) = West Country
54.1%	Speaker 3 (*from Hastings, South East*) = London
27.0%	Speaker 4 (*RP*) = Oxbridge/South/RP
18.4%	Speaker 5 (*from Birmingham, Midlands*) = Oxbridge
13.6%	Speaker 4 (*RP*) = London/South East
5.4%	Speaker 5 (*from Birmingham, Midlands*) = Midlands

Table 5.4.2 Salience rates for South West England respondents

	81.6%	Speaker 2 (*from Bristol, South West*) = West Country
	71.1%	Speaker 3 (*from Hastings, South East*) = London
	60.5%	Speaker 1 (*from Cardiff, Wales*) = Wales
	42.1%	Speaker 4 (*RP*) = Oxbridge/South/RP
	18.4%	Speaker 5 (*from Birmingham, Midlands*) = Midlands
	18.4%	Speaker 5 (*from Birmingham, Midlands*) = Oxbridge/South/RP
	13.2%	Speaker 4 (*RP*)= London/South East

Table 5.4.3 Salience rates for South East England respondents

	85.0%	Speaker 3 (*from Hastings, South East*) = London/South East
	60.0%	Speaker 2 (*from Bristol, South West*) = West Country
	55.0%	Speaker 1 (*from Cardiff, Wales*) = Wales
	45.0%	Speaker 4 (*RP*) = Oxbridge/South/RP
	30.0%	Speaker 4 (*RP*) = London/South East
	25.0%	Speaker 5 (*from Birmingham, Midlands*) = Oxbridge
	15.0%	Speaker 5 (*from Birmingham, Midlands*) = Midlands

Table 5.4.4 Salience rates for Midlands respondents

	92.9%	Speaker 3 (*from Hastings, South East*) = London/South East
	78.6%	Speaker 1 (*from Cardiff, Wales*) = Wales
	71.4%	Speaker 2 (*from Bristol, South West*) = West Country
	57.1%	Speaker 4 (*RP*) = Oxbridge/South/RP
	42.9%	Speaker 5 (*from Birmingham, Midlands*)= Midlands
	28.6%	Speaker 5 =(*from Birmingham, Midlands*)= Oxbridge/South/RP
	14.3%	Speaker 4 (*RP*) = London/South East

For Welsh respondents, the Welsh speaker was most consistently identified as such; the two southern English accents were also well identified; the RP and 'no accent' Midlands speaker elicited significantly fewer positive attributions. For the respondents from the West/South West, the speaker from their own region was also the most frequently positively identified. The neighbouring South East and Welsh accents were also strongly identified. In contrast to the Welsh respondents, nearly 50% of respondents from the West/South West placed the RP speaker as coming form the South/South East of England.

Similarly high numbers of South East respondents correctly identified the South West and South East speakers, with a strong recognition also for the Welsh speaker. As with the respondents from the South West, the South East respondents also placed the RP speaker within the South/South East. The 'accentless' Midlands speaker was once again the least accurately identified in terms of regional origin. The figures from the Midlands respondents have, due to their low numbers, been included as a rough guide only; however, in general terms we can see a similar pattern to that observed by respondents from other regions, in the three most frequently accurately identified speakers are those from the South East, the Wales and the South West. The Midlands speaker and the RP speaker are once more the least frequently identified.

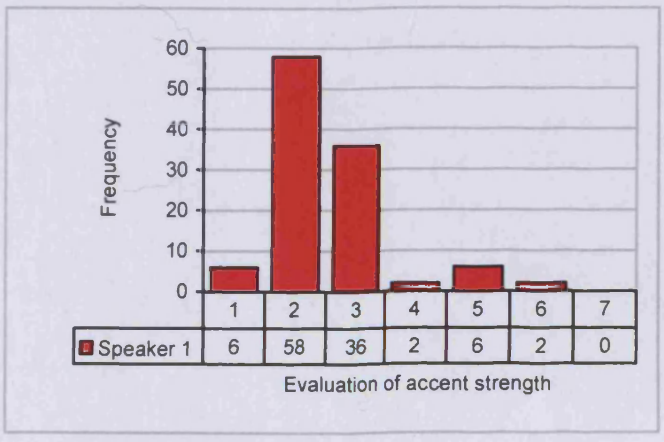
5.5 Accent 'strength': regional/no accent/well-spoken

The respondents were presented with an instrument, upon which they were invited to plot their intuitions regarding the strength of the speakers' accents. As discussed in Section 3.3.7.3, this scale was a composite of two scales, devised to enable respondents to plot their intuitions regarding regionality, and what in non-linguist terms may be referred to as 'accentless' and 'poshness'. The scale was based on the hypothesis that in the minds of the respondents, regional accent and RP are mutually exclusive, but are both intersected by the non-linguist concept of 'no accent'. This design was intended also to be cognitively 'light' and followed the 7-point scale as had been used for previous questions. The use of the scale investigated the commonly held non-linguist belief that speakers 'have an accent (usually regional/national)', 'don't have an accent' (i.e.

the accent is southern, standard English sounding, but without overt regional markers) or 'have a posh/well-spoken accent' (i.e. an accent associated with private schooling and/or ancestral financial privilege). On the scale, '1'='regional accent' and '7'='well spoken', and the scale is intersected at '4' by 'no accent'. What is perhaps unusual about the scale is the attribution of an additional meaning to the mid-point. In the 1-7 scales used elsewhere in the two stages, '4' was unlabelled and is likely to have been understood as a mid point, upon which respondents can plot their evaluation of an absence of strongly marked features either way, or a 'medium amount'. In the no accent scale, 4 was not left to imply absence of regional or well-spoken accent, or no strong opinion, but was denoted as specifically 'no accent'. The scale therefore was an attempt to address three concepts in one.

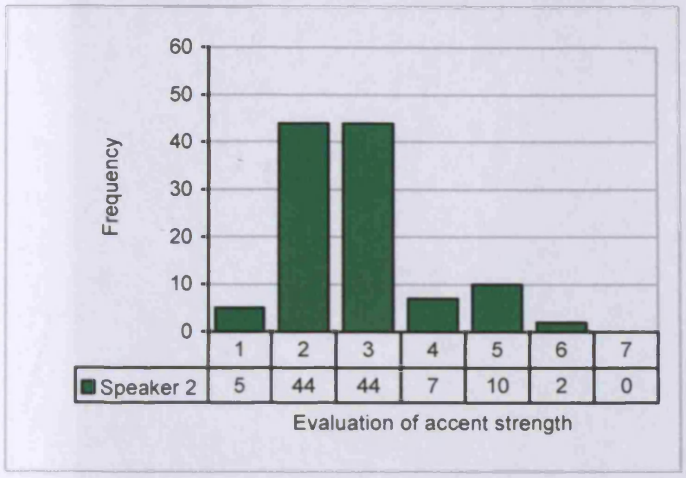
The presentation of the seven-point accent scale generated remarkably consistent patterns of response. The distributions for each speaker are given below, with bar charts for illustrative purposes, followed by a table summarising the means for all speakers and finally an ANOVA to control for experimental variability, that is, to determine whether any differences between the mean ratings for these speakers were statistically significant.

Figure 5.5.1 Attributions for Speaker 1 'accent strength'



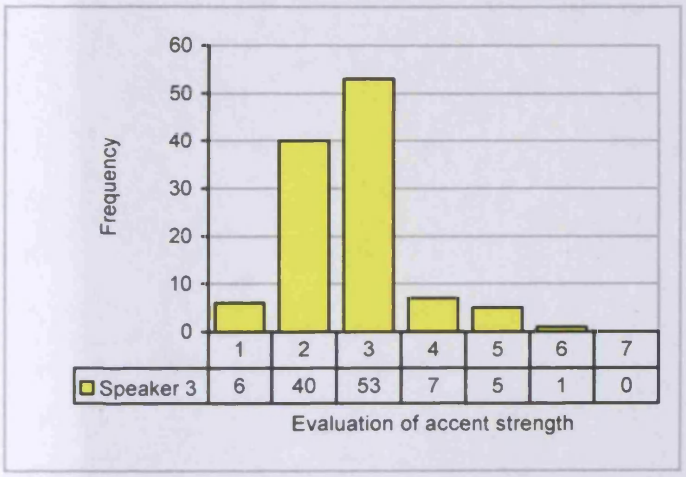
n=110
Mean =2.55
Std. Dev.=0.973

Figure 5.5.2 Attributions for Speaker 2 'accent strength'



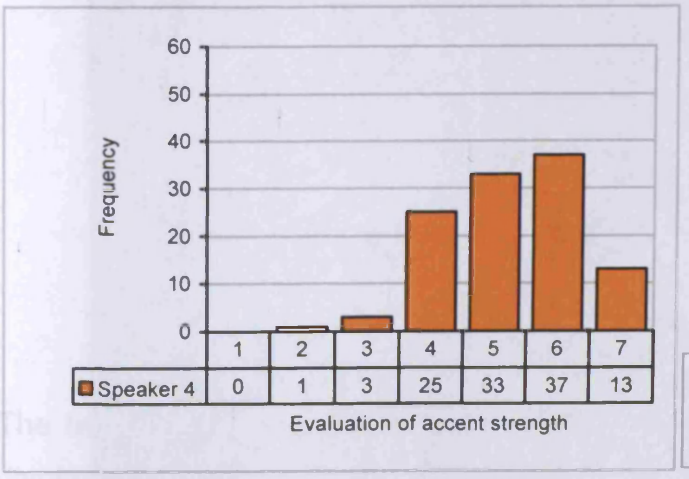
n=112
Mean =2.81
Std. Dev.=1.061

Figure 5.5.3 Attributions for Speaker 3 'accent strength'



n=112
Mean =2.71
Std. Dev.=0.905

Figure 5.5.4 Attributions for Speaker 4 'accent strength'



n=112
Mean =5.26
Std. Dev.=1.072

Figure 5.5.5 Attributions for Speaker 5 'accent strength'

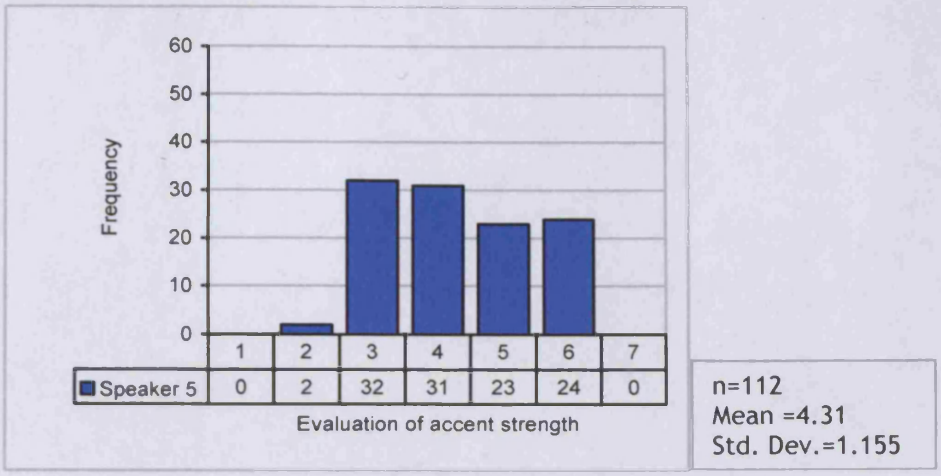


Figure 5.5.6 Table of means for 'accent strength'

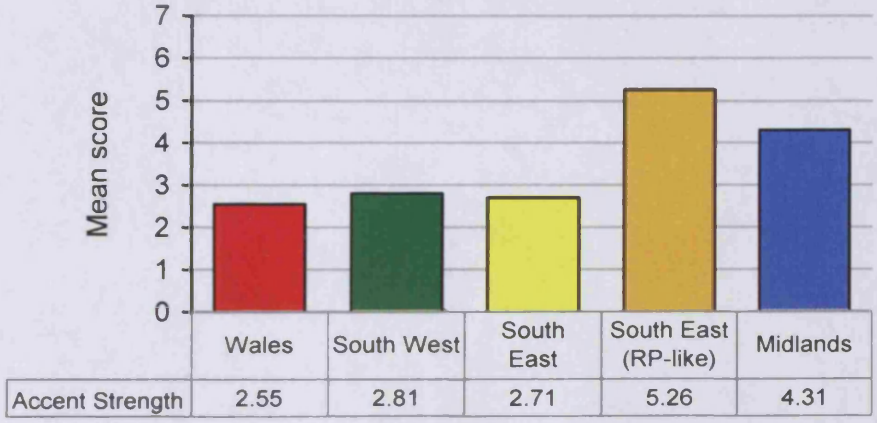


Table 5.6.1 Repeated measures ANOVA for 'accent strength' ($\rho < 0.05$)

	Sig.
Speaker 1 - Speaker 2 - Strength of accent	.02
Speaker 1 - Speaker 3 - Strength of accent	.08
Speaker 1 - Speaker 4 - Strength of accent	.00
Speaker 1 - Speaker 5 - Strength of accent	.00
Speaker 2 - Speaker 3 - Strength of accent	.43
Speaker 2 - Speaker 4 - Strength of accent	.00
Speaker 2 - Speaker 5 - Strength of accent	.00
Speaker 3 - Speaker 4 - Strength of accent	.00
Speaker 3 - Speaker 5 - Strength of accent	.00
Speaker 4 - Speaker 5 - Strength of accent	.00

Adjustment for multiple comparisons: Least significant difference

The table distribution of evaluations, the means and results from the ANOVA display rather striking differences between the speakers, giving us grounds to believe that the treatment, i.e. the use of different voices as stimuli, has had an

effect, and that the differences observed were not a coincidence of random sampling.

The 'accent strength' attributions for Speakers 1, 2 and 3 are remarkably loaded towards one end of the scale and the results of the ANOVA confirm that not only could the majority of respondents identify the origin of the speakers, as seen in previous sections, but they also qualified their evaluation of Speakers 1, 2 and 3 being Welsh, from the South West and South East respectively, as these accents were seen neither well-spoken or 'neutral', but firmly regional. That these speakers are considered to be using 'non-standard' English most probably carries with it some educationally negative connotations, as was evidenced in the ratings for intelligence of the speaker in the results for Stage 1 research (Chapter 4) and in previous attitudes research (Giles 1970; Bishop et al. 2005).

With regards to Speakers 4 and 5, we see how respondents appear to have used the scale to convey well-spokenness for Speaker 4, whose evaluations are significantly different from those for all other speakers. Not only is the speaker not seen as regional, but some 86% of respondents place him on the 'well-spoken' side of the scale. This suggests that not only did respondents find it difficult to place the speaker regionally, but they also made what might be interpreted as a value judgement by indicating that they felt the speaker was not just 'neutral'/'non-regional' but also 'well-spoken', i.e. speaking 'better' than the three 'regional' speakers. The mean score of 5.2 is perhaps not as extreme as it could have been, but serves as indicator that the type of RP being used by Speaker 4 was that which might be distinguished as *general* RP, the form of RP used most commonly in the UK, just short of *conservative* or *advanced* forms of RP (Gimson 1980:91), which we may have expected to attract even higher scores. Although we must consider the possibility that these responses may be a result of the shape of the strength of accent scale and the mutual exclusivity of regional and RP in their arrangement along the scale, there is strong evidence to confirm the researcher suspicion that the speaker's voice was receiving evaluations based on his sociolect rather than regional dialect.

Speaker 5 is the only example of the five speakers where neither points 1 nor 7 are employed by the respondents. The data in previous sections have suggested that Speaker 5 was viewed somewhat differently from other speakers; some identify some regional influences (some Midlanders recognising another Midland speaker); some find the speaker's region to be indeterminable; others feel that his speech has similarities with that associated with Oxford/Cambridge. It is no surprise, then, to see a wide spread of attributions for the 'strength' of Speaker 5's accent. However, over 75% of the respondents rated the speaker around the middle three points on the scale, suggesting that Speaker 5 is seen as somehow less 'well-spoken' than speaker 4, but also less 'regional' than Speakers 1, 2 and 3. The voice donor, who was selected for the study on the basis of his self-evaluation and the evaluation of the researcher that he would be identified as having 'no accent', was evaluated almost exactly thus, receiving a mean evaluative score around the '4' mark. The results also suggest that the 'no accent' scale was meaningful to the respondents and methodologically valid, and that it may have a role to play in any future audiofit.

5.6 Voice features 1 – Bipolar semantic scales

In the following two sections, I set out the details of the findings in response to the research question set out in 2.14.2. regarding how respondents describe their perceptions and conceptions of the speakers' voices using a set of non-linguist-derived audiofit descriptors.

With the aim of developing an audiofit of terms and scales that non-linguists could find useful for articulating their perceptions of unknown voices, voice descriptors suggested by respondents in Stage 1 were checked against the categories suggested by professional forensic linguists, outlined in Chapter 2, such as Hollien (2002) before being reformulated and recycled into an audiofit of bi-polar semantic differential scales, and non-polar descriptors (Section 5.7) for presentation to a second cohort of non-linguists in Study 2.

The features were: speaker speed, speaker hesitancy/confidence, speaker pitch, speaker monotone/melodious, speaker loudness and speaker steadiness. The respondents in both Stages were still exposed to the same voices, however, in Stage 2, respondents are presented with the audiofit framework with which to express their impressions.

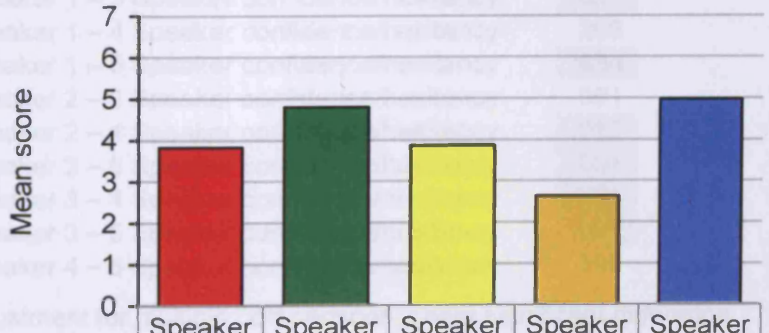
At this point, an important difference should be flagged between this section and other sections that have used the 7-point bipolar semantic scales, such as intelligence, guilt, etc. The previous sections set out to investigate possible relationships between non-linguist evaluations and their attitudes towards speakers of different varieties. However, the current section does not relate voice features to attitudes, but sets out to record respondents' descriptions of voice features, divorced from attitudinal features.

5.6.1 Voice features: comparison of means features by feature

The data are first presented in Section 5.6.1. according to individual voice features, with tables of means presented for each speaker, feature by feature. The means were then subjected to a repeated measures ANOVA to view any significant differences between them. The results of the ANOVA for each voice feature are presented directly underneath the means for each speaker. A commentary follows at the end of the section.

5.6.1.1 Speaker speed

Figure 5.6.1.1 Mean scores for 'speaker speed'



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker Speed	3.83	4.8	3.87	2.64	4.98

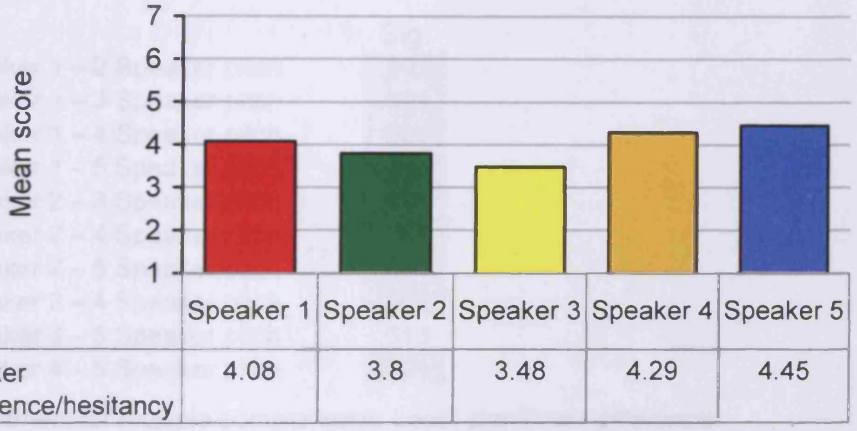
Table 5.6.1.1 Repeated measures ANOVA for 'speaker speed' ($\rho < 0.05$)

	Sig.
Speaker 1 – 2 Speaker speed	.001
Speaker 1 – 3 Speaker speed	.001
Speaker 1 – 4 Speaker speed	.045
Speaker 1 – 5 Speaker speed	.038
Speaker 2 – 3 Speaker speed	.011
Speaker 2 – 4 Speaker speed	.001
Speaker 2 – 5 Speaker speed	.001
Speaker 3 – 4 Speaker speed	.001
Speaker 3 – 5 Speaker speed	.001
Speaker 4 – 5 Speaker speed	.001

Adjustment for multiple comparisons: Least significant difference

5.6.1.2 Speaker confidence/hesitancy

Figure 5.6.1.2 Mean scores for 'speaker confidence/hesitancy'



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker confidence/hesitancy	4.08	3.8	3.48	4.29	4.45

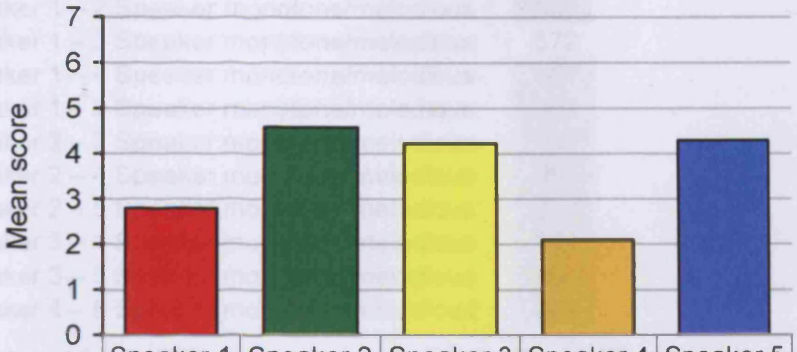
Table 5.6.1.2 Repeated measures ANOVA for 'speaker confidence/hesitancy' ($\rho < 0.05$)

	Sig.
Speaker 1 – 2 Speaker confidence/hesitancy	.173
Speaker 1 – 3 Speaker confidence/hesitancy	.001
Speaker 1 – 4 Speaker confidence/hesitancy	.266
Speaker 1 – 5 Speaker confidence/hesitancy	.034
Speaker 2 – 3 Speaker confidence/hesitancy	.081
Speaker 2 – 4 Speaker confidence/hesitancy	.013
Speaker 2 – 5 Speaker confidence/hesitancy	.001
Speaker 3 – 4 Speaker confidence/hesitancy	.001
Speaker 3 – 5 Speaker confidence/hesitancy	.001
Speaker 4 – 5 Speaker confidence/hesitancy	.390

Adjustment for multiple comparisons: Least significant difference

5.6.1.3 Speaker pitch

Figure 5.6.1.3 Mean scores for 'speaker pitch'



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker pitch	2.8	4.59	4.22	2.1	4.3

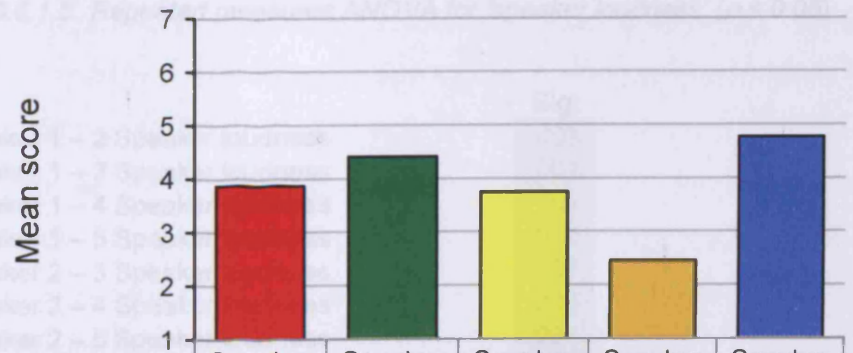
Table 5.6.1.3 Repeated measures ANOVA for 'speaker pitch' ($\rho < 0.05$)

	Sig.
Speaker 1 – 2 Speaker pitch	.001
Speaker 1 – 3 Speaker pitch	.001
Speaker 1 – 4 Speaker pitch	.001
Speaker 1 – 5 Speaker pitch	.001
Speaker 2 – 3 Speaker pitch	.025
Speaker 2 – 4 Speaker pitch	.001
Speaker 2 – 5 Speaker pitch	.037
Speaker 3 – 4 Speaker pitch	.001
Speaker 3 – 5 Speaker pitch	.613
Speaker 4 – 5 Speaker pitch	.001

Adjustment for multiple comparisons: Least significant difference

5.6.1.4 Speaker monotone/melodious

Figure 5.6.1.4 Mean scores for 'speaker monotone/melodious'



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker monotone/melodious	3.87	4.41	3.75	2.45	4.77

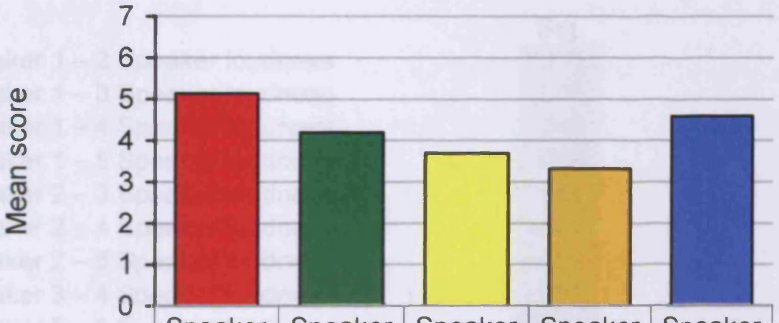
Table 5.6.1.4 Repeated measures ANOVA for 'speaker monotone/melodious ($\rho < 0.05$)

	Sig.
Speaker 1 – 2 Speaker monotone/melodious	.009
Speaker 1 – 3 Speaker monotone/melodious	.572
Speaker 1 – 4 Speaker monotone/melodious	.001
Speaker 1 – 5 Speaker monotone/melodious	.001
Speaker 2 – 3 Speaker monotone/melodious	.001
Speaker 2 – 4 Speaker monotone/melodious	.001
Speaker 2 – 5 Speaker monotone/melodious	.024
Speaker 3 – 4 Speaker monotone/melodious	.001
Speaker 3 – 5 Speaker monotone/melodious	.001
Speaker 4 – 5 Speaker monotone/melodious	.001

Adjustment for multiple comparisons: Least significant difference

5.6.1.5 Speaker loudness

Figure 5.6.1.5 Mean scores for 'speaker loudness'



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker loudness	5.14	4.2	3.7	3.32	4.59

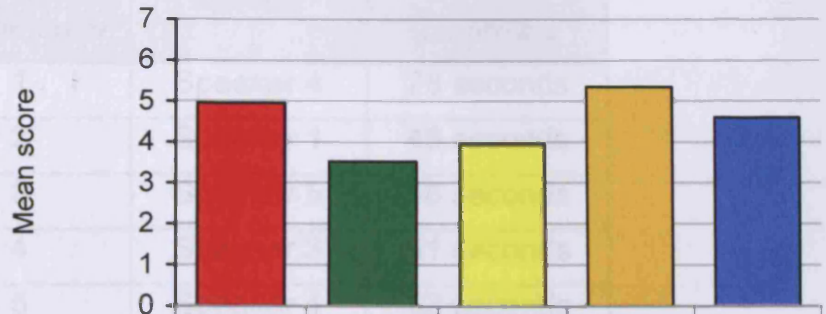
Table 5.6.1.5 Repeated measures ANOVA for 'speaker loudness' ($\rho < 0.05$)

	Sig.
Speaker 1 – 2 Speaker loudness	.001
Speaker 1 – 3 Speaker loudness	.001
Speaker 1 – 4 Speaker loudness	.001
Speaker 1 – 5 Speaker loudness	.001
Speaker 2 – 3 Speaker loudness	.001
Speaker 2 – 4 Speaker loudness	.001
Speaker 2 – 5 Speaker loudness	.024
Speaker 3 – 4 Speaker loudness	.019
Speaker 3 – 5 Speaker loudness	.001
Speaker 4 – 5 Speaker loudness	.001

Adjustment for multiple comparisons: Least significant difference

5.6.1.6 Speaker steadiness

Figure 5.6.1.6 Mean scores for 'speaker steadiness'



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker steadiness	4.95	3.52	3.95	5.35	4.6

Table 5.6.1.6 Repeated measures ANOVA for 'speaker steadiness'

	Sig.
Speaker 1 – 2 Speaker loudness	.001
Speaker 1 – 3 Speaker loudness	.001
Speaker 1 – 4 Speaker loudness	.045
Speaker 1 – 5 Speaker loudness	.038
Speaker 2 – 3 Speaker loudness	.011
Speaker 2 – 4 Speaker loudness	.001
Speaker 2 – 5 Speaker loudness	.001
Speaker 3 – 4 Speaker loudness	.001
Speaker 3 – 5 Speaker loudness	.001
Speaker 4 – 5 Speaker loudness	.001

Adjustment for multiple comparisons: Least significant difference

The results presented in the tables above make for interesting viewing as for almost every feature, the means tell us that the speaker was evaluated significantly differently from the others, indicating that the labels and scales provided were salient to the respondents. The largest number of differences occurred in the data for speed, loudness and speaking steadiness; the least significant differentiation is to be seen in the confidence/hesitancy ratings, although there are still patterns of differentiation to be seen.

Table 5.6.1.7 Ranked order for 'speaker speed'

Respondent ranked order	Speaker	Length of recording
1	Speaker 4	28 seconds
2	Speaker 1	46 seconds
3	Speaker 5	48 seconds
4	Speaker 3	51 seconds
5	Speaker 4	52 seconds

If we view the ratings in terms of rankings, as shown in Table 5.6.1.7, we see some clear patterns. For 'speaker speed', we see that the Speaker 4 is evaluated as the fastest and Speaker 2 the slowest. What is particularly remarkable here is that the ratings for 'speaker speed' precisely match the length of the recordings. It would appear that respondents were highly attuned to the length of time that the speaker was speaking and the speed at which they imparted their information.

For 'speaker confidence', we see both Speakers 4 and 5 rated most highly and Speakers 2 and 3 rated lowest; for 'speaker pitch' we see Speaker 2 rated as having the highest voice and Speaker 4 having the lowest; for 'speaker monotone/melodious' we see Speaker 5 as being the most melodious and Speaker 4 being rated the most monotone; for 'speaker loudness' we see a significantly differentiated ranking from Speaker 1 the loudest, then Speakers 5, 2, 3, with Speaker 4 being rated the softest speaker; for 'speaker steadiness' we again see a significantly differentiated ranking from Speaker 4 being the most confident to Speakers 1, 5, 3 to 2 being the least steady. How these features can be translated into 'speaker profiles' and how the intra-speaker associations of the ratings are discussed in the following section.

5.6.2 Voice features: comparison speaker by speaker

Section 5.6.2 presents the data for voice features, speaker by speaker, enabling a 'speaker profile' view of the data. First the data for each speaker are

presented in a table of frequencies, including mean and median, allowing us to see how, on average, the respondents rated each voice feature presented. Additionally, in order to endow us with a view of the distribution of the data or consensus on the ratings, boxplots have also been employed, displaying the dispersion of the evaluations, the median interquartile range and the lower and upper quartiles and any outliers. By viewing the data on a boxplot, we may see where respondents felt that a particular feature or features were distinctive, the 'stand-out' features that diverge from a presumed default norm.

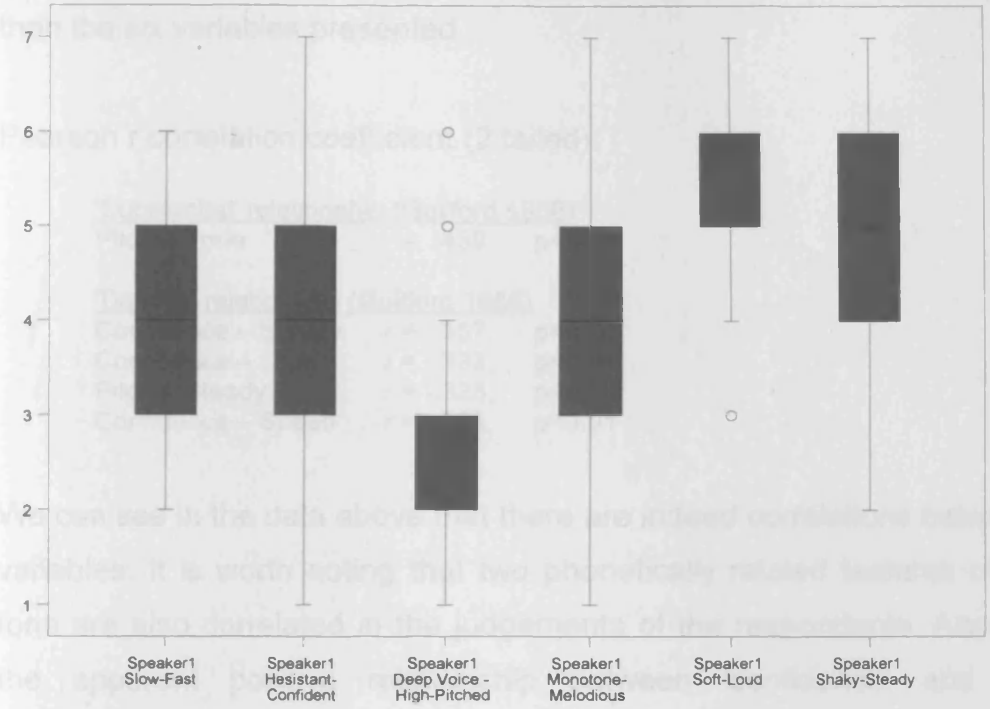
When viewing the data in terms of a 'speaker profile', the six voice features observed are set out adjacent to one another for ease of view. It should be remembered, however, that each feature represents a different evaluative dimension of the speaker's voice and direct normative, within-speaker comparisons are disfavoured. We would not state, for example, that the speaker's voice is significantly deeper than his speed is fast, as such a comparison is illogical. With this in mind, the statistical treatments in this section are also different: it is not appropriate to use an ANOVA, t-test or other analysis of the equality of means, as to do so would not be comparing like with like. What statistical analysis can demonstrate here is whether any of the independent voice features are associated or co-vary with each other and how the six voice features under investigation may load onto a smaller number of factors. To accomplish this, two statistical treatments have been used: a two-tailed Pearson *r* correlation indicates any potential cross-feature relationships, followed by a varimax rotation with Kaiser normalisation, which enables us to identify any factorial loadings.

5.6.2.1 Speaker 1 – Wales

Table 5.6.2.1.1 Means and medians for Speaker 1 voice characteristics

	Speaker 1 Slow-Fast	Speaker 1 Hesitant-Confident	Speaker 1 Deep voice-High-pitched	Speaker 1 Monotone-Melodious	Speaker 1 Soft-Loud	Speaker 1 Shaky-Steady
N Valid	105	104	105	103	96	95
Missing	7	8	7	9	16	17
Mean	**3.88**	**4.02**	**2.81**	**3.94**	**5.11**	**4.95**
Median	4.00	4.00	3.00	4.00	5.00	5.00

Figure 5.6.2.1.1 Boxplots for Speaker 1 voice characteristics



The boxplot below demonstrates that three of the features, 'slow-fast', 'hesitant-confident' and 'monotonous-melodious' do not appear to have any strong tendency either way, and two of them have whiskers extending across the full scale. This indicates neither a great deal of consensus, or strong opinion. However, 'deep voice – high-pitched' does provoke a very clear reaction, with 50% of respondents placing the voice between points 2 and 3, and the 25th and 75th quartiles landing only one point either side. This is a strong indication that the respondents felt Speaker 1 to have a deep voice. We can observe the direct opposite where the speaker's 'soft – loud' rating is concerned, with 50% of the respondents placing the speaker between points 5 and 6 and the remaining 25th and 75th quartiles landing between point 4 and 7. In summary, we can observe that thus far in the results for Stage 2, Speaker 1 is characterised by the respondents as being Welsh, probably from Cardiff, with a regional accent, a deep, loud and quite steady voice.

However, we should also investigate whether these features are all acting individually or whether they betoken similar traits for the respondents, using a two-tailed Pearson *r* correlation to establish any potential cross-feature relationships, followed by a data reduction test through a principal component

analysis, in order to investigate whether features are indeed loaded onto fewer than the six variables presented.

Pearson r correlation coefficient (2 tailed):

'Substantial' relationship (Guilford 1956)
Pitch – Tone r = .439 p<0.01

'Definite' relationship (Guilford 1956)
Confidence – Steady r = .357, p<0.01
Confidence – Loud r = .333, p<0.01
Pitch – Steady r = -.326, p<0.01
Confidence – Speed r = .323, p<0.01

We can see in the data above that there are indeed correlations between the six variables. It is worth noting that two phonetically related features of pitch and tone are also correlated in the judgements of the respondents. Also of note is the apparent positive relationship between 'confidence' and increased 'steadiness', 'speed' and 'loudness'. Is this relationship constant across all speakers or local to one? And how in a future Stage would these features relate to those investigated in Stage 1, 'intelligence', 'honesty' and 'aggression'? Of immediate interest, however, is the factorial relationship between these features. A confirmatory test was run, a principal component analysis, plotting the Eigenvalues.

Table 5.6.2.1.2 Principal component analysis for Speaker 1 voice characteristics

Total Variance Explained

Component	Initial Eigenvalues	
	Total	% of Variance
1	**1.936**	**32.271**
2	**1.608**	**26.807**
3	.805	13.421
4	.713	11.891
5	.535	8.919
6	.401	6.691

Extraction Method: Principal Component Analysis.

Each is shown to have an Eigenvalue higher than 1 in the table below, accounting between them for over 50% of the variance (59%), confirming the results of the scree test, which also identified two underlying factors. Having identified more than one factor underlying the relationships between the variables, a varimax rotation with Kaiser normalisation was performed in order

to identify the factor loadings – the strength of each variable in defining the factor:

Table 5.6.2.1.3 Rotated Component Matrix for Speaker 1 voice characteristics

	Component	
	1	2
Speaker 1 Confidence	**.768**	.072
Speaker 1 Loud	**.715**	-.032
Speaker 1 Speed	**.624**	.356
Speaker 1 Steady	**.594**	-.531
Speaker 1 Pitch	-.064	**.854**
Speaker 1 Tone	.269	**.694**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 3 iterations.

The rotated component matrix suggests that loadings that group together are 'confidence', 'loudness', 'speed' and 'steadiness' of voice. In statistics, labelling a factor is always the subjective decision of the researcher relative to our common sense interpretations of the group relationship. In this context a suggested label is 'force and control'.

We can also observe a second factor at work, containing the variables 'pitch' and 'tone', which we may say are loading on to a factor called 'heard pitch' after Hollien (2002). Furthermore, the component variables of 'heard pitch' were also identified as the strongest correlation above. These factors in many ways replicate the data described in the boxplots above.

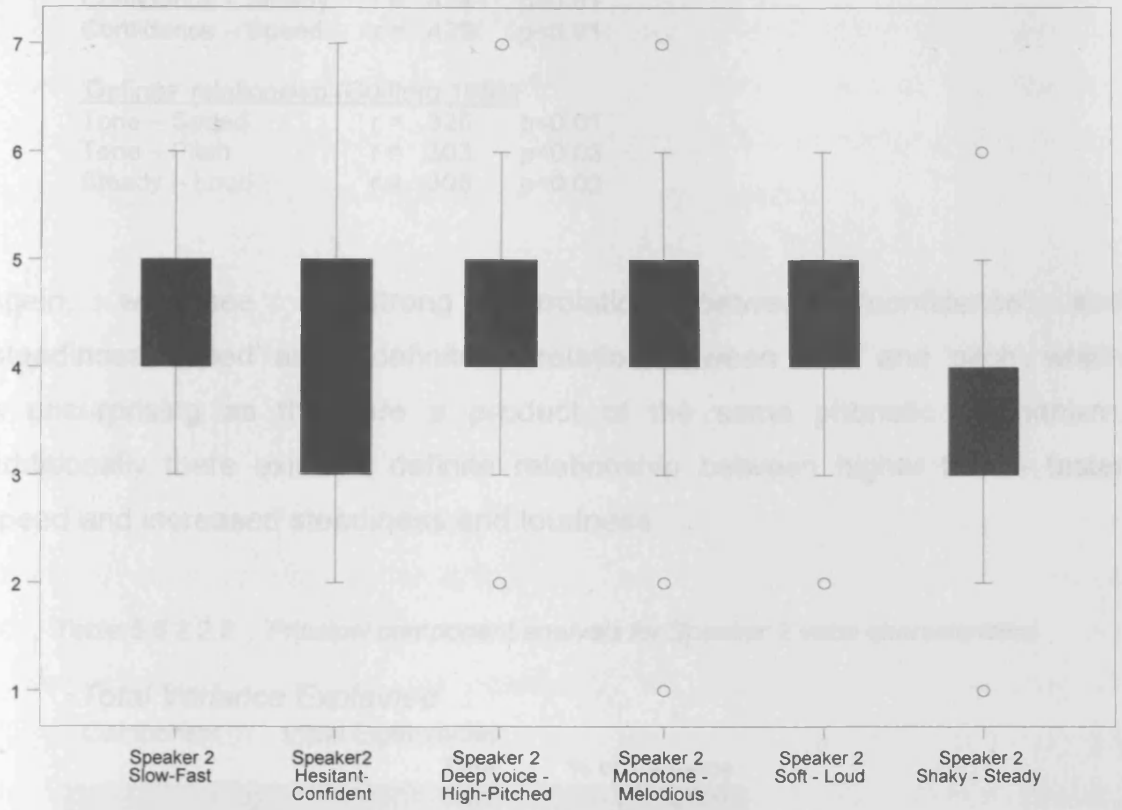
5.6.2.2 Speaker 2 – South West

With Speaker 2, we can observe some differences across the means, with speed, pitch, melody and loudness all presenting counts of >4, and hesitance and shakiness both having frequencies of <4.

Table 5.6.2.2.1 Means and medians for Speaker 2 voice characteristics

		Speaker 2 Slow-Fast	Speaker 2 Hesitant-Confident	Speaker 2 Deep voice-High-pitched	Speaker 2 Monotone-Melodious	Speaker 2 Soft-Loud	Speaker 2 Shaky-Steady
N	Valid	105	108	101	96	95	102
	Missing	7	4	11	16	17	10
Mean		4.74	3.69	4.50	4.30	4.20	3.53
Median		5.00	3.00	5.00	4.00	4.00	3.00

Figure 5.6.2.2.1 Boxplots for Speaker 2 voice characteristics



The inter-quartile ranges for 'slow – fast', 'deep voice – high-pitched', 'monotone – melodious' and 'soft – loud', demonstrate the tendency for the respondents to place these ratings on the higher side of the rating scales. A contrasting effect is observed with the data regarding 'shaky – steady', with the interquartile 50% being located on the lower half of the scale. Opinions on 'hesitant – confident' appear broad, with 25th and 75th quartiles extending across the full range, and the interquartile range straddling the middle point 4.

It would appear, then, that respondents have tended to characterise Speaker 2 as a speaker from the South West of England, with a clearly identifiable strong, regional, South West accent, many specifying Bristol, and with generally faster, higher, melodious, louder voice and a little shaky. However, are the characteristics described in the statistics above individual characteristics, or is there a relationship between them in relation to the voice heard? The following correlations were observed in the Pearson r correlation coefficient (2 tailed):

'Substantial' relationship (Guilford 1956)

Confidence – Steady	r = .474	p<0.01
Confidence – Speed	r = .429	p<0.01

'Definite' relationship (Guilford 1956)

Tone – Speed	r = .325	p<0.01
Tone – Pitch	r = .303	p<0.03
Steady – Loud	r = .303	p<0.03

Again, we see a strong correlation between 'confidence' and 'steadiness'/'speed' and a definite correlation between 'tone' and 'pitch', which is unsurprising as they are a product of the same phonetic mechanism. Additionally there exists a definite relationship between higher tone - faster speed and increased steadiness and loudness.

Table 5.6.2.2.2 Principal component analysis for Speaker 2 voice characteristics

Total Variance Explained

Component	Initial Eigenvalues	
	Total	% of Variance
1	**1.969**	**32.818**
2	**1.433**	**23.879**
3	.840	14.003
4	.748	12.474
5	.628	10.462
6	.382	6.363

Extraction Method: Principal Component Analysis.

The principal component analysis illustrates that there are two components with Eigenvalues over 1, which account for more than 50% of the variance (56%). This confirms that there are two main factors at work, specifically:

Table 5.6.2.2.3 Rotated Component Matrix for Speaker 2 voice characteristics

	Component	
	1	2
Speaker 2 Steady	**.787**	-.146
Speaker 2 Confidence	**.762**	.318
Speaker 2 Loud	**.601**	-.026
Speaker 2 Tone	.059	**.760**
Speaker 2 Pitch	-.296	**.670**
Speaker 2 Speed	.424	**.649**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 3 iterations.

Of interest here are the loadings for Speaker 2 which differ from those for Speaker 1. What we see in these factorial loadings is that 'confidence'-'steadiness', 'steadiness'-'loudness' correlations observed earlier load onto the same factor, a variable we might label 'controlled presence'.

The 'pitch'-'tone', 'tone-speed' correlations observed above load onto an alternative factor which we might label 'prosody'. This difference between the factors present for Speakers 1 and 2 suggests the respondents are interacting with the speakers on a multi-dimensional basis: not only do they differentiate between speakers in terms of their accent, but they perceive differences in the individual voice qualities of the speakers. These results suggest that in studies that observe differing evaluations between speakers (such as speaker accent versus attributions of guilt, for example), the speaker evaluations may be mediated by more than an accent-level response to speakers, and may be predicted by looking beyond the accent of the speaker to more person-specific features.

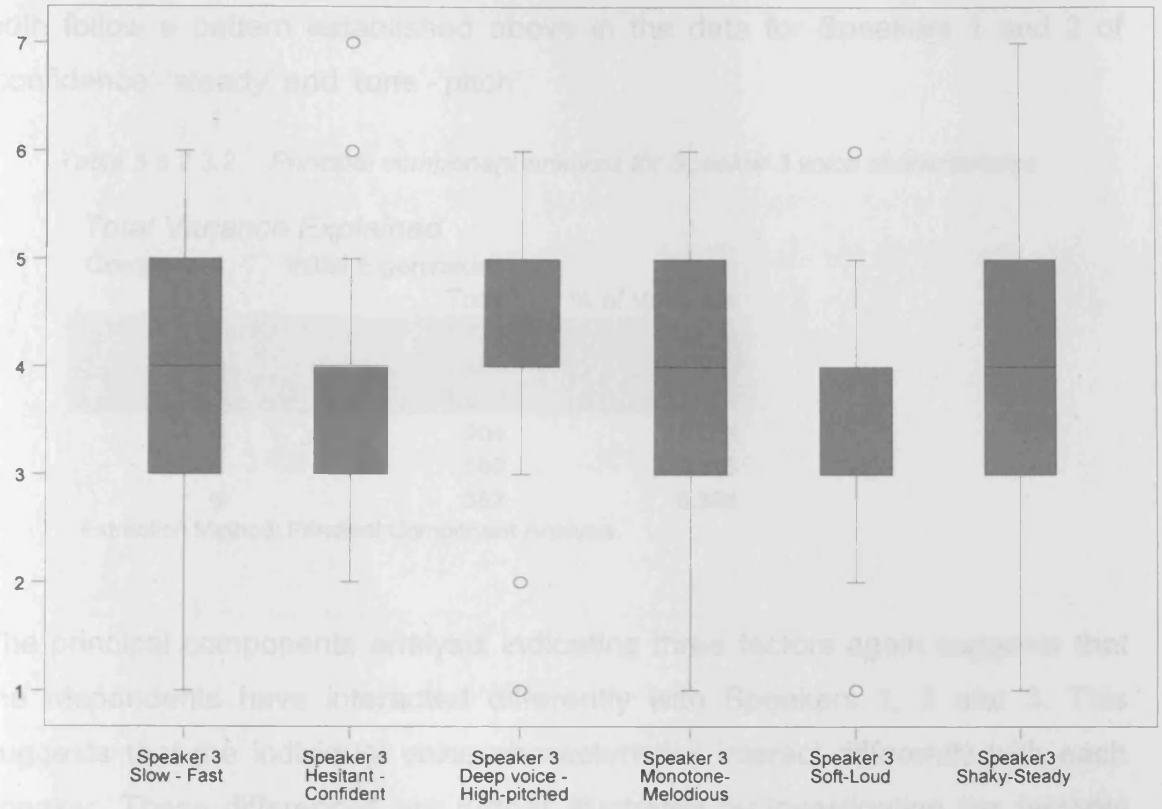
5.6.2.3 Speaker 3 – South East

The means for Speaker 3 demonstrate more variation between the ratings for the individual voice characteristics.

Table 5.6.2.3.1 Means and medians for Speaker 3 voice characteristics

	Speaker 3 Slow-Fast	Speaker 3 Hesitant-Confident	Speaker 3 Deep voice-High-pitched	Speaker 3 Monotone-Melodious	Speaker 3 Soft-Loud	Speaker 3 Shaky-Steady
N Valid	103	104	105	95	95	94
Missing	9	8	7	17	17	18
Mean	3.89	3.58	4.12	3.71	3.71	4.07
Median	4.00	3.00	4.00	4.00	4.00	4.00

Figure 5.6.2.3.1 Boxplots for Speaker 3 voice characteristics



The means suggest that Speaker 3 was generally regarded as having a fairly high-pitched voice and was a little hesitant. The boxplot gives us further insight, illustrating how the central interquartile rests around the 3-4 points for hesitancy and speaking softly, and around 4-5 points for high-pitched voice. Other characteristics are firmly central and with quite extended whiskers, suggesting there was, for the majority, they were not overly distinctive. Speaker 3 is therefore characterised as from the South East of England, possibly London, but with a fairly strong regional accent. On the recording, he speaks with some hesitancy, softly and with a fairly high-pitched voice.

Pearson r correlation coefficient (2 tailed):

'Substantial' relationship (Guilford 1956)
Confidence – Steady r = .500 p<0.01

'Definite' relationship (Guilford 1956)
Tone – Pitch r = .287 p<0.01

The Pearson *r* correlation coefficient reveals two correlations, although they both follow a pattern established above in the data for Speakers 1 and 2 of 'confidence'-'steady' and 'tone'-'pitch'.

Table 5.6.2.3.2 Principal component analysis for Speaker 3 voice characteristics

Total Variance Explained

Component	Initial Eigenvalues	
	Total	% of Variance
1	**1.723**	**28.710**
2	**1.299**	**21.657**
3	**1.022**	**17.030**
4	.904	15.065
5	.589	9.815
6	.382	6.363

Extraction Method: Principal Component Analysis.

The principal components analysis indicating three factors again suggests that the respondents have interacted differently with Speakers 1, 2 and 3. This suggests that the individual voice characteristics interact differently with each speaker. These differences are further illustrated by investigating the factorial loadings for each variable in a rotated component matrix:

Table 5.6.3.3.3 Rotated Component Matrix for Speaker 3 voice characteristics

	Component		
	1	2	3
Speaker 3 Steady	**.881**	-.065	.066
Speaker 3 Confidence	**.672**	.363	-.090
Speaker 3 Speed	-.097	**.887**	.021
Speaker 3 Loud	.252	**.607**	.107
Speaker 3 Tone	.287	.099	**.819**
Speaker 3 Pitch	-.392	.027	**.714**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 5 iterations.

The table above suggests that 'steadiness' and 'confidence' load onto one factor, which we might call 'exhibiting confidence'. 'Speed' and 'loudness' load onto a second factor we might label 'punchiness', and 'tone' and 'pitch' remain together in Hollien's 'heard pitch' factor. We may also note that of the three factors present, two of them were also the only correlations present in the Pearson *r* test, and that across all three speakers there appears to be variation

in the interaction, but with the same 'rump' of variables loading onto the same factors.

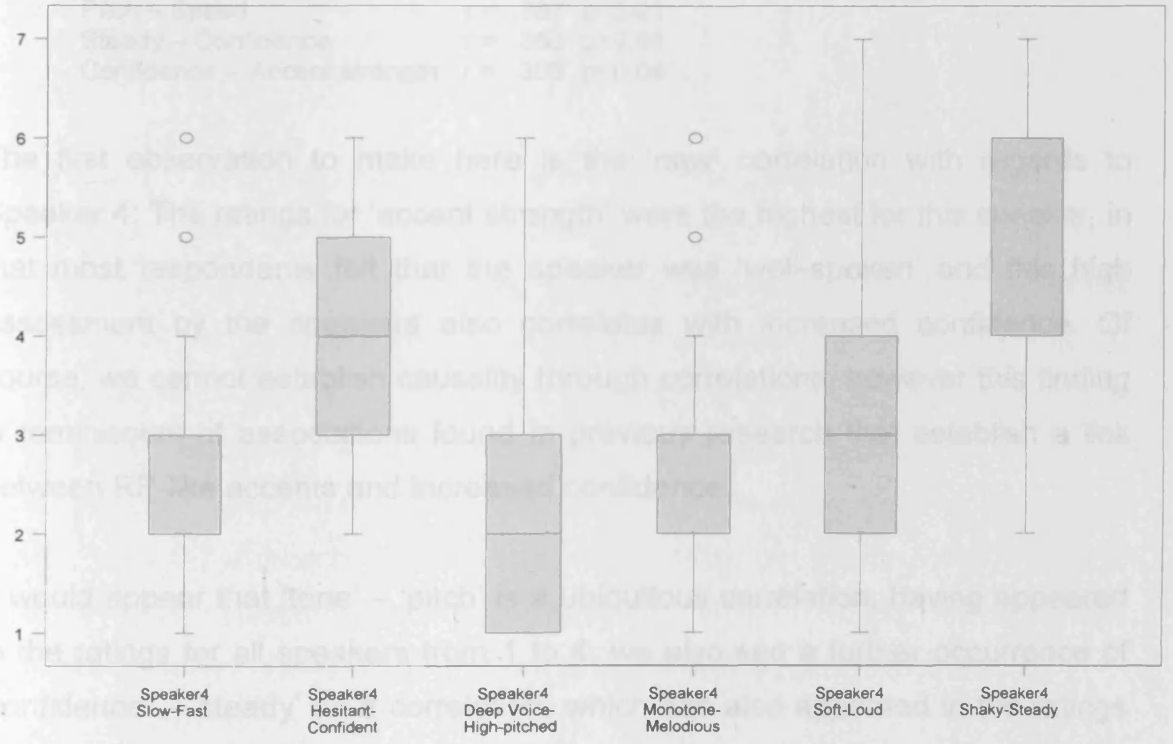
5.6.2.4 *Speaker 4 – South East (RP-like)*

The means for Speaker 4 are remarkable in their variation, demonstrating low scores for 'slow', 'deep voice' and monotone, and high scores for 'steady'

Table 5.6.2.4.1 Means and medians for Speaker 4 voice characteristics

		Speaker 4 Slow-Fast	Speaker 4 Hesitant-Confident	Speaker 4 Deep voice-High-pitched	Speaker 4 Monotone-Melodious	Speaker 4 Soft-Loud	Speaker 4 Shaky-Steady
N	Valid	109	100	107	106	97	94
	Missing	3	12	5	6	15	18
Mean		2.62	4.24	2.15	2.53	3.31	5.36
Median		2.00	4.00	2.00	2.00	3.00	6.00

Figure 5.6.2.4.1 Boxplots for Speaker 4 voice characteristics



The boxplot reveals the distribution of opinion regarding these voice features, and we can see that for 'slow' and 'monotone', the interquartile range rests between points 2 to 3 with the 25th and 75th percentile whiskers extending only one point in either direction. And whilst there is an extended 75th percentile for 'deep voice - high-pitched' the interquartile range is between points 1 to 3, indicating a very strong consensus regarding the pitch of the voice. There is

some indication also that the respondents generally felt Speaker 4 to be speaking softly in the recording, although the 75th percentile whisker does extend fully to point 7.

By contrast, the interquartile range rests between points 4 to 6 for 'steady', although these data are skewed towards point 4. The data on Speaker 4 appear to indicate the respondents felt him to be demonstrating a 'well-spoken' 'Oxbridge' accent, from the South of England, speaking relatively slowly, with a deep, monotonous, relatively softly-spoken but steady voice.

Pearson r correlation coefficient (2 tailed):

'Substantial' relationship (Guilford 1956)		
Tone – Speed	r = .459	p<0.01
Tone – Loud	r = .433	p<0.01
Tone – Pitch	r = .415	p<0.01

'Definite' relationship (Guilford 1956)		
Pitch – Speed	r = .387	p<0.01
Steady – Confidence	r = .363	p<0.01
Confidence – Accent strength	r = .303	p<0.04

The first observation to make here is the 'new' correlation with regards to Speaker 4: The ratings for 'accent strength' were the highest for this speaker, in that most respondents felt that the speaker was 'well-spoken' and this high assessment by the speakers also correlates with increased confidence. Of course, we cannot establish causality through correlations, however this finding is reminiscent of associations found in previous research that establish a link between RP-like accents and increased confidence.

It would appear that 'tone' – 'pitch' is a ubiquitous correlation, having appeared in the ratings for all speakers from 1 to 4; we also see a further occurrence of 'confidence' – 'steady' as a correlation, which has also appeared in the ratings for all speakers. There is also evidence of 'tone' correlating strongly with 'speed', 'loudness' and 'pitch' and other mutual correlations between those variables. A principal components analysis demonstrates that there are only two factors present amongst the six variables with values over 1 accounting for more than 50% of the variance (58%).

Table 5.6.2.4.2 Principal component analysis for Speaker 4 voice characteristics

Total Variance Explained

Component	Initial Eigenvalues	
	Total	% of Variance
1	**2.098**	**34.967**
2	**1.383**	**23.044**
3	.792	13.204
4	.690	11.503
5	.636	10.595
6	.401	6.686

Extraction Method: Principal Component Analysis.

Table 5.6.2.4.3 Rotated Component Matrix for Speaker 4 voice characteristics

	Component	
	1	2
Speaker 4 Tone	**.812**	.096
Speaker 4 Speed	**.698**	-.150
Speaker 4 Pitch	**.660**	-.057
Speaker 4 Loud	**.644**	.025
Speaker 4 Confidence	.181	**.850**
Speaker 4 Steady	-.242	**.797**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 3 iterations.

The Table 5.6.2.4.3 suggests that we are observing a factor seen in the data for Speaker 3, with 'confidence' and 'steady' loading onto one factor entitled 'exhibiting confidence'. Moreover, the factor analysis almost entirely replicates the results of the above Pearson *r* correlation for Speaker 4. However, there is an even larger parallel to be observed: the larger of the Pearson *r* correlations found above related to the features of 'loud', 'pitch', 'speed' and 'tone', all four of which appear to load on to a new factor featuring the more physical aspects of speech rather than psychological performance features; this factor is given the label 'physical performance'. This presents us with strong evidence for understanding how the respondents felt towards and assessed Speaker 4.

5.6.2.5 Speaker 5 – Midlands

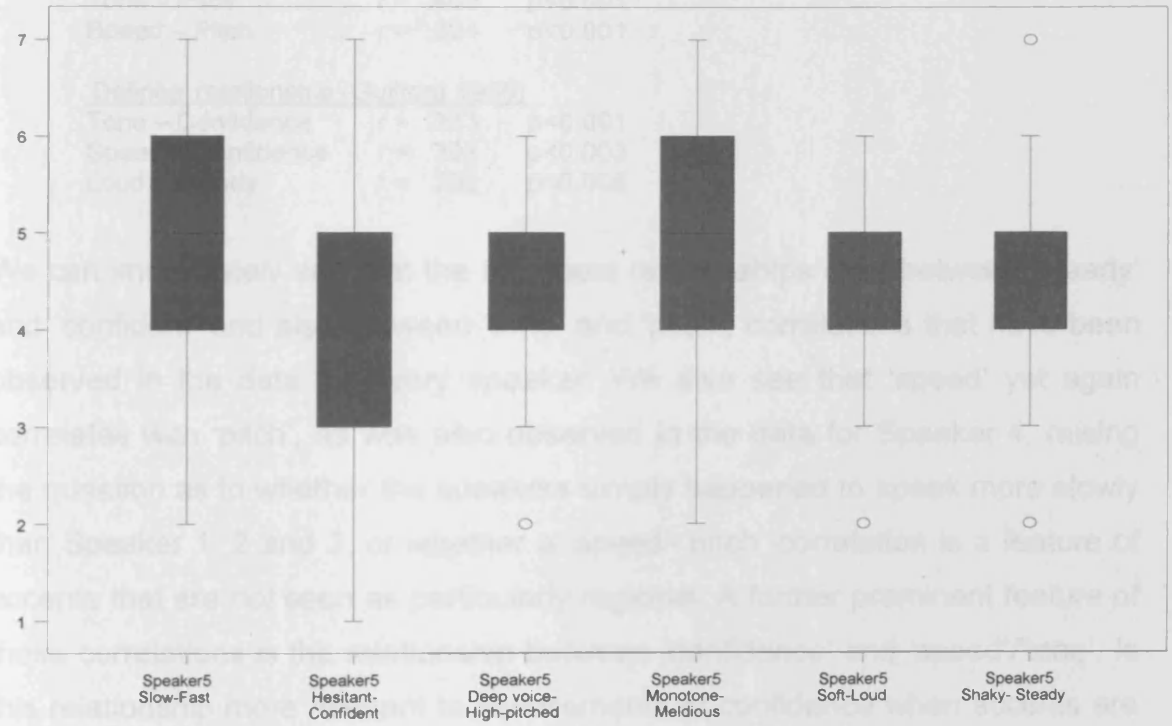
The data in the previous sections have suggested that Speaker 5 has presented respondents with some difficulty in terms of identifying his accent on a regional basis. The question here is whether the difficulty in identifying and labelling his accent obscures other vocal features? Can we expect to see a selection of

means centring on an indistinct point 4? The table of means suggest this is not the case. Regardless of difficulties attributing a regional identity to Speaker 5, respondents were still able to give information regarding the speaker's voice.

Table 5.6.2.5.1 Means and medians for Speaker 5 voice characteristics

		Speaker 5 Slow-Fast	Speaker 5 Hesitant-Confident	Speaker 5 Deep voice-High-pitched	Speaker 5 Monotone-Melodious	Speaker 5 Soft-Loud	Speaker 5 Shaky-Steady
N	Valid	104	109	101	93	93	97
	Missing	8	3	11	19	19	15
Mean		5.01	4.41	4.28	4.78	4.57	4.57
Median		5.00	5.00	4.00	5.00	5.00	5.00

Figure 5.6.2.5.1 Boxplots for Speaker 5 voice characteristics



We can see that all of the means are above 4 and some of them noticeably above 4, such as 'fast' and 'melodious'. However, this does not reveal where the majority of respondents were placing the speakers or whether there was a widespread distribution of scores. The boxplot in Figure 5.6.2.5.1, demonstrates how the interquartile range of scores from respondents rests between points 4 to 6 for 'fast' and 'melodious', with a central median, suggesting that the data are not skewed in either direction. Although not so high on the scale, the interquartile ranges for 'high-pitched', 'loud' and 'steady' are concentrated around 4 to 5 points with 25th and 75th percentile whiskers extending only one

point in either direction. Only the scores for 'hesitant-confident' appear to be unremarkable, based around the central 2 scores with whiskers extending the full length of the scale. Despite presenting some difficulty for many respondents trying to attribute a regional identity to Speaker 5, the data in this and previous sections suggest, therefore, that respondents generally characterised Speaker 5 as speaking with a 'neutral' to slightly 'well-spoken' accent, from the southern half of the UK, but not from Wales, a fairly fast, melodious speaker with a relatively high, loud steady voice.

Pearson r correlation coefficient (2 tailed):

'Substantial' relationship (Guilford 1956)

Steady – Confidence	r = .499	p<0.001
Tone – Pitch	r = .405	p<0.001
Speed – Pitch	r = .394	p<0.001

'Definite' relationship (Guilford 1956)

Tone – Confidence	r = .313	p<0.001
Speed – Confidence	r = .301	p<0.003
Loud – Steady	r = .292	p<0.005

We can immediately see that the strongest relationships exist between 'steady' and 'confident' and also between 'tone' and 'pitch', correlations that have been observed in the data for every speaker. We also see that 'speed' yet again correlates with 'pitch', as was also observed in the data for Speaker 4, raising the question as to whether the speakers simply happened to speak more slowly than Speaker 1, 2 and 3, or whether a 'speed'-'pitch' correlation is a feature of accents that are *not* seen as particularly regional. A further prominent feature of these correlations is the relationship between 'confidence' and 'speed'/'tone'. Is this relationship more relevant to assessments of confidence when accents are viewed as largely 'neutral'? Does an absence of regionality lead respondents to base their attribution on other voice characteristics? Alternatively, do regional accents and the social stereotypes attached to them obscure such characteristics? Principal components analysis confirms the findings of the scree test, demonstrating that there are two factors with Eigenvalues over 1 accounting for more than 50% of the variance (57%).

Table 5.6.2.5.2 Principal component analysis for Speaker 5 voice characteristics

Total Variance Explained

Component	Initial Eigenvalues	
	Total	% of Variance
1	**1.990**	**33.173**
2	**1.409**	**23.478**
3	.871	14.512
4	.815	13.579
5	.539	8.975
6	.377	6.283

Extraction Method: Principal Component Analysis.

The rotated component matrix shows us that the loadings for these two factors are very similar to those for Speaker 2:

Table 5.6.2.5.3 Rotated Component Matrix for Speaker 5 voice characteristics

	Component	
	1	2
Speaker 5 Pitch	**.811**	-.100
Speaker 5 Tone	**.680**	.188
Speaker 5 Speed	**.679**	-.021
Speaker 5 Steady	.019	**.869**
Speaker 5 Loud	-.065	**.688**
Speaker 5 Confidence	.473	**.560**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 3 iterations.

We can observe from Table 5.6.2.5.3 that there are once again two factors. Firstly we can see 'prosody' as previously observed in the data for Speaker 2. 'Prosody' comprises the same three variables that were observed to be almost mutually correlating in the Pearson *r* data for Speaker 5 - comprising 'pitch', 'tone' and speed'. We also see a second factor re-emerging of 'controlled presence', comprising 'steady', 'loud' and 'confident', with 'steady' and 'confident' also being the strongest correlates above and 'steady' and 'loud' also being seen to correlate.

We can therefore observe that throughout all of the data for the speakers, there are multiple examples of correlations, recurring in identical or related patterns. Furthermore, most of those correlates have also been observed through factor analysis, confirmed in number by scree plots and Eigenvalues, to be consistently loading onto factors, in frequently occurring patterns. The possible meanings of these observations will be discussed in the next chapter.

5.7 Voice features 2 – Non-polar features

5.7.1 Small data sets

A second section of the qualitative data collected from Stage 1 did not lend itself to a bipolar design and so was presented to respondents as 'other distinctive voice features'. Although these data were also parametric, being based on frequency counts, the questions appeared as an optional section which respondents should only tick if they felt there was anything distinctive worth noting and as such, may have attracted a lower response rate than questions formulated on bipolar scales, where respondents can note the lack of strong opinion by selecting the central mark. In this context, there is no scale, no central mark and no gradation. The respondents are simply presented with a selection of possible voice characteristics that may or may not apply and they tick the box, or pass over as they feel appropriate.

This lower frequency presents its own problems as the statistical tests employed previously are unlikely to be feasible with very low frequencies. However it would be unwise to reject data out of hand simply because of a low frequency. For example, the field of forensic linguistic analysis is familiar with dealing with smaller data samples, from suicide notes to snatched recordings of speakers. One approach in such circumstances maybe to distinguish between what we could call '*criterion-based*' analysis and '*normative*' analysis. All of the previous criterion-based analysis in Stage 2 has involved comparing respondents' feelings towards a set of speakers against a pre-selected set of criteria – region of origin, strength of accent, bipolar measurements of voice features. It can be argued that criteria-based analysis fundamentally leans towards statistical analysis involving as it does the use of sufficiently high frequencies in order to check for statistically significant results.

The alternative normative approach does not rely on statistical significance in order to reach satisfactory analyses. Rather, analysis of smaller data sets can lean more towards comparison of data sets against each other. The combination of these two approaches is illustrated by Eagleson (1994): In presenting a text-based case study, Eagleson demonstrates an approach that is

fundamentally normative, identifying distinctive features, grouping by type and then performing comparisons against candidate texts. In this section, we have a case of criterion-based features requiring a normative analysis. What is different and what is the same? Which features are shared by all speakers, and which are not? Which features cluster together around one speaker, and not others? Which features are simply too low in their scores to be meaningful? The analysis below utilises normative comparisons to identify those features of each speaker deemed to be distinctive by the respondents. This inevitably relies on the skill and subjective judgement of the researcher.

5.7.2 Analysis

Table 5.7.2 below contains frequency counts for all non-polar voice features, for all 5 speakers, including the sum, mean and median for the features and for each speaker. 'X' denotes categories within which there is little between-speaker variation in the evaluations of the respondents:

Table 5.7.2 Non-polar voice features for all speakers

	Spkr 1	Spkr 2	Spkr 3	Spkr 4	Spkr 5	Sum	**Mean**	Median	
dry	7	7	4	24	8	50	**10**	12	
whispery	0	0	5	17	0	22	**4.4**	8.5	
sharp	5	5	6	2	6	24	**4.8**	3	x
shrill	0	3	3	2	1	9	**1.8**	1.5	x
effeminate	1	16	9	1	2	29	**5.8**	8	
rough	19	6	7	13	2	47	**9.4**	9.5	
croaky	9	16	0	5	13	43	**8.6**	8	
raspy	4	5	1	5	3	18	**3.6**	2.5	x
lisp	0	8	2	1	1	12	**2.4**	4	
stammer	3	12	9	0	10	34	**6.8**	6	
stutter	3	18	3	0	8	32	**6.4**	9	
Sum	51	96	49	70	54				
Mean	**4.6**	**8.7**	**4.5**	**6.4**	**4.9**				
Median	8	9	4.5	12	7.5				

Taking first the individual voice features, we can see that all speakers received some descriptions of 'dry', and not being distinctive to any individual speakers, this is entirely possible. However, we can observe that Speaker 4 receives a

much higher attribution of 'dry' than the other speakers (highlighted in yellow). This pattern is repeated with the feature, 'whispery'. On the other hand, although all speakers receive attributions of 'sharp' and 'shrill', they are all low in frequency and almost of the same score. It is not clear whether this is noise in the data, present at a low level in most descriptions of voices, or a small feature shared by all five speakers. No matter, the scores are similarly low for all five speakers, suggesting that either all five speakers are distinct from the norm and/or they are not distinct from each other in these regards.

The category of 'effeminate' shows Speaker 2 to be distinct from Speakers 1, 4 and 5, and possibly 3. This is emphasised by the difference between Speaker 2's count and the mean for all speakers - nearly three times lower. Speaker 3 also shows some evidence of being felt to be effeminate, and it is interesting to note that both speakers received relatively high ratings on the 'high-pitched' bipolar scale. We can speculate that the two are related and this would make for interesting further research.

For the feature of 'rough', two scores of 19 and 13 stand out against the rest. 13 may be meaningful, although the mean is 9.4, whereas 19 can, in the context of these data, be interpreted as being likely to be more meaningful. 'Croaky' attracts a diversity of frequencies, from 0 to 16, with the scores of 16 and 13 perhaps being meaningful. 'Raspy' has a low and homogenous count. There may be something distinctive in the score of 8 for 'lisp' compared with almost 0 for other speakers. The category of 'stammer' shows great contrast between Speakers 1, 4 and 2, 3, 5, with low attributions for 1 and 4 and higher attributions for 2, 3 and 5. Speakers 1 and 3 were also given the highest ratings in the bipolar voice features for 'steady' and there may be a link here, whereas, the interquartile ranges for Speakers 2, 3 and 5 were usually in the two central points 3 to 5 and the higher attributions of 'stammer' may represent those respondents who responded in the lower 50%. For 'stutter', one score of 18 stands out at nearly three times the mean and higher than those ratings for other speakers. We should remember, however, that 'stammer' and 'stutter' are likely to be used synonymously.

If we analyse the data relative to each speaker, we can see not only which features seem to be prominent, but also identify how one voice may differ from another voice in terms of clusters of features, forming the basis for individual profiles.

Speaker 1

Scores that stand out as more prominent than others are that Speaker 1 has a 'rough' voice and maybe a little 'croaky'.

Speaker 2

Scores that stand out are that the speaker's voice may be effeminate, with maybe a lisp, croaky, and a slight stammer/stutter.

Speaker 3

There are, normatively speaking, not many scores that stand out in Speaker 3's profile: maybe effeminate and a stammer may be present.

Speaker 4

Three scores stand out prominently, compared with the rest for Speaker 4: The speaker's voice on the recording may be described as 'dry', 'whispery' and 'rough'.

Speaker 5

The scores for Speaker 5 are less distinct. Those features that might stand out could be a slightly 'croaky' voice and stammer.

5.8 Soundalikes

Included in Stage 2 was the invitation to respondents to identify anyone known personally to them or famous, that the speaker may sound like. Several questions arise: Do they identify similar people, if indicating someone well known? What features are salient to them and present in the 'soundalike' speakers that they identify?

The responses were collated and coded according to:

1 = Well-known, regionally identified

2 = Personal acquaintance, possible regional connection

3 = Unknown acquaintance, connection unknown

4 = Famous, non-regional connection

5.8.1 Speaker 1 – Wales

Respondents provided the following 'soundalikes':

Table 5.8.1 Soundalikes for Speaker 1 – Wales

Code	Soundalike	Frequency
1	Goldie Lookin' Chain (Rap group from Newport, South Wales)	2
1	Dirty Sanchez (South Wales 'stunt' tv show)	1
1	Kelly Jones from 'Stereophonics' (South Wales pop group)	1
2	A friend (Respondent is from South Wales)	1
2	My brother (Respondent is from South Wales)	1
3	A flatmate	1
3	My friend's boyfriend	1

For those coded as '1', the first three people on the list represent speakers known to have identifiably South Walian accents. Kelly Jones has an accent associated with the South Wales 'Valleys' – a former heavily industrialised area north of Cardiff – and this is likely to convey a perception of a more generic South Wales accent. However, the first two soundalikes on the list both have accents associated with the adjacent urban centres of Cardiff and Newport, which share an accent on a dialect continuum. It is highly likely that it is specifically this accent feature that the respondents are trying to convey in their selection, as Speaker 1 is from the urban centre of Cardiff and was recognised as such by many respondents in Studies 1 and 2.

For the soundalikes coded as '2', we can only speculate that this is a reference to the speaker's accent, as the soundalike is not known to the researcher. Of course, in the police context, it would be simpler to obtain contact with the person in question to establish how they spoke. For those coded as '3', the researcher is unable to obtain any further information, however, a similar

procedure as that followed for soundalikes coded '2' could be followed in the forensic context.

5.8.2 Speaker 2 – South West

Respondents provided the following soundalikes:

Table 5.8.2 Soundalikes for Speaker 2 – South West

Code	Soundalike	Frequency
1	Gareth from 'The Office' (British tv comedy)	4
1	Vickie Pollard (Bristol character from 'Little Britain')	1
1	Little Britain woman (Tv comedy)	1
1	A farmer	1
1	Goldie Lookin' Chain (Rap group from Newport, South Wales)	1
2	People from my hometown (Respondent was from South West)	1
2	A friend (Respondents are from the South West)	4
3	A friend of a friend	1
3	A flatmate's friend	1
3	Kate's friend	1

The regional identification data in the earlier sections confirm that most respondents believed the speaker to be from the South West of England, many of them form Bristol. For those respondents coded with '1', an obvious connection appears in that all the speakers identified are famous for their regional accents, in particular their realisation of the postvocalic [r] shibboleth that can so readily identify those southern English speakers that come from the West or South West of England. This accent is also readily stereotyped as a 'farmer's' accent an example of which also appears. These associations should come as no surprise to those familiar with the Bristol accent, which is regularly stigmatised and parodied in the media and popular culture for realising the rhotic [r] - an almost direct contrast with Labov's (1966) findings re: the status of [r] on the other side of the Atlantic, in New York.

We could also speculate whether the comparison of a man's voice with that of a girl is purely related to the pronunciation or perhaps to do with other features. In the non-polar voice descriptions, some respondents also articulated that they felt the speaker to have a moderately high and somewhat effeminate voice. Could the use of Vickie Pollard as a comparison be conveying information regarding other vocal features beyond accent?

The Goldie Lookin' Chain (known as 'GLC') reference is technically inaccurate, as they are from Newport, South Wales, and may represent a simple mistake in placing the speaker regionally, or a misunderstanding regarding the origins of GLC. There is, however, also a plausible phonetic explanation: The dialect continuum of Cardiff-Newport is, whilst identifiably Welsh, also heavily influenced historically (through immigration) by English from the South West of England. A phonetic feature that becomes more prominent as you head from Cardiff towards Newport and Bristol is the increased realisation of the diphthong [ai] as approximately [oi], a realisation particularly associated with rural England, and the West Country, especially. This realisation is particularly pronounced by members of GLC as an almost a comic caricature of themselves and their origins and it may well be this feature that leads to mistaken comparison with GLC.

For those soundalikes coded as '2' we can only speculate in this research as to the referent, however, as all were from the South West, it is likely that this is a reference to an accent primarily. Those coded as '3' would require further contact.

5.8.3 Speaker 3 – South East

Respondents provided the following soundalikes:

Table 5.8.3 Soundalikes for Speaker 3 – South East

Code	Soundalike	Frequency
1	David Beckham (English footballer)	36
1	Frank Lampard (English Footballer)	1
1	Mike Skinner (Rapper from 'The Streets' pop group)	2
2	My parents (from London)	1
4	Michael Owen (English footballer)	1

Relative to Speaker 1 and 2, there is a smaller range of suggested soundalikes. For Speaker 3, however, there is a large amount of agreement, as the use of David Beckham as a soundalike occurred 36 times. This is indeed a very high figure for a non-essential category within the questionnaire. But what is it in Speaker 3's voice that is so salient to the respondents? It would be

uncontroversial to say that Beckham, a famous footballer from just outside London (Essex) is well known for his strong regional, estuary English speech variety. And these are the very same features that have been identified with Speaker number 3, throughout both Stage 1 and Stage 2. However, there could be even more within the selection of Beckham as a soundalike: He is also well known for having a soft, high, slightly effeminate voice; and these are three features described by a number of respondents in the polar and non-polar voice descriptions. It would appear then that a soundalike can convey more than an intuition about an accent.

The choice of Frank Lampard may echo the same reasoning as for David Beckham. It may be that the respondent perceives Lampard to share similar, high-pitched estuary English tones; alternatively there may be a non-linguistic motivation behind this choice: both are high-profile England-playing footballers who may have been collapsed into the same stereotypic category. The choice of the relatively high-voiced Mike Skinner may be similar to the reasoning behind selecting Goldie Lookin Chain for Speaker 2. Mike Skinner's performing style has lead many people to believe that he is from London, when he is in fact from Birmingham in the Midlands, a feature only conveyed in his everyday speech. In the meantime, many who hear him performing continue to mistake him for someone from further south.

For the soundalike coded '2', it is the respondents themselves who indicated that their parents are from London. The appearance of the footballer, Michael Owen could be motivated by a number of reasons: Firstly, he could also be described as having a relatively high voice and a regional accent. However, he is not from London, rather he is from Cheshire, south of Liverpool. It could be then that either the respondent is focusing more on the voice characteristics than the accent, or alternatively, one of the few people who misidentified the provenance of Speaker 3 as being from the North West of England may also have invoked Michael Owen as their way of articulating their impressions. Alternatively, this may be a case of the respondent mistaking the identity of their referent, confusing, for example, one footballer for another.

5.8.4 Speaker 4 – South East (RP-like)

Respondents provided the following soundalikes:

Table 5.8.4 Soundalikes for Speaker 4 – RP-like

Code	Soundalike	Frequency
4	Stephen Fry (TV comedian/actor)	1
4	Jack Dee (TV stand-up comedian)	1
4	David Walliams (TV comedian from BBC comedy 'Little Britain')	1
4	Guy from 'Blur' (English pop group)	1

Speaker 4 did not produce a wide range of soundalikes, as can be seen in the table above. But we should ask if there is a common link between those four suggested soundalikes. Do they share any common speech characteristics that illustrate what the respondents are trying to convey?

In terms of accent, all four of them are from the south of the UK, and the south of England. One of them, Stephen Fry, has an accent that could be described as RP. Jack Dee does not speak with an RP accent although it is not strongly regional either. He is famous for his comic style of delivery, which is rather deep-voiced, monotonous and dead-pan. David Walliams is famous for portraying a range of comedy characters in his work life and so without asking for further details we are not sure which character this refers to. However, in real life, his manner of speech might be described as rather monotone and deep. Presumably, the 'guy from Blur' is the lead singer, Damon Albarn. If this is the case, his speaking voice may be described as rather deep, southern English and monotone. He has in the past been criticised for speaking 'mockney' - being 'posh' but attempting to speak with an affected 'cockney'-flavoured accent. The real life police context would of course afford greater opportunity to expand and clarify exactly which features the respondents are referring to. However we can infer that there is a general theme connecting the speaker of fairly 'well-spoken', monotone and understated delivery, characteristics that have surfaced in the data in previous sections, the bi-polar and non-bipolar voice descriptions, the 'regional'–'well spoken' scale and the speaker origin data.

5.8.5 Speaker 5 – Midlands

Respondents provided the following soundalikes:

Table 5.8.5 Soundalikes for Speaker 5 – Midlands

Code	Soundalike	Frequency
1	Frank Skinner (Comedian from near Birmingham)	1
4	Scott Mills (Radio 1 dj)	1

It would appear that those speakers with less 'regional' accents attracted fewer soundalikes. But what do the two voices suggested here tell us about the non-linguists' perceptions regarding the voices?

It would be uncontroversial to say that most people who are aware of Frank Skinner will also be aware of his association with the West Midlands, which is strongly evidenced in his very distinctive Black Country (the area immediately to the west of Birmingham) accent. We can speculate that it is this feature that is salient to the respondent that suggested this soundalike.

Scott Mills, however is not from the Midlands but has an accent that many might say is not strongly regional, other than to say it is southern English. Many may also describe him as having an 'average' to 'high' pitched voice, and it may be these non-overtly regional and slightly high voice that are the salient features noted by this respondent that suggested this soundalike.

If we look at the data from the previous sections, these two possible soundalikes are congruent with previous suggestions and the range of opinion. Some respondents, mostly from the Midlands were able to recognise the regional influence in Speaker 5's voice, even though it has been 'diluted' through living in the south of England and working as a lawyer. On the other hand, many respondents indicated that they thought the speaker was not strongly regional in his pronunciation. It was also noted that he had a fairly high-pitched voice.

5.9 Attributions of Guilt

Stage 1 sought to investigate the non-linguist cohorts' attributions of guilt for the crimes of assault and robbery and sexual assault. Stage 2 suggested another criminal context to the next cohort, the crime of fraud. Remembering that no real crime had been committed, that the voice donors had no idea at the time of how the recordings would be used, that they all used the same picture memory interview task and conveyed the same factual information using a verbal guise technique, it is not possible that speakers were covering up, lying or overcompensating. It would therefore not be possible for respondents to detect signs of 'leakage' in the stories presented by the speakers as they were telling no lies, simply narrating a picture memory task. Given this, theoretically, there should be no observable differences between the means for the attributions of guilt for each speaker. However, as we saw in Stage 1, there were indeed great differences in the attributions of guilt and there appears to be a link between other personal features and these attributions. Would the same be repeated with a new cohort and a different crime?

The question asked was:
(Read to the respondents by the research administrator) "These speakers were recorded during police investigations into the crime of fraud. One or none of the speakers may be guilty of the crime. Please mark on the sheet how you likely you feel it is that each speaker may have committed the crime."

Figure 5.9.1.1 Distribution of attributions of guilt – Speaker 1

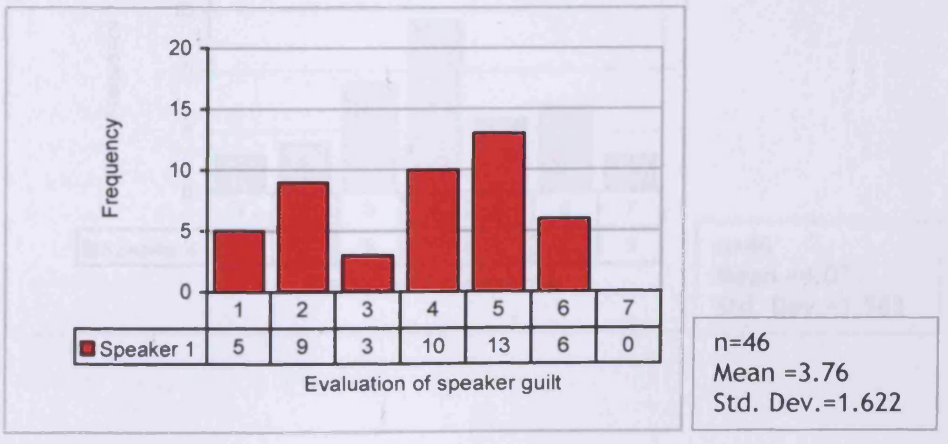
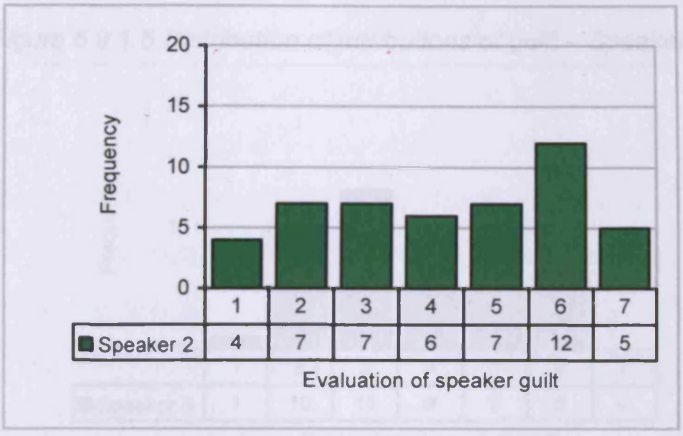
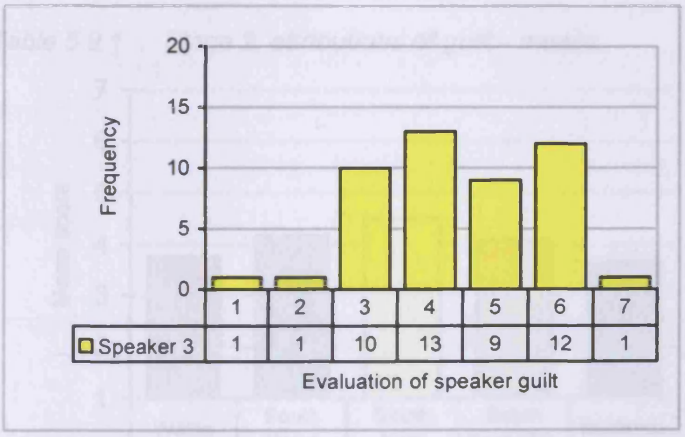


Figure 5.9.1.2 Distribution of attributions of guilt – Speaker 2



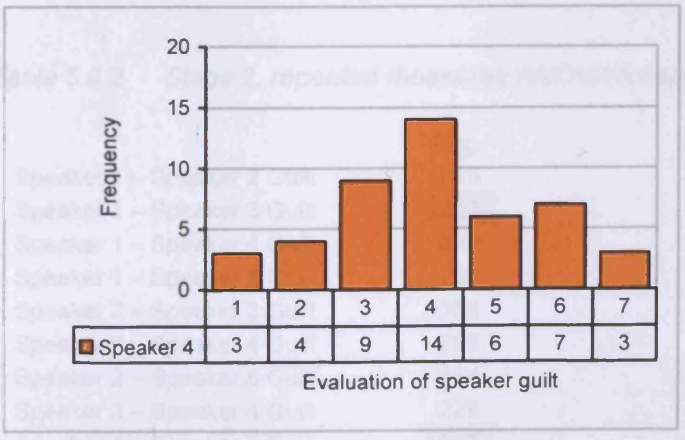
n=46
Mean =4.27
Std. Dev.=1.888

Figure 5.9.1.3 Distribution of attributions of guilt – Speaker 3



n=46
Mean =4.07
Std. Dev.=1.583

Figure 5.9.1.4 Distribution of attributions of guilt – Speaker 4



n=46
Mean =4.07
Std. Dev.=1.583

Figure 5.9.1.5 Distribution of attributions of guilt – Speaker 5

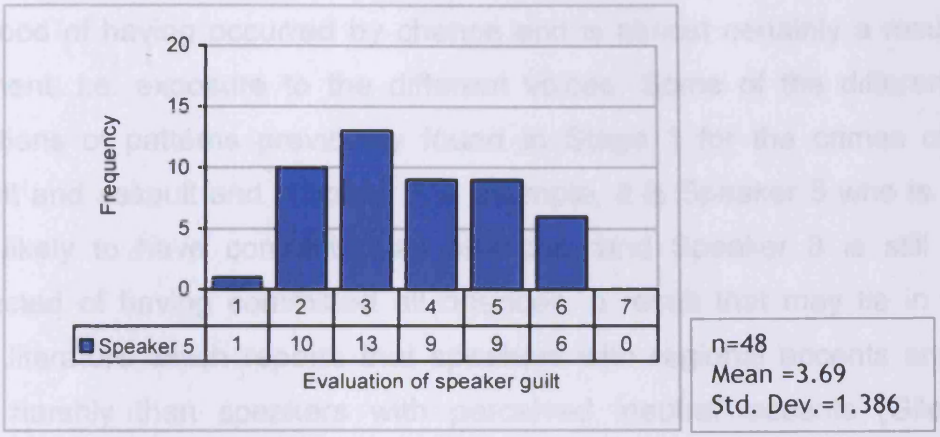


Table 5.9.1 Stage 2, attributions of guilt - means

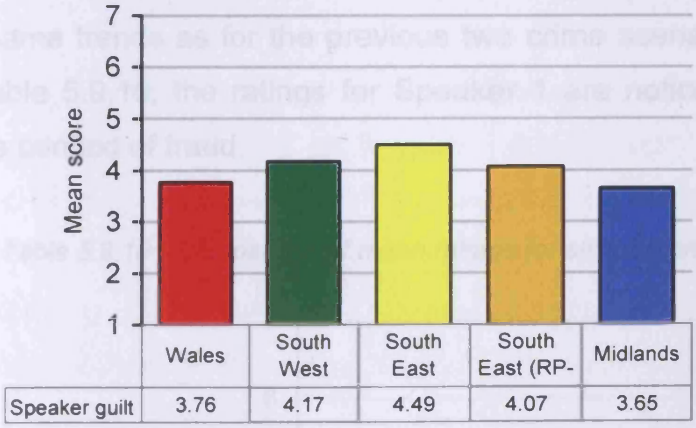


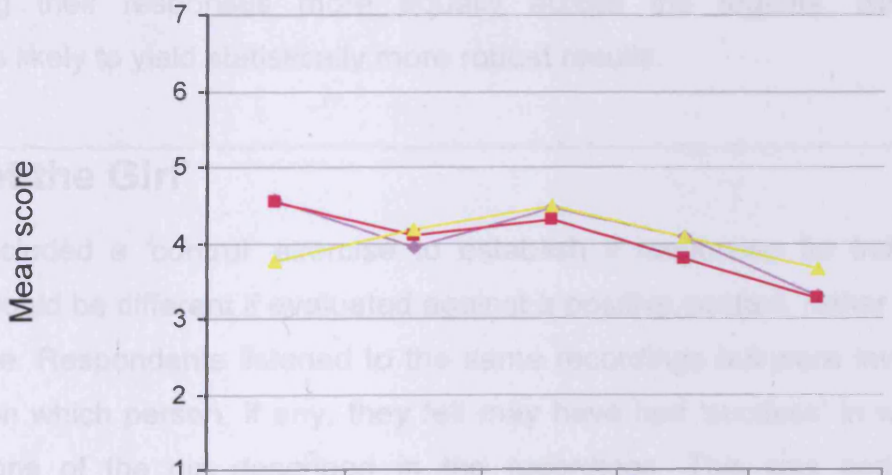
Table 5.9.2 Stage 2, repeated measures ANOVA for speaker guilt

	Sig.
Speaker 1 – Speaker 2 Guilt	.313
Speaker 1 – Speaker 3 Guilt	.021
Speaker 1 – Speaker 4 Guilt	.398
Speaker 1 – Speaker 5 Guilt	.755
Speaker 2 – Speaker 3 Guilt	.366
Speaker 2 – Speaker 4 Guilt	.795
Speaker 2 – Speaker 5 Guilt	.124
Speaker 3 – Speaker 4 Guilt	.229
Speaker 3 – Speaker 5 Guilt	.005
Speaker 4 – Speaker 5 Guilt	.249

We can see from Tables 5.9.1 and 5.9.2 that significant differences emerge between the means for Speakers 1, 3 and 5, a pattern that has a very small likelihood of having occurred by chance and is almost certainly a result of the treatment, i.e. exposure to the different voices. Some of the differences are repetitions of patterns previously found in Stage 1 for the crimes of sexual assault and assault and robbery. For example, it is Speaker 5 who is seen as least likely to have committed all offences, and Speaker 3 is still strongly suspected of having committed all offences, a result that may tie in with the wider literature which reports that speakers with regional accents are judged more harshly than speakers with perceived 'neutral' accents (Giles 1970; Bishop et al. 2005).

What is anomalous here is that whilst the ratings for all other speakers follow the same trends as for the previous two crime scenarios in Stage 1, illustrated in Table 5.9.10, the ratings for Speaker 1 are noticeably more positive in the crime context of fraud.

Table 5.9.10 Comparison of mean ratings for all three crime scenarios



	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Sexual assault (n=68)	4.55	3.93	4.46	4.07	3.28
Assault & robbery (n=76)	4.53	4.09	4.29	3.79	3.26
Fraud (n=47)	3.76	4.17	4.49	4.07	3.65

As these data are collected from three separate cohorts, when cannot statistically pursue this occurrence and can only speculate as to what may be causing this more positive rating for the Cardiff speaker: first, the anomaly may be an artefact of a smaller number of respondents for the questions of guilt in Stage 2 (n = 47); second it may due to a higher proportion of respondents from Wales in Stage 2, who have provided more positive evaluations of the speaker; third, it may be due to a change in the way that Cardiff and Wales is viewed in the Welsh and wider UK context in the intervening year between Stage 1 and Stage 2; fourth, and inline with previous research on accent and crime, it may be a result of the change in crime context and the fact that the speaker's voice and or accent do not provoke the same stereotypic associations with the crime of fraud as they did with the more violent crimes.

Finally, it may be worth noting that the question as to whether there were any differences between the attributions of guilt from different regions towards speakers from their own and other regions merits further investigation: Future research should consider obtaining a larger number of respondents than $n = 45$, sub-dividing their responses more equally across the regions. Such an approach is likely to yield statistically more robust results.

5.10 'Got the Girl'

Stage 2 included a 'control' exercise to establish if responses for individual speakers would be different if evaluated against a positive context, rather than a criminal one. Respondents listened to the same recordings but were invited to comment on which person, if any, they felt may have had 'success' in winning the affections of the girl described in the recordings. This was coded for simplicity as 'got the girl'. Their scores were marked on the same design of seven-point bipolar scale used throughout the two Studies. '1' = 'got the girl' and '7' = 'didn't get the girl'

The data from the respondents suggest that there is some difference between the two scenarios although again they must be handled with caution: the number of respondents is still relatively small (n= 62).

Figure 5.10.1 Distribution of evaluations for 'Got the girl' – Speaker 1

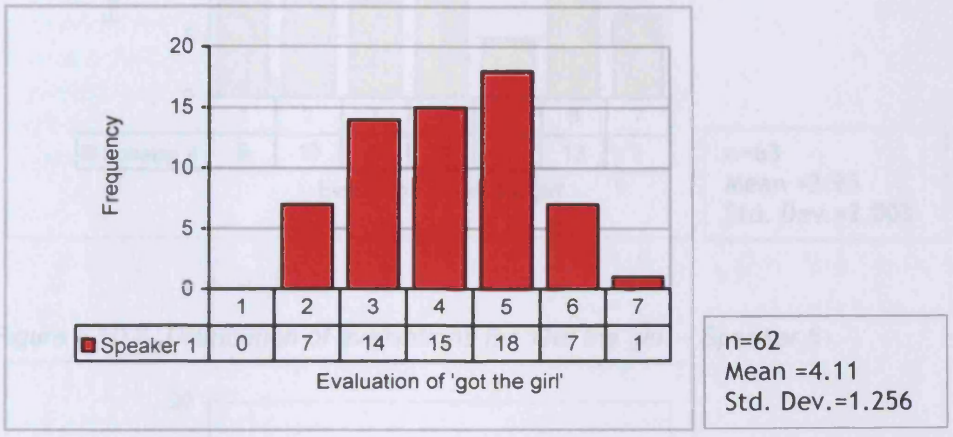


Figure 5.10.2 Distribution of evaluations for 'Got the girl' – Speaker 2

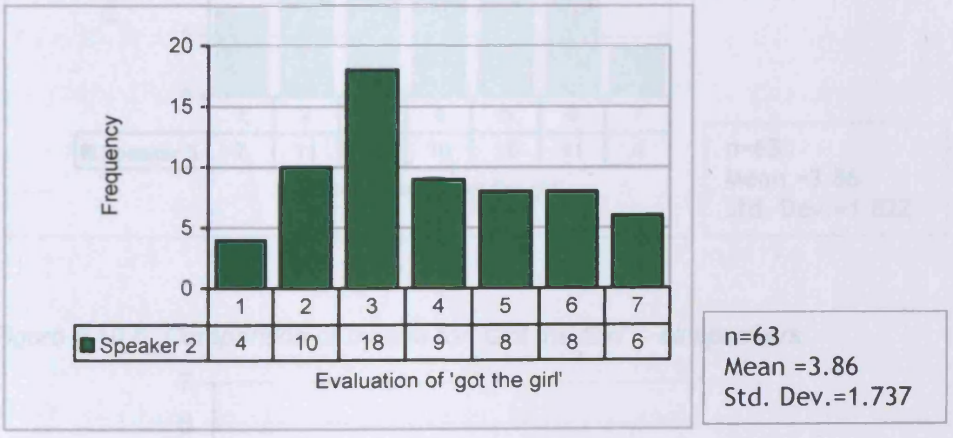


Figure 5.10.3 Distribution of evaluations for 'Got the girl' – Speaker 3

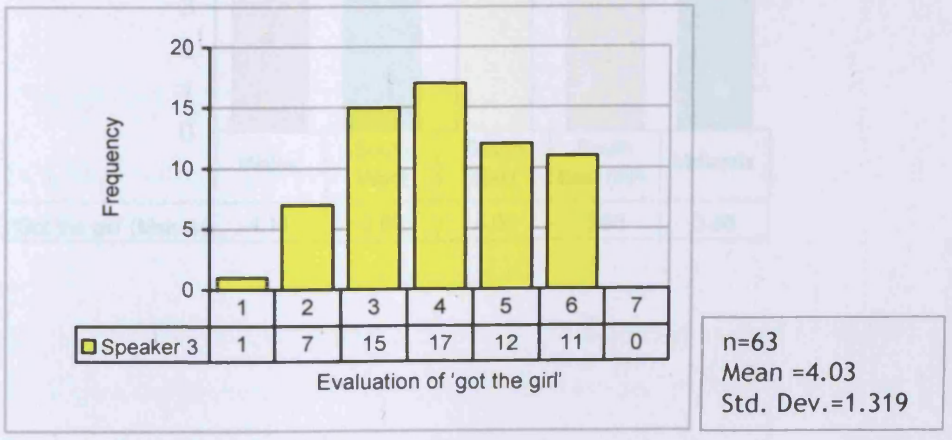


Figure 5.10.4 Distribution of evaluations for 'Got the girl' – Speaker 4

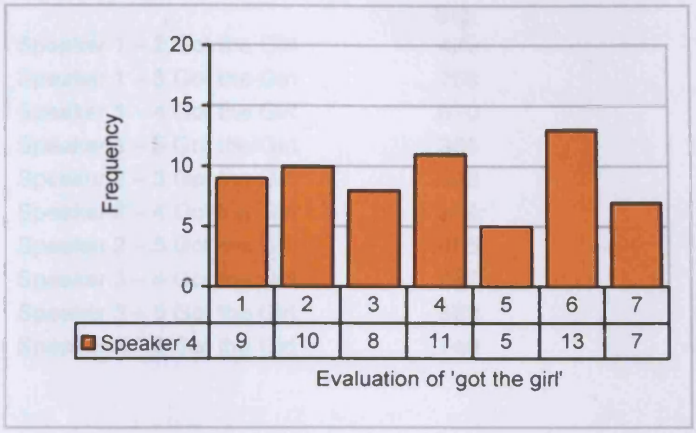


Figure 5.10.5 Distribution of evaluations for 'Got the girl' – Speaker 5

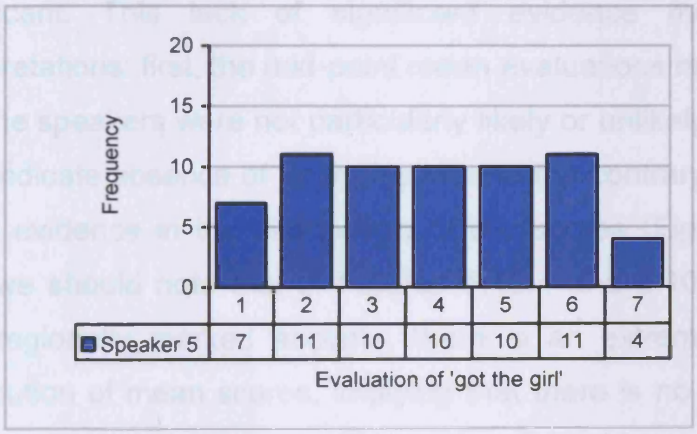


Figure 5.10.6 Comparison of means for 'Got the Girl' – all speakers

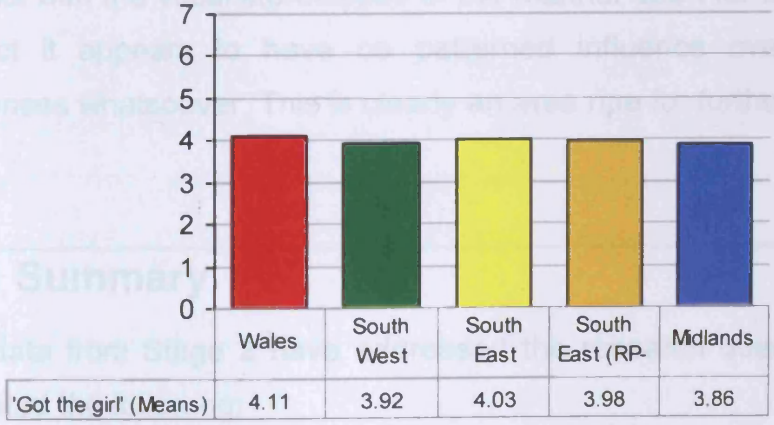


Table 5.10.1 Repeated measures ANOVA for 'Got the girl'

	Sig.
Speaker 1 – 2 Got the Girl	.479
Speaker 1 – 3 Got the Girl	.758
Speaker 1 – 4 Got the Girl	.670
Speaker 1 – 5 Got the Girl	.365
Speaker 2 – 3 Got the Girl	.673
Speaker 2 – 4 Got the Girl	.864
Speaker 2 – 5 Got the Girl	.855
Speaker 3 – 4 Got the Girl	.887
Speaker 3 – 5 Got the Girl	.529
Speaker 4 – 5 Got the Girl	.740

We can observe from Figure 5.10.6 and Table 5.10.1 that the means are all around the mid-point of 4 and that statistically, any small differences are not significant. This lack of significant evidence may have a number of interpretations: first, the mid-point mean evaluations may indicate strong opinion that the speakers were not particularly likely or unlikely to have 'got the girl', or it may indicate absence of strong opinion to the contrary. However, whilst there is some evidence in the distribution of the scores (Figures 5.10.1-3) to support this, we should note that in Figures 5.10.4 and 5.10.5, the speakers with the less regionally marked accents, there is an extremely wide and almost flat distribution of mean scores, implying that there is no overall consensus on the speakers' romantic chances. This makes it hard to accept the first interpretation exclusively. The alternative explanation is that the romantic context does not interact with the vocal stereotypes in the manner seen for the criminal contexts; in fact it appears to have no patterned influence over the non-linguists' responses whatsoever. This is clearly an area ripe for further investigation.

5.11 Summary

The data from Stage 2 have addressed the research questions in 2.14.2 and indicated the following:

1/. In terms of what factors may influence the estimations of speaker origin, it is likely that guided perceptual mapping has had an influence in that there was increased accuracy of the Welsh and South West regional speakers' accents and an increase in the estimation of Speaker 3 being from London and the

South East. Moreover, the inclusion in the guided perceptual mapping of the cities of Oxford and Cambridge has diverted previously inaccurate estimations of Speakers 4 and 5 being form London and the South towards a location that can betoken 'well spoken' speech. Other patterns regarding the salience of accents and the identification accuracy of each region are largely consistent with the findings of Stage 1.

As an alternative or accompaniment to guided perceptual mapping, the provision of a 'strength of accent scale' for the description of the speakers' accents appears to elicit significantly different evaluations for each speaker. Respondents appear to evaluate the speakers along lines which broadly coincide with the evaluations in perceptual mapping, suggesting that the scale may be tapping into non-linguists' conceptions of language variation and ideology in the UK.

2/. When provided with an audiofit, non-linguists appear to utilise it to make distinctions between respondents, but only on specific dimensions. There are, however, distinct patterns in the evaluations provided on the scales, suggesting that the terms are meaningful and of assistance in articulating perceptions of voice features. The non-polar items have a lower uptake but nevertheless display patterns indicating that they may be of use for describing voice features.

3/. In terms of alternative ways of articulating their perceptions of voices, the use of soundalikes has shown that, depending on the voice and available people for comparison, it may be possible to convey perceptions of a voice without having to engage with any technical voice descriptions.

4/. With respect to attributions of guilt and the 'got the girl' control, the data suggest that respondents differentiate between speakers when assessing the crime presented and the likelihood of speakers being guilty. They did not, however, evaluate the speakers significantly differently when presented with the non-crime, 'got the girl' scenario.

6 - DISCUSSION

6.1 Introduction

The chapter discusses the results in the following order: Dealing first with the research questions in 2.11.2, 6.1 and 6.2 discuss the estimations of speaker origin and age, with 6.1 addressing the estimations of age which were asked only of respondents in Study 1; 6.2 deals with estimations of speaker origin, and is broken down into three sections: the first section deals with the estimations of the origins of Speakers 1, 2 and 3, who were identified by the respondents as being from specific British regions; the next section deals with Speaker 4, the RP-like speaker, who was identified as 'well-spoken' by the respondents; the final section deals with Speaker 5, who was identified by most respondents as having 'no accent'. The discussion of these five accents and the estimations of their origin is centred on exploring factors that may account for the choices that the respondents made. As van Bezooijen and Gooskens (1999) have argued, it is important to know to what extent naïve listeners are able to identify language varieties and also on what cues their identification is based.

Section 6.3 deals with the research questions regarding the evaluation of the speaker voice features: the audiofit voice descriptions, and in particular the responses in Stage 2 to the bipolar and non-polar adjectives presented in an audiofit format. The Discussion concludes with Section 6.4, discussing the evaluations of speakers by reviewing the role of language attitudes and non-linguist evaluations of the five speakers.

6.2 Estimations of speaker age

"How well do respondents recognise and estimate the ages of unknown speakers with selected British regional and social accents? What factors may influence their estimations?"

Following the evidence from Bull and Clifford (1984), Huntley et al. (1987) and Braun (1996) presented in Chapter 2 (p.108), regarding non-linguists estimations of age, the use of under 16 and over 40 year-old respondents was disfavoured. This may be a contributory factor to the fact that almost all age estimations were at the correct end of the spectrum with very few respondents over-estimating the speakers' ages by more than one age span. But there are other considerations that may account for the relative accuracy of the respondents' age estimations. Their accuracy may simply be a reflection of the respondents basing their estimations around their own ages (all between 18 and 21), as was found by Huntley et al. (1987). Alternatively, the respondents may have been influenced by certain assumptions made regarding the lifestyles of the speakers - all five are apparently single, and out consuming alcohol with friends. This is a pattern of behaviour most likely linked to a group of younger men than older speakers, contributing perhaps to the virtual unanimity amongst the respondents in their estimations that the speakers were at the younger end of the spectrum.

However, this is perhaps insufficient explanation for the considerable accuracy in the selection of the specific age band, with Speakers 1, 3 and 5 all having their ages correctly guessed as 26-35 by 51%, 76% and 52% of respondents, respectively. Was this simply the results of respondents picking a likely age band for the 'group' and generalising it onto all heard voices? This seems unlikely as Speaker 2, was estimated by 77% of respondents as belonging to the 18-25 age group, younger than Speakers 1, 3 and 5. This suggests that although respondents' estimations were at the low end of the spectrum, they did nevertheless differentiate between speakers on the basis of age and not simply generalise a group age onto all members. Moreover, the differentiation between speakers according to their age suggests that it is unlikely the age estimations are predicted solely by the age of the respondents themselves.

An interesting point worth noting is the apparent lack of relationship between voice pitch, age estimation and accuracy of age estimation. The two speakers that received the evaluations of having the highest pitched voices, Speakers 2 and 5, were seen respectively as the youngest and the oldest speakers. The youngest voice donor in real life, Speaker 4, aged 21 was seen as having the deepest voice and was estimated by 59% of respondents to be 26+. This is an area that could support further research, until which point, researchers should exercise extreme caution in assuming any relationship between voice pitch and age estimation.

Finally, there exist problems regarding the system and estimator variables: in terms of system variables, whilst respondents were broadly accurate in their estimation of ages, there is a problem with the age spans provided. Braun (1996) recommends the use of age spans, although she provides no specifics for numerical divisions; thus, the researcher-derived decision was taken to provide age bins from 18 (the legal drinking age in the UK)-25, and thereafter in ten-year spans from 26-35 as it was felt that these would be cognitively light. The problem with this decision is that if the speaker's age is 25, and the respondent estimates '24', they are one year away and judged to be correct, whereas if they estimate '26', they are one year away but judged to be incorrect. Moreover, if someone estimates '26' or '35', whether they are one year or nine years away, in terms of age bins, they are judged to be equally inaccurate. Such anomalies compromise the validity of the exercise. With regard to estimator variables, the data may be telling us that respondents simply were not able to consistently identify speaker ages to within a ten-year time frame. Future researchers may consider addressing these issues by experimenting with:

i/. a different categorisation of ages bands, such as 18-30, 30-40, etc.
ii/. wider or narrower bands for perceptual discretion – ten years may simply be too narrow for reliable estimations, e.g. 18-35
iii/. the use of lexical labels for age groups in preference to numbers, such as 'young man', 'middle aged man', etc.
iv/. allowing free choice for age estimations

6.3 Estimations of speaker origin (Stages 1 and 2)

"How well do respondents recognise and estimate the linguistic backgrounds of unknown speakers with selected British regional and social accents? What factors may influence their estimations?"

One of the main aims of this research was to investigate non-linguist respondents' ability to recognise regional and social accents in the UK. Preston (1996) has indicated in the US context that, to a large extent, non-linguist respondents are able to differentiate between voices with regional phonology but with no further lexical or grammatical cues. Both stages of the present research revealed a large degree of accuracy in identifying the origin of the accents used, in addition to uncovering new data regarding how respondents articulate their estimations in the UK.

In order to discover more about community-salient referents, the methodological choice was made not only to conduct one study into the estimations of respondents with regard to speaker origin, but also to conduct a second Stage, with the same voices but new respondents and methodological variations, such as the use of a perceptual map based on regional divisions suggested by respondents in Stage 1. As the research has been dealing with both geographical and social accent variation, the accents employed are discussed in separate sections. Section 6.3.1 deals first with the estimations of regional accents (Speaker 1, 2 and 3); 6.3.2 deals with any further factors related to estimating the origins of the RP speaker (Speaker 4); and finally, Section 6.3.3 focuses on any outstanding factors that relate to the estimation of the 'no accent' (Speaker 5).

6.3.1 Estimations of the origins of traditional regional accents

The voice donors from Cardiff, Bristol and Hastings were recruited on the basis of their regional accents, as judged by the researcher. As a linguist I felt that the regional origins of these three speakers would be easily located by any professional linguist. However, a fundamental aim of this research was to investigate *non-linguists'* perceptions and conceptions and the accuracy of their

estimations of regional origin. The focus of this discussion will be the non-linguist respondents and the factors that may have influenced their estimations of speaker origin.

A departure point for reviewing non-linguist conceptions of these regional accents is the composite 'strength of accent' scale identified in Stage 2, section 5.4, in which we are able to observe that the non-linguists themselves firmly believed all three speakers to be producing regional accents: For Speaker 1, 91% of respondents placed his accent left of the 'neutral' mid-point of '4' on the strength of accent scale; for Speaker 2, the figure was 83% and for Speaker 3, 88%. We can also see from the ANOVAs in 5.4.6, that there is no significant difference between the regional evaluations for Speakers 1 and 3, and for Speakers 2 and 3, indicating that in the view of the non-linguists, the speakers were very similar in their overall level of 'regionality'.

Turning to the specific locations that the non-linguists selected, the data from Stage 1 suggested that there was a high degree of accuracy in the non-linguists' estimations of regional origin for Speakers 1, 2 and 3, from Wales, the South West of England and the South East of England. It would appear that the majority of respondents were accurate in their identifications for these regional speakers, at a very similar level of accuracy for each voice (66-71%), the majority of respondents. This is a high figure when we consider that the respondents were given no clues as to the likely regional origin of the speakers, that there are many other British regions (or possibly international native speaker regions) that the voices could have been drawn from, that isoglosses are not fixed determinants that predict speaker's accent, and finally, that the isoglosses employed to divide the regions were idealised. It is fair to conclude, therefore, that the answer to the question, 'how well do a group of non-linguists recognise and estimate the origins of speakers' accents?' must be sufficiently meaningful to non-linguists to warrant further enquiry. These same high recognition rates were seen in Stage 2, using a separate cohort of respondents, where an even higher 71% of respondents correctly identified the region of origin for all three traditional regional accents. Across the two Stages, ten factors may have mediated the identification process, explaining the accurate

and inaccurate estimations. Some of these fall into the category of *system variables* and others, *estimator variables,* as discussed in Chapter 2, a division which will be referred to later in this chapter.

1/. System variables: Laboratory conditions

The present research was conducted in laboratory conditions and represents, in terms of the regional accents used, a best case scenario, for the purposes of theoretical investigation (see Section 2.3.1, p.42) The data in this research may, therefore, show us the theoretical upper limits of non-linguist potential given highly marked accents.

2/. System variables: Geographical scale

The present study was designed to include speakers from within the Cardiff University catchment area - adjacent regional areas to Wales and Cardiff - rather than British regions further afield. With almost all respondents being drawn from these nearby regions, the data indicate that in the southern British context, respondents are largely familiar with the speech from their own and nearby regions; that is, nearby accents are available and salient. This is worth highlighting as it also relates to the diversity of regional accents in the UK. In terms of scale, unlike many of the former large British colonies where English has gone on to become the majority language, such as Australia, New Zealand, Canada, etc, one does not need to travel any great distance to encounter a form of speech that is distinctive to a nearby region and identifiable not only by those who live in the region but by outsiders also. The concept of a region and diversity of regional accent are in the UK context therefore, on *microscale* in relation to the context in the larger, 'new world' countries. In forensic terms, this may be advantageous when seeking to identify the origin of a speaker, as they may be located to within, for example, a 50-100 miles radius, rather than a 500-1000 miles radius. Indeed, the British microscale may be a facilitating factor to the current research and to this line of police investigation, thanks to the smaller distances involved in noting sociolinguistic variations. The converse, however, may also act as a practical deterrent to such investigation in countries on a larger scale.

3/.System variables: Selection of speakers from focused communities

Returning to the reasons for the high 70% recognition rates for all three regional speakers, it may be that the voice donors, or in the forensic context, the suspects, have been taken from *focused* (Le Page 1980), socially stable linguistic communities, and that they are seen to be highly typical, perhaps even stereotypical of their respective regions, and therefore easier to identify. The voice donors for this study were not selected at random, but rather they were chosen with the strength of their accents in mind. We must therefore remember to qualify any identifications made by respondents, by acknowledging that this research did not establish how linguistically focused the regions were beforehand; it is highly likely that a large number of speakers will have less marked or salient accents, with some showing bi or multi-dialectal or lingual features.

Indeed, it is not known to what extent the speakers living in any one area use the traditional speech variety related to that region. That is, per 1000 people in one city or village, for example, it is not known how many speakers are full regional dialect speakers, how many have shades of the dialect, code switching according to social context, and how many do not share the regional features or are not from the region at all. A representative cross-section of speakers from a community may therefore be substantially more difficult for our respondents to identify, and we must not assume a mutual relationship between the speakers of a regional speech variety and all speakers from regions using that speech variety.

4/. Estimator variables: Respondent region

Observing the data for respondents who correctly identified the speakers' regions of origin reveals a strikingly influential factor in the respondents' estimations – their own region of origin. The data in Sections 4.3 and 5.3 indicate that there are significant associations between speaker and respondent origin when identifying Speaker 1, 3 and 4, with Welsh respondents most frequently identifying Speaker 1 being from Wales, South East respondents identifying Speaker 3 being from the South East (also Midlands respondents' in Stage 2) and South East respondents identifying Speaker 4 as being from the

South/South East/London, This is an association between respondent and speaker origin which was also observed by Williams et al. (1999) amongst Welsh teenagers identifying speakers from the regions within Wales and aligns with previous research on the evaluations of non-community speakers (Williams et al. 1996; Niedzielski and Preston 1999; van Bezooijen and Ytsma 1999; Kerswill and Williams 2002; Garrett et al. 2004). The data suggest that this perceptual asymmetry is in part a function of geographical proximity and that sociolinguistic dislocation from a speaker's region may in some cases impair non-linguists' ability to recognise that regional accent, with, for example, respondents from the South East being weakest at identifying Speaker 1's Welshness and reciprocally, the Welsh respondents least often identifying Speaker 3 and Speaker 4 as being from the South East.

However, we should be cautious of concluding that success in identifying a speaker's origins is related solely to respondent origin. That is not the whole story. The data show that whilst speaker accent and region is not symmetrically salient, this is not simply a matter of Euclidian distance. We must consider social and perceptual space alongside physical space (Britain 2002:609). On the one hand knowledge of other varieties will be mediated by physical features such as mountains, marshes, motorways, lack of roads or public transport; however, other social factors will also be relevant, such as political subdivisions and history. In Wales, for example, there is a separate language, a Welsh Assembly Government with its own legislative franchise, and significant cultural memory that maintains negative sentiment towards English colonisation and unwanted English governance from London very much apparent in popular Welsh discourse. Such social divisions are further maintained through the political and media infrastructure, Wales having its own Welsh language television service. Social and perceptual distance may therefore also be cited as a relevant predictor of accent salience.

Likewise, social factors may explain not only absence of symmetrical salience, but also presence of symmetrical salience. For example, across the two Stages, Speaker 2 is identified as being from the South West by high frequencies of respondents from all regions, but respondent region is never significantly

associated with correctly estimating the speaker origin. This suggests that the Bristolian voice donor is equally salient to respondents from the South West as he is to respondents from other regions and that his Bristolian accent is somehow symmetrically salient across the southern half of the UK, in a way that the South East or Wales accents are not. There may be many explanations for this but clearly Euclidian distance is not the only one. The cause may be social in that in modern Britain, the stereotypical South West accent may have hurdled geographical distances to be become symmetrically salient thanks to national media coverage and popular social stereotypes associated with speakers of with a rhotic South West, such as the popular Bristolian tv character Vickie Pollard from the comedy show, Little Britain (also referred to in the Soundalikes section (5.8)).

The implication for an audiofit is that what is required is an appreciation of the lack of perceptual symmetry across the country and the need for a localised aspect to any elicitation of accent description. Linguists researching this area must be cautious in the assumptions they make with regard to non-linguists' awareness of any linguistic diversity that may not be immediately around them. Language variation is asymmetrically embedded at a local level and we need to explore community-salient differentiations and labelling of local and national accents. The non-linguist police as interviewers and data gatherers need to be aware that what is available (Preston 1996) to one non-linguist officer is not necessarily available to the non-linguist earwitness, and what is available in one community is not necessary available in another.

5/. <u>System variables: The elicitation framework</u>

In Stage 1 of the study, where open questions were used to elicit estimations of speaker origin, the recognition rates were 66% for Wales, 67% for South West and 69% for London/South East (excluding 'South' for the sake of equal comparison between Stages). In Stage 2, guided perceptual mapping and a perceptual list were employed (see 2.11.1).The utility of this framework was evident on the respondents' complete questionnaires: firstly, respondents completed both the list and the map, and second, that there was no evidence of contradiction between the two, one appearing to support the other. The data for

Stage 2, the respondents using the perceptual map, indicate a slight increase in the recognition rates for Speakers 1 and 2, and all three speakers receive above 70% accurate estimations of origin.

Given that, as discussed above, that the Welsh respondents' recognition of or areal taxonomy of English dialects appears to be weaker than other respondents, it may be that the group that benefitted most from the use of perceptual mapping is the Welsh respondents, and that the provision of a perceptual map supports cases of weak salience. Likewise, the opposite effect may be seen for the English respondents, where the responses from the English improved to the point where respondent region was not significant. Naturally, any observed differences between recognition rates in Stages 1 and 2 need to be treated with caution as they have been drawn from separate cohorts in each study. We may, however, infer that the provision of some form of guided perceptual mapping may assist in disambiguating the provenance of other accents and that the use of perceptual mapping in some ways reduces the perceptual asymmetry of the respondents from different regions.

6/. Estimator variables: linguistic experience and perceptual incongruity. Considering the respondents on a more individual basis, we must consider the role that both non-linguists' geo-spatial awareness and linguistic experience have in the recognition and description of speakers from other regions. How well does each non-linguist know the region from which the speaker is drawn? Do they have personal experience of visiting this region or meeting people who identified themselves as being from this region? Does the sociolinguistic maturity and life experience of the judges (Kerswill and Williams 2002) inform their predictions of what speakers from this region sound like? An example may be that someone living in the South East of England may have heard 'a Welsh accent' in the media, but have very little experience of Wales or its speakers. Our South East respondent is likely to have expectations as to what a Welsh person sounds like but their mental model of a Welsh accent may be rather one dimensional and based on only a few examples and little personal experience. When encountering Speaker 1, who was identified by many (especially Welsh) respondents as using the distinctive Cardiff accent, it may well be that this

Cardiff variant of the Welsh accent does not align with the respondents' expectations of a Welsh person's accent, as they are not familiar with the linguistic diversity of Wales. As a result, of this *perceptual incongruity* the respondent either misidentifies Speaker 1's region or fails to provide any suggestion at all, or makes a misidentification that other speakers or hearers with similar linguistic experience might make. Unmet linguistic expectations are therefore important aspects of salience and availability and merit further investigation.

7/. Estimator variables: Speaker ability/sensitivity

We must remember the diversity of the adult populations and not assume homogeneity amongst our non-linguist respondents, acknowledging that there may have been variations in respondent speech sensitivity. In the same way that some people have an aptitude for learning languages and some for impersonating accents, others simply do not appear to have an ear for recognising accents. Pahn and Pahn (1991) demonstrated some between-speaker variation in ability to perceive accents, and in the discussions that came up with respondents in the post-questionnaire de-brief, a significant minority, in both Stages, remarked that they are 'not good at recognising accents'. This should be seen as part of the natural patchwork of human strengths, skills and weaknesses, the implication being that in forensic terms, speaker description may not be productive or reliable for some individuals.

Alternatively, other respondents may pick up on selected parts of the speech rather than the whole of an utterance, noting the sounds that are for them emblematic of the speech of a particular social group. In this case, respondents may mistakenly assign the speaker to a particular reference group (see 2.10.7.1 for further details).

8/. Exposure-recognition

To date, no research has been conducted on the relationship between exposure to an accent and the amount of text required for recognition of the regional origin. The question is: did the respondents have sufficient text to listen to or sufficient exposure to the text to recognise the accent? Would the recognition

rates have improved if they had had another 30 seconds or a minute of monologue, or five repetitions? Would those respondents that provided region-level identifications have been able to provide urban level estimations? Further research is needed to establish if the recognition of accents has an optimal exposure time.

9. Estimator variables: non-linguist language ideology

The non-linguist belief that there *is* a mutual relationship between speaker and regional variety may distort the results seen. For example, consider non-linguists' awareness of the internal linguistic diversity of other regions: We know that the identification and application of labels to regional varieties is a simplification of the complexity of language variation. Regions are not linguistically internally homogenous, and within one region there are many dialect sub-strata, with each region subject to its own internal regional variations; Yet, for non-linguists, the *matryoshka paradigm* mentioned in 2.12.5 persists and linguistic homogeneity within a region is a common non-linguist conception of language variation; people will talk of a Welsh accent, a West Country accent, etc, without concerning themselves with that within that region there may be considerable local variations.

The belief that all speakers from one region speak the same way, a possible example of *outgroup homogeneity bias* (Section 2.12.6), may have lead respondents to be more strident in their evaluations once they felt they have detected enough clues. Whether this led them to make more or less accurate evaluations is a matter for future research.

10/. Estimator variables: Salient locators - regional vs. urban-level identification

Earlier in this section, the discussion centred on the evaluations of speaker origin in terms of speaker region, but what of speaker city/town? What we see in the data from the estimations by city/town is that a much smaller number of respondents used city or town as the salient locator, with more than 50% of non-linguists not offering city-specific estimations. Aside, of course, from the non-linguist's knowledge of the area in question, speaker sensitivity, sociolinguistic experience or training, what factors may have mediated the

selection of an urban or regional label as the salient locator, and what factors shaped the level of global or specific detail (Preston 1996) that respondents provided? A number of not necessarily mutually exclusive possibilities arise:

i/. *Size of urban area*: the speaker may not be from an urban area with any overt linguistic status or prominence. Cardiff and Bristol are, in the UK context, substantial urban centres, regional capitals of great regional prominence and sizeable populations, and at least four times the size of Hastings, the origin of Speaker 3. Some urban areas will simply not be *available* (Preston 1996) possessing neither sufficient national profile or identity to be viewed as linguistically distinct, blurring into the linguistic landscape.

ii/. *Vacuum of prominent urban centres*: it may be that in one particular region, there simply are no substantial urban areas of any distinct salience. This may well be the case for the South East, where Hastings is situated. Across Kent and East Sussex, there is not one urban area with a population greater than 100,000.

iii/. *Asymmetrical status/prominence*: although there does not appear to be any evidence of this in the current research, it should be remembered that in certain areas, the urban centres within the region may have more overt linguistic status or prominence than the region that surrounds them. For example, one may not be surprised to hear the label 'Liverpool' being used in preference to 'Merseyside', or 'Newcastle' being used in preference to Northumberland'.

iv/. *Region-specific linguistic phenomena*: Although not one accurate urban evaluation was elicited for Speaker 3 across the two Stages, at least one urban-level evaluation was provided, and the majority of respondents were in agreement as to the origin of the Speaker: across the two Stages, 52% and 62% of non-linguist judges estimated the speaker as having come from London. The rise of estuary English has lead to a change in the linguistic landscape around London. In the past two generations, or so, as

a result of commuting into and migrating out of the city, the speech of London has rapidly spread or influenced the varieties of the regions immediately surrounding it, virtually 'merging' the accents of speakers from both areas with the resultant 'variety' commonly identified as estuary English (Rosewarne 1984). This *dialect levelling* has been described by Kerswill and Williams (2002) as a process of loss of regional dialects, resulting in compromise new varieties, with a geographical spread that is larger than the original dialect, in some cases functioning as regional standards. The 21st century speech of the South East of England is, as a result of this process, largely homogenous to the ears of the non-linguist (from within and without the region) with few specific urban differentiations to be observed. Moreover, speakers from the South East may process Speaker 3's speech, as simply one point along a spectrum that leads to strongly identified London speech rather than a similar but distinct way of speaking. If this is the case, there may be no conceptual isogloss in the minds of South East respondents that enables them to readily distinguish estuary speakers from London speakers. It is interesting to note that similar tendencies have been observed – to a lesser extent – amongst at least 50% of respondents from other regions, which may tell us more about the availability of estuary English across the south.

vi. *Urban upscaling:* It may be that when respondents were not able to invoke a specific urban-level label, they may have simply named the largest urban centre rather than the region, as we saw with the very high frequencies of respondents estimating Speaker 3's origin to be London. This compares with Kerswill and Williams' (2002) *perceptual dislocation*, in that, based on the accent heard, the majority of judges estimate the speaker to be from a location other than their real place of origin. A further cause for upscaling could be that the metropolis of London acted in this case as a distractor from its surroundings, dwarfing all smaller urban centres around it and the that pull and familiarity of the London association is too strong for non-linguists to contemplate a more specific level of detail. Alternatively, those judges with a slightly more sophisticated geo-linguistic taxonomy may be using the term 'London' not only to mean genuine

estimations of 'London' but also to serve as non-linguist shorthand for 'near London/London influenced' or even as a non-linguist synonym for the linguists' term, 'estuary English'. Future research should be aware of this semantic duality.

6.3.2 Estimations of the origin of the RP-like accent

Speaker 4 was specifically selected for what was deemed to be an RP-like accent. The thinking behind the inclusion of an RP-like speaker was to investigate how people would define him in the context of questions as to his origin, particularly bearing in mind the geographical basis to most non-linguists' ideology of language variation. The salience of this accent for the non-linguists, and its RP-ness was therefore to be determined only by the judges' responses. The aim here was not that respondents would be able to correctly identify his real home town, but rather, to establish how non-linguists articulate their feelings regarding his accent. This section discusses any supplementary factors that may mediate RP accent description beyond those discussed in Section 6.3.1.

With regards firstly to the attempts at eliciting estimations of speaker origin, what we see in the data is that in Stage 1, where only open questions were used, some 44% of respondents stated that they did not know where the speaker was from, a much higher frequency than for the previous regional speakers. However, whilst many may find it difficult to place a specific regional estimation on an RP accent, we can be sure that if the respondents had been asked if the speaker was Northern English, Scottish, Welsh, Irish, American, Australian, etc, the response would have been 'no', as RP has no particular associations with the varieties of English spoken in those regions. And in fact there were respondents, some 49% in Study 1, who did attempt at least a region-level estimation of regional origin.

What the 49% did identify was that the speaker was from the South/South East of England, an estimation approaching that provided for Speaker 3 (70%). This, we could say, is not unreasonable, as the two speakers grew up ten miles

apart. In fact, given the high number of estimations for both speakers, and the knowledge of their similar regional origin, we could conclude that, to the non-linguist respondents, the two speakers simply sound very similar.

However, if one looks past the respondents' estimations of regional origin and considers the data from another point of view, one discovers a glaring anomaly. Also included in the study was the strength of accent scale (Section 5.5), intended to offer respondents the opportunity to differentiate between speakers using social as well as geographical criteria. The data on the strength of accent scale indicate that the evaluations of Speaker 4 were quite high: 74% of respondents placed him to the right of the mid-point, at 5, 6 or 7, on the 'well-spoken' side of the scale, with a mean of 5.26. These estimations for Speaker 4 on the strength of accent scale are significantly different from those for Speaker 3; in fact they are at the opposite end with Speaker 3 having a mean score of 2.71. This is, then the clearest possible indication that the two speakers, despite being seen having similarity in regional background are not heard as speaking with the same accent. On the one hand, we have evidence that the two speakers are identified as being from the same region by large numbers of respondents, yet on the other, respondents are saying that their accents could almost not have been more different. How do we reconcile the anomaly?

First, we must conclude that areal-only estimations of regional background are simply inadequate in the context of British language variation as they do not take into account accent variation across different social groups or the social connotations of accents. Without other instruments such as the strength of accent scale, which affords the non-linguist a more multi-dimensional way of expressing their impressions of an accent, rather than relying on geographical evaluations alone, we may easily misinterpret the data and results of statistical analyses.

Second, it is well-documented that RP speakers are often associated with specific social stereotypes, such as prestige, a good job, money, etc (Garrett et al. 2003), and within the British Isles it is widely accepted that a large amount of wealth, expensive property, etc is situated in the South East of England and

London. In both stages of this research we see descriptions of Speaker 4's South East origins including references to 'Windsor/Eton' (famous and very expensive public schools), 'Oxford/Cambridge' (prestigious universities often associated as the natural destinations for the alumni of famous public schools) the county of 'Surrey' and the 'Home Counties' (a phrase not necessarily applied with geographical precision), both terms stereotypically used to convey wealthy and politically conservative suburbs of the South East. It could be suggested then, that for a proportion of the respondents, the terms 'South East' and 'London' are also mobilised to articulate a social association and judgement, indicating financial power/status/success, and the capital of the UK. There is an argument in the literature that supports this view: for example, Bishop et al. (2005) note the possible distinction between 'London' as a label compared with 'Cockney', with 'Cockney' presenting a narrower set of traditional working class connotations compared to 'London' which may also include connotations of power and wealth. We may see then, how the apparent anomaly of both a strongly regional and strongly 'well-spoken' speaker having the same geographical origins can occur. In the minds of many of our non-linguist respondents, the two forms of speech, a geographical and a social variety, are seen in the South/South East region to cohabit and many labels related to 'London', 'South East' and the 'South' most probably have, in the minds of the non-linguists, semantic duality, conveying a relationship with the traditional working-class speech habits or alternatively, high-status, educated and powerful speakers. This allows for the existence of a one-way association between RP and the South/South East of England: an RP speaker may be associated with the South/South East of England, but as we have seen with the data from Speakers 2 and 3 above, being from the South/South East of England does not mean people will always identify you with RP.

The salience of the social meaning of Speaker 4's RP accent is further evidenced when we consider the results from Stage 2 where guided perceptual mapping task was employed. The data in this stage show an almost 50% drop in the number of 'don't know' responses and an increase in the broadly 'correct' responses: In Stage 1, 37% estimated the speaker's origin as 'London/South East', (48% including the more general 'South'). In Stage 2, 56% of respondents

indicated 'Oxbridge, London, South East' (60%, of respondents including the general 'South'). It would appear that the differences observed are yet again likely to be due to the use in Stage 2 of guided perceptual mapping and the inclusion of labels to indicate speakers of RP. Here, however, we arrive at a critical observation. The critical point here relates to the use of Oxford and Cambridge as cities of origin. Although Oxford and Cambridge are marked on the map, the motivation for respondents to use these labels is likely to be an *association* rather than an estimation related to correctly guessing the urban origins of the speaker. RP has a particular urban association in the minds of non-linguists, even if this association is not based on linguist fact, as more than 39% of respondents selected these two urban locations to articulate their sense of RP-ness. What we are seeing is that Oxford and Cambridge work as effective geographical non-linguist shorthand for associations with the 'RP' accent and it is the provision of a perceptual framework that is assisting non-linguists in articulating these associations.

The implication is that any audiofit taxonomy has to allow for socially marked as well as regionally marked accents, and some kind of framework is needed to assist and focus the non-linguists in their evaluations, such as the strength of accent scale. But perhaps most important of all is that the elicitation framework should consider not asking explicitly where non-linguists believe a speaker is from, as non-linguists encounter difficulty trying to estimate an accent's geographical origin when the accent is a product of social rather than regional background. To ask non-linguists to guess the geographical origin may even be counter-productive, as it may, for example increase inaccuracies or 'don't know' responses. What non-linguists are able to provide is indicators of the social or regional background they *associate* the accent with. A response such as 'this is the sort of accent I associate with Oxford/Cambridge, middle class, or private education' is likely to present less risk of respondent error than 'I know the accent but I don't know where people who speak like this come from, so I'll make a guess because I have been asked to'. By asking non-linguists about their associations, we are more likely, in the context of UK language variation at least, to exploit what non-linguists do know and avoid relying on non-linguists' weaknesses in spatial geography, areal taxonomies or lack of knowledge of

regional and social variation in any linguistic community, and reduce the effects of response biases.

A final comment must be made with regard to respondent origin. The data in Sections 4.2 and 5.2 provide evidence that in the minds of the non-linguists, RP seems to have a broad regional 'home' somewhere in the South of England, most commonly in the South East. Not only is it mostly located in the South East of England, but the data is Section 4.3 and 5.3 suggest that it is most usually located there by respondents from the South East themselves, and many from the South West. It appears then, that the co-existence of RP alongside traditional regional dialects is particularly prevalent in the minds of our young, Southern English, non-linguists, who presumably have close-up, first hand experience. In contrast, we can observe that this perception is at its weakest, in both stages of the research, amongst the only non-English respondents in the study, that is, the Welsh non-linguists. As with the estimations of speaker origin for Speaker 3, the RP accent seems less socially and geographically marked amongst Welsh respondents, who respond with a consistently higher rate of 'don't knows' to the question of Speaker 4's origin. It is likely that factors of social and perceptual space are once again mediating the estimations of speaker origin alongside Euclidian distance, increasing the need for further investigation of language perception at the local level.

6.3.3 Estimations of the origin of Speaker 5

Speaker 5 was selected on the basis of his Midlands origin, in line with the use of speakers from the adjacent areas to Wales, but was also selected on the basis of his particular accent, which was felt to have the potential to be seen as either 'no accent' or 'neutral' in non-linguist perceptual terms. As mentioned in Chapter 3, the inclusion of this speaker took the research away from the common *regional varieties vs. RP* paradigm, which has been the mainstay of much previous language attitudes work, because of the need in the forensic context to understand how it is that accents that can be neither overtly regionally or socially marked are perceived by non-linguists.

Speakers 1, 2 and 3 were always placed within their own regional area by the majority of respondents, and Speaker 5 has been placed within his area by only 14% and 18% of respondents across the two Stages, which relative to the data for the other non-RP speakers, is dramatically low. We may identify this as perhaps one reason *not to rely* on the evaluations of a professional linguist when selecting voices for research. I am confident that had I requested an identification of the speaker's origin, a professional linguist would have identified the speaker as from Birmingham or the West Midlands. It is only by asking the non-linguists directly that we come to understand what is salient to them rather than to the trained ear, and whatever evidence there was of a Midlands accent it was only salient to the few - Speaker 5's accent was insufficiently salient to be identified as from the Midlands. In cases such as these, it is vital to ask the non-linguists what it is that they hear, rather than rely only on the analysis and classifications of the professional. The question arises, therefore, where the respondents *did* believe Speaker 5 was from.

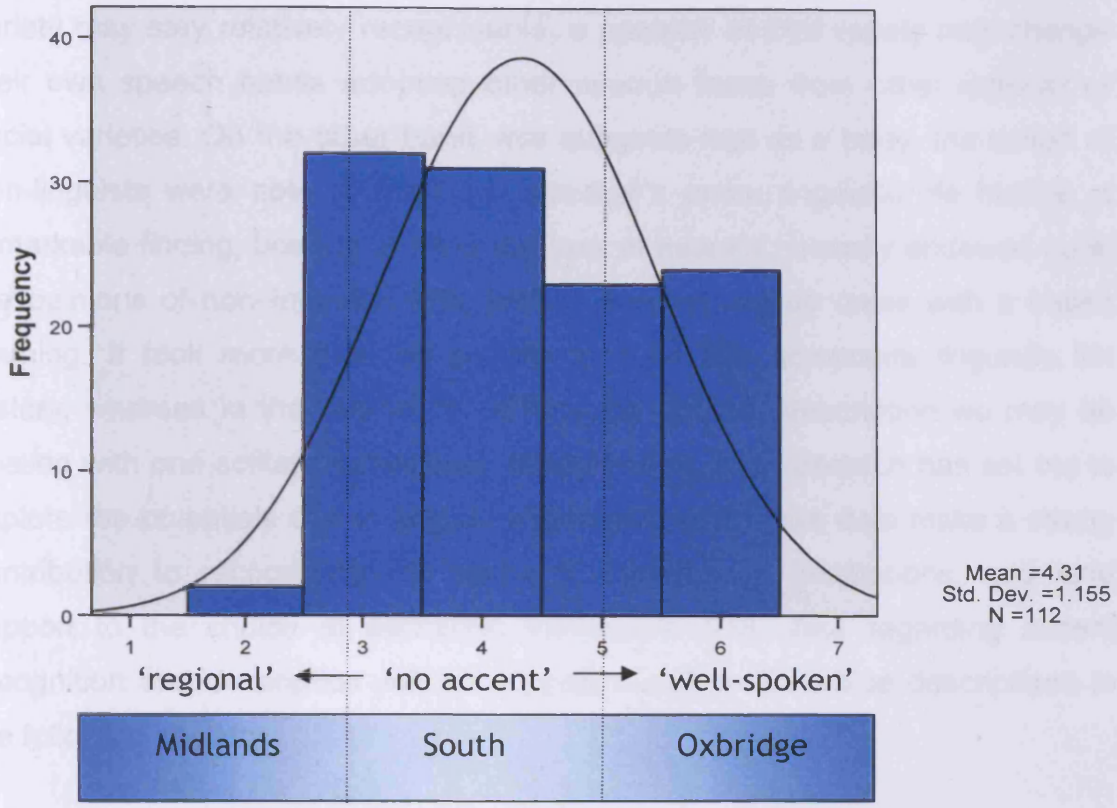
Firstly, when we view the data in 4.2 and 5.2, we may say that there is a pattern of general southernness evident in the estimations for both Speakers 4 and 5's origin. However, in Stage 2, in particular, where guided perceptual mapping was used, we can observe that there is a contrast in the distribution of the estimations: whilst the estimations for Speaker 4 focused largely on the South East, with what might be interpreted as social judgement associations, such as 'Windsor/Eton', 'Oxford' or 'Cambridge', the distribution for Speaker 5 is fairly evenly divided across London & South East (13%), Bristol and South West (9%), the South (12%), demonstrating less consensus amongst the respondents and with few 'RP-connoting' labels. This appears to suggest that those respondents that attempted an estimation were not viewing Speaker 5 as having the same kind of accent as Speaker 4, primarily South Eastern, and associated with the speech of the 'wealthiest' corner of the country. Moreover, in Stage 1, 34% of respondents estimated the speaker to be from a location within the South of England and in Stage 2, with the use of guided perceptual mapping, this figure increases to 51%. It appears that Speaker 5's accent may be salient in terms of its *southernness* as respondents identify a generally neutral southern English accent. This is not inaccurate as the speaker has lived

in the South for some nine years. With no real evidence of a high amount of social stereotyping, either regionally or socially, we may have encountered a phenomenon rarely seen in language attitudes research: responses to a person who is not regarded as particularly regional, or 'posh', but just 'English' (if you are Welsh) or 'no accent' if you are from the south of England.

Furthermore, if we take the data from the strength of accent scales, (5.5) what we see for Speaker 5 the evidence confirms that he was regarded as largely 'no accent'. For finer detail, if one observes the histogram in 5.5.5, showing the mean of 4.3, we can see that there was a small tendency towards the right of the scale, indicating overall that the majority of respondents believed the speaker to have no accent, but that he was perhaps a little 'well spoken'. This 'southern English', slightly well-spoken accent becomes more clearly defined when we consider the use in Stage 2 of the use of Oxford and Cambridge by 24% of respondents as a geographical shorthand to the articulate social associations of the speaker's accent. And indeed, Speaker 5 has, due to his work in the legal profession, made a conscious attempt to move his own accent away from the Midlands towards a more formal-sounding RP-like variety.

These data are highly revealing in terms of what the strength of accent scale can tell us about the wider conception of British accents: the use in the strength of accent scale was, as detailed in Chapter 3, intended as a composite of *regional ↔no accent*, and *well-spoken ↔no accent* scales. It was conjectured that such a scale provided respondents with a model that closely matched the non-linguist conception of accents in the UK. What the data across these five speakers suggest is that for the young non-linguist cohorts used in this study, 'well-spoken' is not the automatic alternative to speaking with a regional accent, and that speaking 'well-spoken' is not viewed as an example of having 'no accent'. An argument for the interpretation of these three shades of identification is that they represent the distribution across the strength of accent scale.

Figure 6.2.3 Interpretation of respondents' evaluations of Speaker 5 'strength of accent'



Taking the data from Stage 2, to the left of the scale are those intuitions that are largely to do with the nature of the 'Midlands' 19% accent (plus a few other small, regional misidentifications); to the right of the scale are those evaluations that represent the well-spokenness or 'Oxbridge' element of the Speaker's accent (24%); and the largest central bulk represents 28% of evaluations of the Speaker being from the 'South'. This would correspond not only with his geographical location since leaving home at 18, but also with the 'Oxbridge' accent so often associated with the legal profession, being based as it is on a southern-derived RP accent.

Moreover, one of the most exciting potentials of these data is that combined, these three elements – 71% of evaluations represent a longitudinal profile of the Speaker's life: Born and raised in the Midlands, at a comprehensive (state) school; moved away at 18, to the South of England; has trained to be and practices in the largely RP-accented world of lawyers and legal professionals. (Speaker 5 was the only speaker to have moved around the country and indeed engaged in social movement in this way.) On the one hand, this has

implications for forensic investigations as it reminds us that whilst a language variety may stay relatively recognisable, a speaker of that variety may change their own speech habits adopting other speech forms from other regional or social varieties. On the other hand, this suggests that as a body, the cohort of non-linguists were able to intuit the speaker's entire linguistic life history, a remarkable finding, bearing in mind the lack of esteem normally endowed upon the opinions of non-linguists. This finding does of course come with a health warning. It took more than 65 people to build this composite linguistic life history, whereas in the real world of forensic speaker description we may be dealing with one solitary earwitness. Nevertheless, this research has set out to explore the potentials of non-linguist description and these data make a strong contribution to recognising the value of non-linguist evaluations, and lend support to the choice of elicitation framework. The data regarding accent recognition and description will be supplemented by the voice descriptions in the following section.

To conclude, this speaker takes the research in a direction away from the common *regional varieties versus RP approach* which has been the mainstay of much previous language attitudes work in the UK. Language attitude needs to do more, therefore, to understand the social value of 'no accent' speakers.

6.4 Audiofit voice descriptions

When invited to provide qualitative comments on voices, what descriptive vocabulary do non-linguists provide? When provided with categories of voice features, how do respondents describe the selected speakers?

6.4.1 Bipolar semantic differential scales

What we saw in Chapter 4 is that many of the non-linguists readily provided comment on the categories provided within the questionnaire, such as 'speaker age' and 'speaker origin'. These were neither areas where respondents felt lacking in confidence to comment, or where they were unable to agree.

Beyond speaker age and accent, however, a substantial part of this research was to investigate the potentials for developing an audiofit, an instrument that police can employ in the elicitation of voice descriptions from earwitnesses. Yarmey (2001:114) has indicated that it is unlikely that at an initial police interview, witnesses would be provided with a set of rating scales, but that it is more probable that the police would ask the earwitness to 'verbally describe the perpetrator's voice'. Yet, as Braun and Künzel argue, a detailed speaker profile may be 'of paramount importance to investigators' in their attempt to narrow down the field of potential suspects (Braun and Künzel 1998:13).

Künzel (1996:146) has suggested that some individuals are able to specify linguistic features for speaker description, some of which correspond 'roughly to phonetic categories'. As was reported in the Results for Stage 1 (Chapter 4), some 1068 qualitative items were collected, all of which the non-linguists had provided in their description of the unknown voices. When these were analysed for themes, it was seen that as a response to a question as general as 'how would you describe the speaker's voice', non-linguists provided descriptions that were detailed in nature and broadly the same as those listed by a professional forensic linguist, Hollien (2002):

-pitch and tone (non-linguists) = 'fundamental frequency or heard pitch' (Hollien)
-voice quality (non-linguists) = 'general voice quality' (Hollien)
-speed, fluency, speaking tone (non-linguists) = 'prosody' (Hollien)
-force (non-linguists) = 'vocal intensity' (Hollien)
-speech impediments (non-linguists) = 'speech impediments' (Hollien).

Certain professional categories, 'articulation' and 'idiosyncratic language patterns and pronunciations' were not included in the comments made by the non-linguists, perhaps because the vocal percepts had not displayed these absent features, or possibly because they were not available to the non-linguists as salient features.

Beyond these categories, we see that non-linguists also engaged in the description of features that the professional linguist would not seek to comment

on, being too subjective or empirically weak to monitor: some non-linguists felt it perfectly appropriate to comment on the speaker's educational background, on the type of person they are, (such as 'a bloke, roughish, laddish'), their state of mind or personality, ('calm', 'aggressive'), or their believability ('genuine', 'coherent', 'vague'), all features which a professional linguist operating according to the IAFPA's code of practice would not discuss. But what we must not ignore is that these are the features that were salient to the non-linguists, and it is upon all of these features, including those unused by the professional linguist, that the subsequent evaluations (see following section) were based.

In summary, what Stage 1 revealed is firstly that non-linguists' observations of voice features are very similar to those made by professional linguists; moreover, whether the professional linguist or the courts consider them acceptable or not, what the Stage 1 research also illustrated is some of the non-linguists' *own* bases for evaluative judgements, as opposed to those suggested or imposed by researchers or legal professionals. Such evaluations will be discussed in section 6.5.

The use of voice descriptors in Stage 1 not only revealed non-linguist conceptions of voice characteristics, but provision by the judges of varied descriptors for each person also demonstrated how non-linguists are able to give vivid descriptions of individual voices; they were able, through their own qualitative remarks, to characterise the speakers according to their respective patterns and clusters of voice features. Table 4.6.1 details how the commonly reported features for each speaker produced an illuminating profile of each of the voices heard, with differences to be observed not only in the speakers' accents, but in their vocal features also. We can see, for example, that judges' responses reflect perceived differences between speakers in terms of *force*, 'loud' (Speaker 1) versus 'quiet' and 'soft' (Speaker 2) or *fluency*, 'hesitant', 'jumpy' (Speaker 4) versus 'flowing', 'controlled' (Speaker 5). Moreover, where features were not particularly prominent or salient to the non-linguists, no comment was provided, such as *force* for Speakers 3 and 4, or *fluency* for speakers 1 and 3. This suggests that the non-linguists were able to perceive

and articulate differences between speakers, and not only at the level of accent (Garrett et al. 2003).

It was the data gathered in Stage 1 that then provided the basis for suggested categories for qualitative voice description in Stage 2 in order to establish whether the non-linguists' own terms are sufficiently salient and operationisable to be used by a separate group of non-linguists to describe the same voices. The voice features, divided into polar and non-polar categories were put to the second cohort of respondents as a preliminary audiofit device to assist the non-linguists in articulating their responses. Respondents in Stage 2 were able to select from six possible vocal characteristics with which they could evaluate each speaker's voice, the intention being to invite respondents to comment on these features noting anything that they felt to be distinctive or different. The respondents evaluated by awarding a score divergent from the midpoint '4' on the audiofit scales.

The first question that arises from the results is whether the respondents evaluated these speakers differently using the features provided. The analyses in Section 5.6.1 confirm that all speakers were seen to be different from each other, with the ANOVAs identifying significant differences between all speakers on most features. The significant differences offer support for the claim that an audiofit derived originally from non-linguist observations and vocabulary is meaningful to other non-linguist judges, and that bipolar semantic differential scales, at least in a set of six, are accessible and meaningful to the non-linguists when given the task of describing unknown voices or highlighting those features that they feel to be most distinctive. Given that the differences were significant, Section 5.6.2 reported the data in terms of a voice profile for each speaker, illustrated in tables of means and through boxplots.

What we see first of all when observing the boxplots of the evaluations for each speaker is that the respondents have not given extreme ratings for every feature for each speaker. This suggests that the respondents have used these six audiofit scales perspicaciously, only evaluating features and indicating divergences on those characteristics as they felt they were relevant to each

speaker. We should also remember, however, that along these scales, placing an evaluation at the centre point of '4' is not always an indicator of lack of strong opinion. It may be that the respondents feel strongly that the speaker has, for example, a medium pitched voice. Nevertheless, there are differences in the means and the distributions of evaluations, indicating that as with the non-linguists in Stage 1 who identified their own voice features, the non-linguists in Stage 2 felt certain features to be distinctive to each speaker.

A further question arises with regard to the respective profiles for each speaker: if the audiofit categories supplied by respondents in Stage 1 are meaningful to the respondents in Stage 2, do we see similar voice features emerging in the profiles for both cohorts? The data suggests that on the one hand, all is not perfect in this respect, as there was some contradiction between the individual qualitative comments in Stage 1 and the data resulting from statistical treatment of the scaled evaluations from the 112 respondents in Stage 2. For example, in Stage 1, Speaker 2 was characterised by one or two individuals as having a low and monotonous voice, whereas the summary of the evaluations of respondents in Stage 1 suggests that he has a slightly high-pitched, melodious voice. How do we reconcile such differences? One suggestion is to view Stages 1 and 2 as being evolutionary. The qualitative comments elicited in Stage 1 were not intended to be statistically representative of the views of the whole cohort; rather, the comments and themes were intended to form the basis of a guide to features that are salient to non-linguists. Stage 2 recycled these semantic themes into bipolar scales, in order to assess their relevance to a second cohort of non-linguists, the effect of the cognitive task load, the ease of administration and generalisability. On the other hand, the data overwhelmingly demonstrated a large amount of agreement between the individual qualitative comments in Stage 1 and the quantitative tendencies using an audiofit with the same vocal categories in Stage 2, which is reported below:

Speaker 1 – Wales
Stage 1: Deep, loud, slow, monotone, expressionless
Stage 2: Deep, loud, steady

Speaker 2 – South West
Stage 1: Stumbles, stammers, quiet, low, medium, like a girl's voice, feminine, monotonous, shaky
Stage 2: Shaky, slightly fast, high-pitched, melodious, loud

Speaker 3 – South East
Stage 1: Hesitant, high-pitched, monotone, fast
Stage 2: Hesitant, high voice, soft speech

Speaker 4 – South East, RP-like
Stage 1: Slow, deep, monotone, expressionless
Stage 2: Slow, deep, monotonous, soft, steady speech

Speaker 5 – Midlands, 'no accent'
Stage 1: Expressive, lively, natural, medium, well-projected, flowing,
Stage 2: Melodious, high, loud, steady, fast

If we compare the data from the evaluative audiofit categories for each speaker in Stage 2 with the original qualitative comments describing each speaker in Stage 1, we see some considerable overlap in the characterisation of each speaker's voice; this, despite the comments having been elicited from different cohorts of respondents. From the possible six choices, those vocal characteristics that respondents selected as being distinctive when describing each speaker's voice, are broadly similar or identical to the characteristics that were described in the qualitative comments of respondents in Stage 1. Moreover, the direction of the evaluations on the scales in Stage 2 mostly matches those qualitative comments in Stage 1. We see, for example, that respondents in Stages 1 and 2 used identical characterisations for Speaker 1, 'deep' and 'loud'; Speaker 2 is 'shaky'; Speaker 3 is 'hesitant' and 'high-pitched'; Speaker 4 is 'slow', 'deep' and 'monotonous'.

More loosely, very similar characterisations are seen for Speaker 1, 'monotone', 'expressionless' (Stage 1) and 'steady' (Stage 2); Speaker 2, 'like a girl's voice', 'feminine' (Stage 1) and 'high-pitched' (Stage 2); Speaker 4, 'expressionless'

(Stage 1) and 'steady' (Stage 2); Speaker 5, 'expressive', 'lively', 'flowing' (Stage 1) and 'melodious', 'fast' (Stage 2). What we see is that when the respondents in Stage 2 do feel that the audiofit contains a particular category that is relevant to a speaker's voice, they frequently comment in the same direction as those qualitative comments elicited in Stage 1. This lends considerable support to the validity and reliability of non-linguist perceptions of voice characteristics as well as to the premise of an audiofit.

What researchers, linguists and the police should take from this exercise is three-fold: first, through a combination of qualitative and quantitative methods, we appear to have isolated a small set of meaningful voice features that are conceptualised by and salient to non-linguists; second, the framework, i.e. the use of bipolar semantic rating scales, appears to have facilitated the reporting of the non-linguists' evaluations in articulating their audiofit descriptions of voice features; third, we have evidence that non-linguists' descriptions of voices can be aligned relatively well with traditional scientific descriptions of voice, speech and language. Views such as those from Yarmey (2001) that witnesses lack the analytic skills or available terms or concepts to describe relevant voice characteristics and Shuy's (1993) view that the non-linguist does not have the meta-language to describe his/her own language need to be revisited and re-evaluated.

6.4.2 Correlations and factors for bi-polar features

The data in 5.5 demonstrate that there are patterns in the ways that non-linguists utilised the bi-polar scales. What is remarkable is the way that certain audiofit factors seem to consistently (if not exclusively) associate with each other.

For example, we find in the evaluation of every speaker that 'tone' and 'pitch' consistently load on to the same factor. Likewise, for the evaluations of every speaker, 'confidence' and 'steadiness' load on to the same factor. Moreover, for every instance of these factorial loadings, 'pitch' – 'tone', and 'confidence' – 'steady' are always found to correlate. The evidence is even more compelling

when we consider that at no point do 'tone' or 'pitch' ever correlate with either 'confidence' or 'steadiness', and nor do they appear in the same factorial loadings. The research reflects that non-linguists have perceived to have isolated two distinct attribution measures in their description of the voices; what has emerged from the data are factorial loadings one related to *voice height* ('tone' and 'pitch'), and one related to *firmness* ('confidence' and 'steadiness'). Furthermore, we must remember that the features were derived from the non-linguists' responses in Stage 1 and used by a second cohort of non-linguists in Stage 2. We can therefore trace a direct link between these two emergent attribution measures and the conceptions and evaluations of non-linguists. This pushes forward our understanding of how non-linguists (and even, perhaps, linguists?) evaluate voices, and how those voices are heard, encoded and evaluated.

With regard to the other dimensions in the bipolar scales, for Speakers 1, 2 and 5, we also see a relationship between 'confidence' and 'speed' in factorial loadings and correlations. The likelihood here is not that these are one attribution measure as for the features above; rather the likely relationship is that the confidence of a speaker can be inferred from their speed (and loudness). That is, the perceived confidence of a speaker is enhanced by a faster speech rate, whereas slower rates of speech can raise doubts over the speaker's confidence (Kimble and Seidel 2005). It should be noted that there is one particular exception to this tendency in the correlational data for Speaker 5. Although the speaker is evaluated as speaking slowly (apparently, the slowest of all five speakers), there is no apparent downgrading of his 'confidence' evaluation; in fact, the non-linguists did not evaluate him away from the central '4' point on the 'confidence' scale. Why would this be so? The answer most probably lies in the influence of one of the two other features that correlate with 'confidence' for Speaker 4: Speaker 4 was the only speaker for whom we see a correlation between 'strength of accent' and 'confidence'; that is, the higher the evaluation on the scale for 'strength of accent', the higher on the scale was the evaluation for 'confidence'. More specifically, a high rating on the scale for 'strength of accent' was a specific indication of the non-linguists evaluating the speaker being 'well spoken', an evaluation that was only attributed to Speaker

4. The strong likelihood is, then, the use of an RP accent makes a significant contribution to obscuring the influence that the slow speech rate may have had on non-linguists' evaluations. That is, accent may be more potent than speech rate in this context.

Finally, with regard to the remaining bi-polar vocal feature, 'loudness', it would appear correlations and factorial loadings vary dependent on the speaker being evaluated. We see, for example, for Speaker 1 a correlation between 'loudness' and 'confidence', for Speaker 2, 'loudness' and 'steadiness', for Speakers 3 and 5, 'loudness' and 'speed', and for Speaker 4, 'loudness' and 'tone'. It would appear that loudness is very much an independent characteristic in the minds of the non-linguists, as its associations vary with each speaker.

6.4.3 Non-polar features

The data for non-polar features were an example of a small data set which does not readily lend itself to tests of statistical significance and distribution. However, the normative analysis presented in 5.6 made the case that rather than selecting distinctive features through a criterion-based approach, and identifying statistically significant features, it is sometimes more practical to make a comparison of features in their context, i.e., identifying which features stand out from others as being more frequently selected, and identifying potential clusters of features.

This normative analysis demonstrated that, as with the bi-polar features, we can also see that most non-polar voice features that were derived from the qualitative comments elicited in Stage 1, have been sufficiently meaningful to a large section of the respondents in Stage 2 for them to have focused on what was salient to them and to have variously attributed characteristics to the speakers, even when no response was required. We can again, then, acknowledge that respondents have been perspicacious in their evaluations of the speakers; they have not uniformly ticked every box, stating that every voice feature was present, and indeed there are certain boxes which remained unticked by all 112 respondents apparently left empty if they were not relevant,

'sharp', 'shrill' and 'raspy' (although we must note that certain of the characteristics may have been ignored as a result of the lexis not being universally transparent or meaningful).

A comparison of the qualitative data from Stage 1 and the non-polar data for Stage 2 (excluding those bipolar features mentioned above) again shows remarkable overlap with certain speakers

Speaker 2 – South West
Stage 1: Stammers, stumbles, like a girl's voice, feminine
Stage 2: Stammer, stutter, effeminate, lisp, croaky

Speaker 3 – South East
Stage 1: Hesitant
Stage 2: Stammer, effeminate

What is apparent from the data in 5.6 is that, although the attribution of non-polar adjectives is at a relatively low rate, it is clear that their provision in the audiofit section of the questionnaire facilitated a further description of the speakers' voices, adding detail that might otherwise have gone unelicited.

The extent to which any other adjectives may be required or any of the current adjectives rendered unhelpful will be a matter for further research, employing a wider sample of voice donors.

In summary, Stages 1 and 2 saw the process of eliciting speaker descriptions through the provision of voice descriptors derived from non-linguists' own comments and observations, and their recycling in the form of bi-polar semantic differential scales and non-polar adjectives. Largely identical to the categories of voice qualities set out by the professional linguist, Hollien (2002), these non-linguist derived voice features have been used with effect by the non-linguists in Stage 2. The conclusion is that the administration of this form of audiofit has yielded some detailed and reliable results. What is required now is for these data and this approach to evolve as a Stage 3, an additional study taking the

bipolar and non-polar audiofit responses of earwitness and assessing their effectiveness in identifying one speaker from a bank of many. On the strength of the data in this study, such an approach has and is bound to continue to yield great potential for the area of forensic speaker description and identification, where a selection of candidate voices is regularly used in voice line-ups, and where the imperative to provide information in pursuit of a criminal may render more detailed descriptions necessary.

6.4.4 Soundalikes

The data in 5.8 are, to the best knowledge of this researcher, the first of their kind. If we consider the frequency of the production of soundalikes, we must first acknowledge that this will not always be the richest source of voice description data. It is entirely possible that the unknown voice encountered simply does not sound like anyone the earwitness recognises, either famous or known to them personally. This should not, however, detract from the highly descriptive data that may be elicited.

The analysis in 5.8 was limited only to those soundalikes that could be identified by the researcher due to their public profile, but even with that restricted data, two important aspects of voice description emerge. Firstly, the soundalike frequently conveys information about the accent that may not have been overtly described or detailed in the rest of the research. For example, those famous names cited as possible soundalike for the traditional regional Speakers 1, 2 and 3 are almost consistently associated with each respective speaker's real home area of Wales, the South West or the South East. More specifically, for Speaker 1, the three soundalikes are known to be from South Wales and whether the respondent realises or not, their intra-regional choices are providing ever more detailed information regarding the speaker, by excluding Mid and North Wales. We also see the use of the social stereotypes to convey accent information. For Speaker 2, the soundalike was not a specific person, but a farmer (suggesting the realisation of the postvocalic [r]) and Speaker 3 was described as a 'chav', a pejorative term which amongst other things, conveys detail regarding the speaker's use of a traditional regional working class accent.

The use of soundalikes to convey accent information is not limited to regional accents. One of the difficulties that we have seen in the data for respondents in both Stages, but especially Stage 1, was the location geographically of an accent that was viewed to be 'posh' and not particularly associated with any one region (Speaker 4, RP-like). The soundalike presents an alternative channel through which the non-linguist can articulate their sense of the speaker's 'poshness', by selecting someone with a similar level of RP-ness.

Secondly, soundalikes appear to convey more than information about the accent. For Speaker 3, for example, the choice of David Beckham, Frank Lampard and Mike Skinner conveys information regarding the pitch of the voice, all three of them sharing voices that are relatively high-pitched, in the same way that Stephen Fry, Jack Dee, David Walliams and Damon Albarn from the group 'Blur' may betoken a deep voice.

Thirdly, there may be a possible social judgement taking place. With the soundalikes for Speaker 3, Beckham and Skinner are all known to be modestly educated, working class boys who as men have gained fame and wealth. This could be a positive stereotype, but given the generally negative assessments of Speaker 3 on most dimensions throughout this study, it is more likely to be derisory way of acknowledging that their wealth and fame is not commensurate with their education or intelligence, i.e. Speaker 3 is not well-educated.

The three dimensions discussed above give rise to the question of whether language researchers, the police and legal professionals have in the past been guilty of approaching research and respondents from too narrow an understanding of how people hear voices. To expect that respondents hear an accent in isolation from other features may be misguided. Whilst we accept that – certainly for sociolinguists – it is fascinating that language varies geographically and socially, and that what we say as humans is powerfully influenced by our geo-social background and context; and whilst we take great interest in the fact that there is an aspect of people's vocal performance that can be separated as extra-physiological and dependent on one's geo-social

context, the isolation of accent alone for analysis may not be based on the realities of how people hear and what they attend to.

What we see in the data provided by the non-linguists when describing soundalikes is not just a description of a person with the same accent; other aspects of vocal quality are also attended to, such as pitch and delivery. Moreover, the soundalikes do not equate to only one aspect of vocal performance each, but may encapsulate *at the same time* many aspects of the speaker's output: the voice, the prosody and the pronunciation. To observe and focus only on one aspect of the aural channel such as accent alone, as the sole distinguishing characteristic of a speaker's repertoire in the forensic context is analogous to stating that when we perceive people through the visual channel, we only notice the clothes rather than the shape, size, colour and features of the physical being.

Moreover, soundalikes offer respondents the opportunity to characterise the voice and accent encountered by describing their *associations* (see Section 6.3) with that particular manner of speaking at the point of encounter, and without having to commit to an exact description of regional origin that could be technically inaccurate, or make statements that could give the impression that the voice used during the encounter is the suspect's only way of speaking. The non-linguist police officer is thereby better equipped to see the line between general and specific detail of description.

Finally, as with all approaches, we must add a note of caution: given the potential for verbal overshadowing (Schooler and Engstler-Schooler 1990) in this context, there is the danger that once a person or social stereotype is invoked, the non-linguist may proceed to describe the voice of the soundalike rather than the voice of the suspect. Given this, it is recommended that soundalikes be sought, as was the case in this research only after any accent or voice description has been elicited.

From the perspective of the language researcher, police officer or legal professional, then, the soundalike question has the potential to supply a rich

vein of linguistic information regarding the voice(s) heard, and is one which should be exploited in future research and in the wider forensic context, and this research strongly advocates its potential advantages.

6.4.5 Forensic voice description: 3 descriptive channels

The previous three sections have discussed the data from both Stages of the research and in particular, what we can learn about the non-linguist conceptions, perceptions and descriptions. But of particular note is the emergence of three distinct channels through which the linguist or legal professional can harvest these perceptions: These three channels are reviewed below.

1/. Geo-social referencing

The first channel in the context of forensic speaker description refers to perhaps its most frequently researched line of investigation in linguistics, and the most frequently employed by the police, usually centred on the question, 'Did he have an accent?' Linguists themselves are familiar with the various social paradigms that exist within a speech community or nation for accounting for speech variation and it appears that researchers and also the police have often seen the verbal cues relating to the speaker's background or origin as a rich vein to be tapped.

This line of enquiry can firstly be oriented towards eliciting clues that may indicate the speaker's supposed regional provenance. The matryoshka paradigm of villages and towns within regions, regions within larger regions, larger regions within countries, forming part of a multi-directional speech continuum, is probably that which is most frequently utilised by linguists when describing language variation. And given the high rates of successful regional identification of speaker origins in the two stages, it would appear that non-linguists also use some form of geo-linguistic referencing to make sense of language variation around them. Moreover, the widespread evidence of non-linguists invoking geo-linguistic paradigms in order to estimate the origins of speaker accents, and doing so to great effect, indicates that the professional

linguists' negative evaluations of those who have no background in language studies and their ability to perceive or articulate their perceptions around the speech variation in their community may be somewhat overstated (see comments by Yarmey and Shuy on Chapter 1, p.22).

However, whilst non-linguists clearly acquire a certain number of geo-linguistic paradigms, a geo-linguistic approach to investigating speaker accent does not always provide a helpful framework for eliciting the required descriptions. The problem is simply that the existence of non-regional, *social* accents, such as RP, can make it hard for non-linguists to respond to questions regarding a speaker's regional origin especially if no assisting framework such as guided perceptual mapping is employed.

In this research, two alternative approaches were investigated: firstly, the researcher uses guided perceptual mapping and includes on the map and in town lists, the names of cities that are commonly used as non-linguist shorthand for social accents. In the case of the UK, Oxford and Cambridge are often used by non-linguists to indicate the RP-like nature of a speaker's accent and these were employed in Stage 2 of this study, with a considerable uptake from the non-linguists when describing Speaker 4 (RP-like). This way, the researcher introduces an avenue for *social associations* into an existing geo-linguistic paradigm, creating a geo-*social* approach. These may lack geographical accuracy, but may be equally or more salient to non-linguist earwitnesses and police officers alike, if they draw upon well-known universally salient social stereotypes. The second alternative geo-social approach, the composite 'strength of accent' scale, enables respondents to indicate whether a speaker's accent is regional or social. Moreover, they can also distinguish between the 'well-spoken' (RP-like) and 'no accent', which we also saw in respondents' evaluation of Speakers 4 and 5 (Stage 2).

It should be noted that geo-social referencing does come with a few notes of caution: firstly, despite the successes of this approach, and the indication that non-linguists can and do use geo-social referencing, the researcher should be careful not to presume the same level of detail is shared between the linguist

and the non-linguist. The paradigms that have been evolved and developed by those who have a professional interest in language variation require a detailed understanding of language variation, which does not reflect the understanding of non-linguists. We saw for example that when evaluating the origins of the regional speakers, strong evidence emerged that non-linguists are good at identifying the origin of speaker accent on a regional level, but weaker on the urban level. Researchers and the police must be sure, therefore, not to project any technical assumptions onto respondents/earwitnesses, through the questions or framework employed, as to do so may lead to respondent error.

Secondly, unless employed contemporaneously with other approaches, geo-social referencing as the sole channel may inhibit the respondent non-linguist from articulating their perceptions surrounding the uniqueness of a speaker's *voice*, as it affords them only a geo-social channel through which to articulate their perceptions. Researchers and legal professionals should be careful to give respondents the opportunity to articulate all of the detail they have noticed about a speaker, utilising multiple channels.

Thirdly, we have seen that there is evidence of respondent origin being associated with the accuracy of identification of speaker origin. Geo-social referencing should not assume perceptual symmetry and is in need of further research, exploring local-level social and perceptual space, in order to catalogue current respective community and non-community salient accents.

2/. Physical speech characteristics

The second channel that has been explored in this research and that is available to non-linguists is the description of physical speech characteristics. Non-linguists will provide their perceptions regarding the quality and manner of a speaker's voice or any speech peculiarities. What we saw was that in Stage 1 respondents provided comments when invited to and that in Stage 2, respondents utilised bi-polar semantic differential scales to characterise the voices heard; non-polar adjectives also provided detail regarding the voice, in what might be regarded as a supporting role.

We saw in both Stages that the qualitative comments elicited were broadly similar to the categories used by a professional linguist and it is clear from the results in Stage 2 that we have arrived at a first set of non-linguist derived bipolar semantic differential scales. Moreover, when provided to the non-linguists as a means of expressing their intuitions regarding the voices, respondents seemed to grasp the descriptors readily and utilise them effectively for reporting the distinctive characteristics of each voice, data which stood up to a rigorous statistical treatment. We saw also that non-polar adjectives were used by non-linguists to report the vocal peculiarities of each speaker and that there were clear patterns emerging. However, this is an area in need of further investigation, especially in terms of how these apparently unconnected adjectives can best be meaningfully presented to the respondents to maximise their uptake and be statistically scrutinised.

The results from this research also indicate that there are certain attribution measurements that appear to be one psychological dimension in the minds of non-linguists (voice height and firmness). This investigation of physical speech characteristics takes us one step closer to fully understanding how voices are mentally represented. What is required next is the use of a wider range of voices in order to expand the non-linguist descriptive categories. Until such time as a complete picture of non-linguist mental representations of voices emerges, the use of non-linguist derived bipolar and non-polar adjectives such as those used in this study should always be supplemented with an open invitation for other qualitative comments regarding the voice, in order that the framework does not restrict non-linguist description. Furthermore, such dimensions merit further investigation also in other forensic fields, in order to establish how they interact with, for example, the credibility of witnesses in court trials (compare with Lind and O'Barr (1979)).

3/. Comparative social associations

The third channel for eliciting forensic voice descriptions utilises non-linguists' comparative judgements to conceptualise and elicit their evaluations. What we have seen in the two channels above is that geo-social referencing and physical speech description rely on common criteria held in the public mind. These

criteria may not be precisely defined, being knowledge acquired formally or informally, consciously or unconsciously. However, there are situations in which the earlier two channels may be of limited assistance. Consider the following:

i/. The non-linguist recognises the accent, but does not know where the accent is from (lack of technical knowledge re: language variation)
ii/. The non-linguist claims that the speaker 'didn't have an accent' (lack of technical knowledge re: language variation)
iii/. The non-linguist recognises something distinct about the voice, but finds it hard to describe (lack of technical knowledge re: articulation)
iv/. The non-linguist recognises the accent, but finds it hard to describe the features (lack of technical knowledge re: language variation and articulation)

We can see the limitations: what point is there in asking earwitnesses to describe the unknown voice that they heard if they have not learnt enough about language variation to hazard a guess where somebody is from? What point is there in asking earwitnesses to describe the unknown voice that they heard if they lack the vocabulary to sketch a description? How does a non-linguist sketch a voice that they claim has no stand-out features? The non-linguist may not be able to provide information through the channel of physical speech description, and may not have available a sufficiently detailed mapped paradigm of geo-social language variation across a community.

The third channel is a response to these limitations and focuses on a more metaphoric and often imaginative set of criteria, in that rather than focusing specifically on any particular aspect of the spoken performance, it invites respondents to elaborate on any comparisons or associations between the speaker's voice and that of someone known to them personally or through popular culture, or any socially stereotyped region or institution. The resultant responses may involve the smallest amount of linguistic vocabulary and on this basis may be easier for the non-technical, i.e. police officers, to employ. For example, a comparative channel may be of assistance in the four points mentioned above by asking, 'Have you heard this accent/voice before? Have

you heard an accent/voice like this before? Who or where do you associate the speaker's accent/voice with? Who did the speaker sound like? Anyone you know personally? Anybody famous?'

The advantages offered are that the non-linguist may not know where the speaker is from, but they *do* recognise the accent and can point out where a similar sounding accent may be heard; the non-linguist may not have the technical vocabulary to describe the features, but they can articulate their perceptions by describing someone who, for example, sounds similar. By focusing on a known speaker, it may be possible to elicit a wider amount of information than simply by focusing on the two channels above.

What we would expect from comparative social associations channel are two likely response types: firstly, a *real life soundalike*: a substantial number of respondents described Speaker 3 as sounding like 'David Beckham', indicating not only region of origin, but also aspect of tone of voice, pitch of voice. There may even be inferences about the speaker's educational background (i.e. academically weak); other examples of this type were discussed in section 6.4.3. This is a good demonstration of how comparative associations can serve to articulate a great deal, with follow-on questions asking the non-linguist to expand, if they can, on what aspects of the voice are so similar.

Secondly, respondents may choose to invoke *social stereotypes* as the basis of their comparisons. In each culture, nation, language community, there exists the widespread use of social stereotypes. They are acquired at various stages of our socialisation and adult life, and while we may not subscribe to their veracity, we nevertheless are aware of their existence. We can compare this with Preston's (1996) argument regarding non-linguists using the term 'nasal' (see p.50). They may not be using it in a technically correct way but it is, for most non-linguists, community salient. The point is that it is a *reliable* criterion, as the majority of native speakers in the community understand it, and it which has widespread currency.

Take for example, the often used (in this research too) term 'farmer'. For many non-linguists, an accent that is widespread throughout the South West of England contains certain features that are especially salient. Although the rhotic accents are present throughout Scotland and Ireland, in certain parts of Wales and north-west England, the realisation of post-vocalic [r] in particular appears for many southern British speakers to function as a shibboleth. When it is 'turned on' (perhaps in combination with a cluster of other salient features - /aɪ/ being realised as [ɔɪ] in words such as 'night' or 'right'), many non-linguists have a very specific way of indexing the speech, and that is to label the speaker as a 'farmer' or of rural origin, with the attendant implication of a lack of educational and cultural sophistication. Furthermore, if one were to invite non-linguists, especially in the South of the UK impersonate or characterise the speech of a farmer, it would be extremely likely that southern British speakers would realise the post-vocalic [r]. However, in the present research, the non-linguists were responding to the voice not of a farmer, but of a man who has lived all of his life in the centre of one of the largest cities in the England, Bristol. Whilst the term 'farmer' may be totally inaccurate when describing the speaker from inner-city Bristol, it is nevertheless a term that is widely used by non-linguists to describe speakers with a certain pronunciation. It may not be valid as a term for describing the speaker's origins, but so many non-linguists use the term 'farmer' to index this speaker's style of pronunciation that the term is nevertheless a potentially reliable one. That is to say that for many non-linguists, the presence of one particular feature, or cluster of features is commonly indexed in the same way. And whichever word is selected to index this perception is to a large extent irrelevant.

Most linguists, of course, are loath to use any such stereotypes in their work as general terms of reference. Accusations of being racist or at best ignorant would have disastrous consequences for the individual towards whom those accusations are made. Yet, the non-linguist in the street is often far less reticent and throughout this research, there have been many examples of non-linguists using learnt social stereotypes to articulate their perceptions regarding the voice, and quite possibly the character of the speaker also.

It would appear then that the case for not pushing the non-linguist towards an exact geographical identification is strengthened; non-linguists rely on social stereotypes that often totally misrepresent the speaker's origin as well as their character. The data that have emerged in this research suggest that, viewed differently, this consistently widespread mis-characterisation via comparative social associations may be seen in more positive terms.

6.5 The role of language attitudes

How do a group of non-linguists evaluate the selected speakers? To what extent are their perceptions, descriptions and attributions of guilt permeable to pre-existing voice and accent stereotypes?

In this section, the role of language attitudes and their interaction with the evaluations of the speakers will be discussed as will the extent to which the evaluations of the speakers acts as a backdrop to the eventual attributions of guilt and romantic success.

6.5.1 Intelligence

Stage 1 invited respondents to rate the speakers according to intelligence, honesty and aggression. Intelligence has been seen as a constituent evaluation of the *superiority* factor of speech evaluation (Zahn and Hopper 1985), now more commonly referred to as *prestige* (Bishop et al. 2005; Coupland and Bishop 2007). In their recent update of Giles' (1970) work, Bishop et al. and Coupland and Bishop revealed that much of the social evaluation of more than thirty-five years ago has remained remarkably unchanged in the UK, and the patterns that arose in this study with regard to speaker intelligence bear more than a passing resemblance to those patterns, old and new. On performing a normative assessment of the 'intelligence' evaluations for each speaker, we see in this research that of the five voices evaluated, the RP-like accent of Speaker 4 performs significantly better than those of the three British accents that were identified by our respondents as being regional. This is entirely in keeping with Giles' (1970) work. Specifically, we see that not only is RP evaluated more highly than those accents identified

by non-linguists as being regional, but that Giles' hierarchy is replicated with Wales and the West Country being ranked mid-table, and the South East accent being in the final position.

Furthermore, the tests for significance suggest that the attitudinal responses fall into three rather than five separate evaluative ratings: the highest set consists of those speakers defined by non-linguists as 'well-spoken' (RP-like) and the 'no accent' (Midlands), who are not evaluated significantly differently; a second set (also not seen as significantly different) consists of the Wales and South West Speakers; third on the list comes the South East speaker. This suggests that non-linguists accorded high prestige, middle prestige and low prestige in the evaluations and again, such a division could be said to replicate the findings in the previous literature.

However, two anomalies arise: In previous studies where speakers from the accent of Birmingham/West Midlands has been presented, either aurally or conceptually, the Birmingham accent has been consistently downgraded (Giles 1970; Bishop et al. 2005; Coupland and Bishop 2007). Indeed, Dixon et al. point out that as a 'third class' urban accent the *Brummie* has been generally evaluated more negatively than rural, regional accents and RP (Dixon et al. 2002:166). But in this study, the speaker from Birmingham/the Midlands (Speaker 5) does not come bottom of the rankings, but in fact comes top.

To find that a speaker from the Midlands is most positively evaluated is not only interesting in itself, but it also serves as a salutary reminder of Preston's (1996) criticism of language attitudes work that has failed to ask of the respondents where they actually *think* the speaker is from. The fact that the speaker is from the Midlands and that professional linguists would be able to identify key features of Midlands pronunciation in his speech is irrelevant to the evaluations of most of the respondents, as most non-linguists did *not* identify him as being from the Midlands; many respondents felt that he was the closest to 'no accent'. Moreover, being identified as a 'no accent' speaker and then receiving the most positive evaluations suggests that for our young non-linguist respondents from around the south of the UK, the 'no accent' variety is currently seen as more

prestigious than RP. We could argue that 'well-spoken' has lost its status as the educated *norm* English (but not as a socially marked *well-spoken* variety), to be superseded by an accent with perhaps less social stigma associated with being 'posh'.

The second anomaly is the ranking of the speaker from the South East. His ranking as lowest on the dimension of 'intelligence' is at odds with the more recent data produced by Bishop et al. (2005) and Coupland and Bishop (2007), which suggests that the London accent is, in fact, rated rather highly on a 'prestige' dimension, especially among current 18-24 year olds. Bishop et al. (2007:151) point out that their finding may be the result of confusion over the concept of a 'London accent', with its allusion to the busy and dynamic capital city, versus Giles' 'Cockney' label, with its traditional working-class connotations. Bishop et al. (2005) acknowledge that this may be a drawback of a conceptual approach, presenting labels and categories as if they were neutral. It seems, then, that the differences observed between the ranking for the conceptual London accent in Bishop et al. and Coupland's work, and the accent heard by the non-linguist respondents in this study, are two distinct attitude objects, subject to semantic duality, and on this basis no great time need be spent on comparisons.

But a second, more fundamental point, is that when investigating non-linguists' descriptions of speakers and their evaluations according to their own conceptualisation of language varieties and attendant ideologies, researcher-imposed conceptual labelling is likely to elicit information related to the researcher's ideologies rather than language as it is heard by the folk. These two anomalies are taken as support for the methodology used in this study, whereby real voice donors were engaged in order to investigate the non-linguist respondents' conceptualisations and evaluations of the accents in their own terms. Moreover, this research asserts that it is essential in the field of forensic speaker description to employ real voices as stimuli, rather than researcher-imposed conceptual prescriptions, especially in earwitness work.

6.5.2 Honesty

The evaluation of 'honesty' falls into Zahn and Hopper's (1985) factor of *attractiveness*, latterly known as *social attractiveness*. In Giles' (1970) original work he found that the rankings for accents were re-ordered most notably with a rise in the esteem of the West Country accent and a slightly lower rating for the South Wales accent. Bishop et al. (2005), in using the West Country conceptual label, also found a higher rating for the West Country accent. The data for the present study demonstrate a similar pattern, with the 'no accent', RP-like and 'West Country' accents all being favoured over the 'South Wales' and 'South East' speakers. Moreover, the ANOVA tests found no significant difference between the evaluations of the 'no accent', RP-like and 'West Country' speakers or between the 'South Wales' and 'South East' speakers, suggesting that the non-linguist respondents made a division between high and low social attractiveness.

One explanation for the rise in evaluation of the West Country accent may be that the South West of England has many retirement towns along its largely unspoilt coastline, and in the interior, large stretches of undeveloped moorland (Exmoor, Dartmoor) or farmland. In the national media, the South West of England is rarely represented as containing large urban areas (ignoring Bristol, Plymouth, Torquay, Exeter, etc) in fact, quite the opposite occurs, with the media frequently portraying those from the South West as fishermen or farmers, and the area itself as being dotted with 'twee', safe, friendly, cosy villages almost always exclusively inhabited by white people. Such is the media's fondness for such representations, that these cosy villages are used as a backdrop for murder mysteries, whose appeal appears to be based on the juxtaposition of grizzly murders with an idyllic rural setting (for example, the perennial series 'The Midsomer Murders' following the work of a detective solving murders in an idyllic rural village in the fictional county of 'Midsomer', whose name is apparently based on an play on the label of the South Western county of Somerset, as presented in Giles' origin work). Moreover, those who are seen to be the 'indigenous' country folk are frequently represented as having a 'country' accent, most obviously represented by the use of the

postvocalic [r] by actors, whilst those who are seen to be the wealthier classes are often represented as having 'neutral' or 'RP-like' accents. This rhotic-local relationship appears to be mutually exclusive in that those with rhotic accents are assumed to be 'locals' and those that are 'locals' are portrayed with rhotic accents. This can be seen in, for example, the long-running radio drama, 'The Archers', the popular television comedy show 'The Vicar of Dibley', and the adult comic Viz with its character 'Farmer Palmer', "A parody of farmers in Britain. He is evidently from the West Country, judging by his accent" (Wikipedia 2007)).

In terms of the responses elicited in this research, it appears that respondents may not always be hearing an urban Bristolian accent, but rather, they are hearing an accent representative in the mind of a stereotypically 'farmer's' or West Country speech variety. This would result firstly in diminishing the Bristol accent's urban integrity, and secondly, in creating the very real possibility that speakers whose entire life experience and values have been shaped by their urban experience in England's sixth largest city are being evaluated according to respondents' belief that they have a rural background from the South West. The anomalous evaluation of the Bristol speaker's voice with higher ratings for 'honesty' than for 'intelligence' may be an example of iconicity, fanned by media and public stereotypes of those with 'local' accents being less-educated, but harmless. This anomaly will continue to be observed in the following sections on 'speaker aggressiveness' and 'speaker guilt.'

Turning to the two other voices for which respondents seemed to identify a strong regional correspondence, South Wales and the South East, both of the speakers appear to be examples of what respondents are hearing as 'urban' voices. The reason for suspecting this to be the case for the Hastings speaker, is that more than 70% of respondents perceived the speaker to be from or near to the largest city in the UK, London, and presumably identified the accent as more 'Cockney' than 'metropolitan urbanite'; and of the nearly 70% of respondents who 'correctly' identified Speaker 1 as Wales, with some 45% estimating the speaker to be either from either Cardiff or South Wales. The common connection here is that Cardiff, South Wales and London carry with

them largely urban connotations (and for those respondents who are only familiar with South Wales and its industrial heritage, it may be that to some extent even the label 'Wales' may have urban/industrial connotations also). It is rare in the UK to see many media representations of or hear much public discourse with people regarding London or Cardiff and their environs as being 'rural', 'idyllic', 'a place to retire to', etc.

These data not only confirm that the non-linguist language ideology has remained largely unchanged in thirty-five years, but also that the respondents were differentiating between speakers on the dimensions of 'intelligence' and 'honesty'.

6.5.3 Aggression

The decision to elicit evaluations of speaker 'aggression' was made in order to establish, firstly if non-linguists would distinguish between speakers on the dimension of aggression using only verbal cues, and secondly, how, in the forensic context, evaluations of physical aggression may be related to 'attributions of speaker guilt' (next section).

The evaluative dimension of 'speaker aggression' was not included in Giles' (1970) original study of language attitudes in the UK, nor was it included in Zahn and Hopper's respected Speech Evaluation Instrument (1985) or any other work since. As a result, there is no pre-defined factor onto which 'speaker aggression' is known to load, and the data in 4.9 suggest no direct alignment between 'speaker aggression' and either 'speaker honesty' (i.e. the factor of *social attractiveness*) or 'speaker intelligence' (i.e. the factor of *prestige*). However, if no direct relationship can be observed, the next option is to consider that 'speaker aggression' may be partially related to one or both of the other dimensions. Of the two, the closest relationship to be observed most probably exists between the evaluative patterns for 'speaker aggression' and 'speaker honesty' than with 'speaker intelligence'. Especially when we consider the correlational relationships in Section 4.11, in which we see that for Speakers 3, 4 and 5, the strongest and most 'substantial' correlational relationships

(Guilford 1956) are to be observed between 'speaker honesty' and 'speaker aggression'.

What we see, then, is insufficient evidence to suggest that 'speaker aggression' is simply a replica of other existing evaluative dimensions or factors. Given this, we must consider the possibility that other vocal features, such as those described in the 'physical speech characteristics' section of the audiofit may have an independent effect on the attributions of 'speaker aggression' and that this instrument may be sensitive to the context, in this case, forensic, in which it is utilised.

Turning to the individual evaluations for each speaker, what we see in the comparison of means and the ANOVAs is that the speakers identified as West Country and 'no accent' are evaluated most positively, and that there is no significant difference between their evaluations. The most negative evaluations were awarded to the two 'urban' accents of South Wales and the South East, who are seen as the most aggressive, again with no significant difference to be observed between their evaluations. This leaves the RP accent in a mid-ranking position, being seen as significantly more aggressive than the West Country and the 'no accent' speakers, but less than the 'urban' voices.

Given the tradition of language attitudes studies that has found RP to be evaluated near the top on most dimensions, these results are somewhat surprising. A number of explanations may be found: compared to, for example ratings for 'speaker intelligence', the significant drop in evaluation ranking for the RP speaker may reveal more about an ongoing loss of status of RP speakers socially, and negative stereotypes emerging as to the speaker's personal qualities. Any stereotypical connotations that have existed historically with regard to RP being the language of the educated, fair and law-abiding and powerful social group may be on the wane. The evaluations of these young non-linguists may be indicative of an ongoing social change in the UK, with an increase in social stigma attached to the use of RP. Moreover, it may be that normative comparisons with the 'no accent' speaker have led to a usurping of RP's status, to be replaced by a less socially marked variety, a variety which

respondents in Bishop et al. (2005) and Coupland and Bishop's (2007) research appear to interpret as 'Standard English'. It may be that without the 'no accent' speaker, RP may have fared better.

Alternatively, it may be that other paralinguistic features of the individual speaker's voice (tone, pitch, etc) are influencing the downgrading of the RP-like speaker in this context. For example, the lack of melodiousness and expressiveness noted in earlier sections may be interacting with negative stereotypes attached to RP. This is a matter that will no doubt resurface in future research.

6.5.4 Attributions of speaker guilt

Attributions of speaker guilt were invited in both Stages of the present research. It is worth reminding ourselves at this point that no actual crime was mentioned in the text of the recordings and that the voice donors committed no real crime; the allusion to crime was made only in a question put to the respondents, a question which varied in accordance with the crime type. The three crime types used were: sexual assault, assault and robbery and fraud. Sexual assault and assault and robbery were presented to an almost 50/50 split of respondents in Stage 1 and fraud was presented to nearly half of the respondents in Stage 2. The same voices were used in all conditions. We must also remember that the evaluations for Speaker guilt relate to a performed action rather than estimations of a speakers attributes, in such cases as aggression, honesty and intelligence and that as the evaluation suggested that only one person could be guilty, there is a competitive angle to this dimension that is not present in the other evaluative dimensions.

The tables of means and results of ANOVAs for the two crimes in Stage 1 showed patterns of very similar attributions of guilt across the five speakers, regardless of crime scenario. In the sexual assault scenario, it was Speakers 1, 3 and 4 that received the highest attributions of guilt, all the means being significantly higher than the attributions for Speakers 2 and 5. Speaker 5 is seen as significantly less likely to have committed the crime than all other speakers.

The downgrading of two of the regional accents - Speakers 1 and 3 - is, based on patterns in the existing literature, perhaps no surprise. What was unpredicted, however, is that Speaker 4, with his RP accent, should be seen as not significantly less guilty than the two regional speakers. For the scenario of assault and robbery, Speakers 1 and 3 again receive the highest attributions of guilt followed by Speakers 2 and 4 with Speaker 5 being seen as significantly less guilty than all other speakers. The lack of clear significant differences between all the ratings for all speakers suggests that the pattern here is complicated in that not all respondents shared the same view of all speakers. This can be confirmed if one views the standard deviations.

However, observing the tables of means (Figures 4.10.1.1 and 4.10.1.3) and the combined plot of means (Figure 4.10.1.5), the data appear to suggest that for both the crimes of sexual assault and assault and robbery, the five speakers' voices are evaluated in remarkably similar ways. There is a strong possibility that the crimes of sexual assault and robbery are not treated dissimilarly in relation to the voices despite the evaluations being drawn from two separate groups. It is as if there were some kind of supra-crime type, perhaps 'violent personal attack,' against which the speakers are being evaluated. If we sacrifice some statistical robustness and merge the two very similar results together to obtain a synthesised data set with double the sample size, we may gain a view of the evaluations of 'violent personal attack'. The synthesised results appear in Figures 4.10.1.11 and 4.10.1.12. Over the 144 responses, Speakers 1 and 3 are seen as significantly more likely to have committed a violent personal attack than Speakers 2 and 4, who in turn are seen as significantly more guilty than Speaker 5. It is the traditional regional accents with the urban associations, South Wales and the South East that are viewed most negatively. Ranking in the middle are the speakers identified as West Country and RP-like; finally, it is once more, the 'no accent' speaker, who received the most positive evaluations, significantly different from the rest of the speakers.

Separately or apart, it is difficult to resist the conclusion that the accents used by the speakers have influenced the respondents in the crime scenarios, and this is in line with previous research (Giles 1970; Bishop et al. 2005; Coupland

and Bishop 2007). For example, it was Speaker 5's accent that presented most difficulty for the respondents to place regionally or socially and it would appear from the evidence in all scenarios that speaking with 'no accent' may contribute significantly to reducing the likelihood of guilt attributions, relative to the other accents presented. What needs further attention is how we might deepen our understanding of such impressions by investigating how aspects of the physiological delivery of the message, such as tone of voice and speed, may influence respondents' evaluations.

The data for Stage 2 are less conclusive, with the only significant differences being found between the most positively evaluated speaker, 'no accent' and the most downgraded speaker, South East. In this respect, although the significant differences are fewer, the results nevertheless replicate the data from the two crime scenarios in Stage 1 in terms of urban South East and 'no accent' appearing at opposite ends of the evaluation spectrum; however, and most probably due to the smaller number of respondents evaluating crime, (n=45) in Stage 2, more precise relationships differentiating between the middle-ranking Speakers 1, 2, and 4 are not observable. Nevertheless, the results for speaker guilt in the two Stages generate at least two further points of discussion.

1/. What are the implications for social justice and future research?
Research in the field of attributions of guilt as a result of verbal cues is still in its infancy. However, Seggie (1983) suggested more than twenty-five years ago that decision-making involved in the legal setting represents potentially fertile area of study for sociolinguists, and this research has been no exception to that prediction.

What we have seen in these data is that non-linguist respondents were just as likely to make *ideological* evaluations of guilt as they were to make evaluations on other dimensions with fewer social implications, such as speaker 'honesty', 'intelligence' and 'aggression'. That is, social desirability biases do not seem to extinguish social ideologies of language or accent prejudice when considering speaker guilt, and if requested, non-linguist respondents can readily provide and apply examples of their own ideology of language, knowingly advantaging

certain speakers and disadvantaging others. What remains to be seen is whether or to what extent such ideologies translate into influences upon the process of reporting and describing an unknown speaker's voice and evaluating the credibility of witnesses in police interviews and court trials, where other evidence may be presented alongside.

A suggested future line of enquiry might be to repeat this experiment, but with a higher rate of respondent accountability, such as informing them that the survey is being conducted with the aim of changing the law; if respondents were told that their responses would be enacted in law change, would they so readily provide ideological evaluations of speaker accent? Furthermore, in an extension of the forensic context, such as the effects of non-linguist language ideology in the courtroom (Lind and O'Barr 1979), if respondents were given the responsibility of sentencing speakers who present themselves using such accents as those employed in this study, could we guarantee social justice for all?

With regard specifically to the status of the accents, we must remember first of all that all conclusions are based on one token of each regional or social accent. We cannot rule out that other factors may play a role in the evaluation process. It seems likely, for example, that voice quality may also be an attitude-triggering cue and that there is interaction between an individual's voice qualities and evaluations of intelligence, honesty, aggression, etc. We can be certain that the relationship between accent, speaker and attitude is more complex than previous research has suggested. (These cues were presented to different cohorts of Stage 1 and 2 and therefore do not lend themselves to valid statistical analysis.) Before any firm conclusions can be drawn as to the associations between speaker voice, accent and speaker evaluations, more than one token of the voice should be used in the study, utilising a repeated measures design on the same group in order to establish if it is the accent or some other features that are interacting with speaker guilt.

The identification of the role of other voice features in the evaluation process does not diminish the importance of the social meaning of accent. What the

current data appear to indicate is that from the point of view of young, southern British non-linguists who find themselves in a role similar to an earwitness or jury member, speaking with a less socially marked 'no accent' that is based on a generic southern England speech variety will contribute to being evaluated more favourably. Contrary to what we might have predicted from previous research into the status of RP, the use of an RP accent is no longer seen to confer an automatic advantage, in terms of eliciting positive evaluations, and certainly not in the context of a suggested violent crime; the RP accent is now considered as the third, centre point in the evaluation rankings, and accent alone should not be assumed to be the single motivating factor behind evaluations. Future research should take care to include a wider range of accents than the traditional regional vs. RP paradigm.

With regard to speakers who are heard to use accents that respondents associate with more traditional, working class areas, these speakers are, perhaps, considering the work of Giles (1970), Bishop et al. (2005) and Coupland and Bishop (2007), more likely to be significantly downgraded and are at risk of being unfairly treated at the hands of social justice. Furthermore, should an urban Londoner be considering committing a violent crime, they would be best advised to do so in an accent that the local populace will associate with rural idylls in the wider West Country, in order to evade ideological accent prejudice or better still, sound as generically 'southern' as possible.

We should also consider the role of social stereotypes. As was discussed in 2.10.7.3, once the respondents have decided which regional/social group a speaker belongs to, there is the possibility that they are no longer attending to the individual's speech, but rather, responding to the characteristics of the social group to which they perceive the speaker to belong or projecting the attributes of the perceived social group on to the speaker, exemplifying iconicity (Irvine 1996).

Finally, we must remember that the non linguists were young and that their opinions may be subject to change with continued life experience and

sociolinguistic maturity. We cannot assume stability of attitude or that the attitudinal evaluations elicited in this research are static and not subject to change. For example, a new tv character that is highly sympathetic to one particular accent may result in effecting a change of attitude towards that accent. Alternatively, the data in this research may be a predictor of the beginning of the end of RP as top dog amongst young people, indicating a sea change in British language ideology, as has been suggested by Coupland and Bishop (2007).

2/. What motivations are there behind attributions of speaker guilt?

In considering the motivations for the attributions of 'speaker guilt', we can compare the data for the synthesised guilt attributions for 'violent personal assault' with those for the previous evaluations of speaker 'intelligence', 'honesty' and 'aggression'. Boxed into groups where no significant difference was found in the ANOVAs, the rankings are as follows:

Figure 6.5.4 Comparison of correlations: 'intelligence', 'honesty', 'aggression' 'guilt'

'Intelligence'	'Honesty'	'Aggression'	'Guilt' (Sex Asslt.)	'Guilt' (Asslt & Rob.)	'Guilt' (Combined)
Spkr 5	Spkr 5	Spkr 2	Spkr 5	Spkr 5	Spkr 5
Spkr 4	Spkr 2	Spkr 5	Spkr 2	Spkr 4	Spkr 4
Spkr 2	Spkr 4	Spkr 4	Spkr 4	Spkr 2	Spkr 2
Spkr 1	Spkr 1	Spkr 3	Spkr 3	Spkr 3	Spkr 3
Spkr 3	Spkr 3	Spkr 1	Spkr 1	Spkr 1	Spkr 1

The ranking for 'speaker intelligence', 'speaker honesty', 'speaker aggression' and 'speaker guilt' show the close, albeit not identical relationship in the rankings of the speakers, suggesting that there *may* be a relationship between the evaluations for all factors. However, they are not identical. Speaker guilt is

the only evaluation scale (for 'sexual assault' and combined 'violent personal attack') where we see Speaker 5 being evaluated as significantly different from all other speakers and so we must conclude that the 'speaker guilt' as an evaluative dimension is at least in part, distinct from those evaluative measures that form the 'social attractiveness' or 'prestige' factors used by Bishop et al. (2005) and Coupland and Bishop (2007).

A further way we can examine the motivations behind attributions of speaker guilt is to examine the evaluative dimensions in terms of their correlations (Section 4.11). We see that of all the available dimensions, 'speaker guilt' using the combined 'violent personal attack' for a larger statistical base, most frequently correlates with 'speaker honesty'. For Speakers 1, 2 and 5, the relationship is defined as 'substantial' (Guilford 1956), and for Speaker 3 and 4 'definite'. 'Definite' correlational relationships are also to be observed in between 'speaker guilt' and 'speaker 'aggression' for Speakers 1, 3 and 4. A small, but 'definite' relationship is identified between 'speaker guilt' and 'speaker intelligence' for Speakers 1 and 3. These patterns suggest once again that the relationship between 'speaker guilt' and other evaluative dimensions is a complex one. An alternative view of the data is to isolate each speaker's 'guilt attribution' and their correlations, presented in their rank order, grouped by significant differences between them:

Speaker 5 ('no accent')	'speaker guilt' is inversely related to 'honesty'
Speaker 2 ('W. Country')	'speaker guilt' is inversely related to 'honesty'
Speaker 4 (RP-like)	'speaker guilt' is inversely related to 'honesty'
	'speaker guilt' is positively related to 'aggression'
Speaker 1 (Wales)	'speaker guilt' is inversely related to 'honesty'
	'speaker guilt' is positively related to 'aggression'
	'speaker guilt' is inversely related to 'intelligence'
Speaker 3 (South East)	'speaker guilt' is inversely related to 'honesty'
	'speaker guilt' is positively related to 'aggression'
	'speaker guilt' is inversely related to 'intelligence'

This affords us an interesting view of the motivations for the attributions for 'speaker guilt'. What we see is that for the Speaker 5 and Speaker 2, the only other evaluative dimension that relates to 'speaker guilt' is that of 'speaker honesty'. That is, those who doubted the speakers' innocence inversely doubted their honesty, as we might expect.

However, what we see with the remaining three speakers is that the motivations for 'speaker guilt' become more complex. For Speaker 4 (RP-like), attributions of 'speaker guilt' are also related to a perception of increased 'speaker aggression'. This is a remarkable finding in that, in previous studies, views of 'speaker aggression' have never previously been elicited, and we have never before seen the potential for an RP-like speaker to be associated with this dimension. The most pressing question in the light of this evidence is whether it is the accent alone that is interacting with 'speaker aggression' or whether there are qualities present in the other channels such as the physical speech characteristics or a soundalikes that may be influencing our non-linguists overall impression of the speaker. With regard to the soundalike possibility, amongst the names provided by the Stage 2 cohort of respondents, there were no specific names mentioned that could indicate anything particularly aggressive about the speaker's delivery; this most likely narrows the field to either the physical characteristics of the individual or the ideological properties of an RP-like accent in the minds of the non-linguists; which of the two (or the proportion, if both are in part responsible) is not something that the current research is able to determine. It does however, have the potential to be a revealing line of future enquiry.

With regard to the most downgraded of the speakers, we see that it is not only the lack of 'honesty' and the increased perception of 'aggression' that is interacting with 'speaker guilt', but also the perception of lower 'intelligence'. Where traditional attitudes studies that have used vocal percepts may have accounted for this phenomenon by focusing on the accent of the speakers presented, this study finds itself once more in the position of having identified a number of other features that may also have contributed to these evaluations.

In line with traditional language attitudes research, one explanation may be that it is the urban nature of the accents and their working class associations that have created this association with lower intelligence and increased aggressiveness. We certainly know that urban accents have been almost consistently downgraded in UK-based studies on the grounds of *prestige*, the factor onto which intelligence is believed to load, and the data in the present research replicate this pattern. We cannot be sure which factor 'aggression' loads onto (*prestige*, *social attractiveness* or even *dynamism* (Zahn and Hopper 1985)), and so cannot make comparisons with the typical evaluations of urban accents upon this dimension. It does, however, seem likely that urban accents are viewed as inherently more aggressive than their rural, or 'no accent' counterparts. Further comparative research in this area is required. If this is the case, we can conclude that the accent and its associations of less honest, more aggressive and less intelligent are, at least in part, motivating the attributions of guilt.

We may also need to look once more at the other voice description channels, the vocal features or soundalikes, to explain what may be contributing to or solely motivating these attributions of guilt. Is there something in the speakers' voices that betokens a lack of intelligence or increased aggression for the respondents? If this is the case, we must consider that it is the voice donors themselves that may have contributed through their vocal performances to the impressions that the non-linguists have had, and the evaluations that they have provided. Moreover, there may be a combined effect, whereby certain vocal characteristics are more prominent when delivered through the medium of an urban accent, exponentially downgrading the speakers' evaluations at the hands of the non-linguists. In one sense, this seems unlikely as the most downgraded accents are both urban in nature and appeared in the audiofit to be quite opposite from each other. Speaker 1 was characterised as low, monotone, loud and steady, whereas Speaker 3 was characterised as hesitant, soft and high voiced. This would appear anomalous and tilt the balance in favour of an accent-based rather than a voice-based motivation for negative attributions of speaker guilt. However, as we saw with Speaker 4, the speaker was not very positively evaluated and was also seen as being aggressive, despite having an

RP-like accent. The present study is unable to resolve this query and it would appear that in forensic speaker description, there is more work to be done to resolve the extent to which accents and voice can contribute to attributions of 'speaker guilt'.

6.5.5 'Got the girl'

The final section in this Discussion chapter refers to the question posed to the cohort of non-linguists in Stage 2, regarding the potential for romantic success, an evaluation measure used as a control against the criminal scenarios presented to other respondents. Once again, the same voices were played to the respondents but the suggested context of the topic of their conversation was altered with respondents being told that one of the speakers may or may not have had a romantic encounter with the girl discussed in the recordings.

The comparison of means in 5.10 shows a disparate range of evaluations for each of the speakers, and in the ANOVA tests, no significant differences were found between the speakers. The standard deviations in the data for each speaker illustrate the lack of strong opinion. Of note here is that the two speakers seen by the respondents as not being strongly regional, Speakers 4 and 5 (RP-like and 'no accent') were subject to a wide range of opinions, with evaluations for Speaker 4, in particular, straddling almost the entire width of the scale.

What is evident is that these data are remarkably different from those for the attributions of guilt, where at least two speakers at opposite ends of the comparison of means were seen to be significantly different, and the data in Stage 1, where very significant differences between the evaluations of the speakers were evident. There may be a number of explanations for this lack of strong tendencies in the data.

Firstly, it may be that, as for the attribution of guilt data in Stage 1, the number of respondents was too small to reveal any clear patterns (n=62). However, the Stage 2 attributions of guilt data was based on a smaller sample (n=48) yet still

revealed a significant difference between the highest and the lowest means. An alternative reason could be the nature of the question. Did respondents feel that they had enough information to evaluate a final denouement? Stereotypes regarding accents and voice types may be less well-formed in the minds of the non-linguists with regard to positive achievements. A further possibility is that the task of evaluating romantic partnership, involving an unknown person (the girl in question never spoke on the recordings) may be cognitively more complex than identifying a 'criminal potential' of the person speaking. In these cases it is also possible that the lack of significant differences in the 'got the girl' compared with the guilt scenarios may indicate that the research was successful in not corralling respondents in to presenting non-attitudes (Ostram et al. 1994: 28) where they don't exist (p.59).

Finally, we must remember the possibility that attitudes consist of beliefs about and evaluations of not only the accent, but quite possibly the speaker, vicariously, through accent. Evaluations of the speaker and the speaker's accent need not necessarily align and it may be that this non-alignment is particularly evident in this scenario. Future research may like to consider re-configuring this aspect of the study to achieve a more definitive outcome.

6.6 Summary

In this chapter, I have discussed the patterns in the accuracy in estimation of speaker origin and the possible explanations for these patterns. What we have seen is that not only is it essential to identify what it is the non-linguists hear rather than the professional, but also to consider that identifications are not mutually homogenous and that regional and social varieties of English in the UK are not symmetrically salient. Neither is it always helpful to the folk to use solely geographically-oriented questions as the sole locator of speaker provenance; alternative methods should always be used and providing a framework within which to operate appears to either encourage or assist (or both) non-linguist respondents in identifying the origin.

We have seen that there is also the potential to elicit fairly specific voice descriptions from non-linguists, if they are provided with a framework, although improving the consistency between descriptions will need further work. This may be achieved by multiple test and refinement or by reformulating the format in which the items are presented to the non-linguists.

Finally, with regards to the speaker evaluations, we have seen some common patterns of speaker evaluation, but also, in the status of RP and 'no accent' speakers, perhaps seen emerging evidence of a restructuring of British non-linguist language ideology. Furthermore, I have linked accent, general speaker evaluations and attributions of guilt for the first time and discussed the underlying mechanisms that may be driving this process. We have seen further evidence of how as a salient marker of identity, speaker accent may have a particular evaluative significance in psycho-legal contexts, developing the argument that language is one of the extra-legal factors that affect forensic impressions attributions and decisions (Dixon and Mahoney 2004)

In the following chapter I will revisit the points discussed in this chapter and draw my conclusions.

7 - CONCLUSIONS

7.1 Introduction

This final Conclusions chapter revisits the context of forensic speaker description, reviewing the contribution this study has made to the field and proposing future work. Principally, this study's contribution to the field relates to four main areas: methodological advances; what non-linguists perceive in a speaker and how they conceptualise language and speaker variation; how non-linguists describe what they hear; and how regionally and socially marked accents are evaluated, and the motivations behind such evaluations. This Conclusions chapter looks at how the present research has contributed to each of these four areas.

7.2 Methodology

The methodology of this study has seen the blending of approaches from established fields coupled with a number of methodological innovations. With regards to forensic speaker description, the research has drawn upon the experience of professional forensic linguists and their own evidence regarding the perception and conceptions of earwitnesses. However, this research has departed from the closely related field of forensic speaker identification in that it has focused on non-linguists' descriptions of speakers' accents, origins and voices, rather than engaging the non-linguist earwitnesses in the process of identifying suspects in a speaker identification line-up.

The study has also drawn upon research in the field of language attitudes and social psychology, by presenting sound-only conditions to non-linguist respondents for evaluation, evaluations which the respondents are invited to

map onto bipolar semantic differential scales. However, despite being by far the most frequently employed sound-only condition in previous language attitudes studies, the matched guise technique was rejected in favour of the use of a verbal guise technique, deploying five genuine speakers of each variety as vocal percepts. The voice donors worked with content-neutral texts, recorded using the Picture Memory Interview (PMI), with the aim of being able to elicit descriptions of voice characteristics in addition to identifications and evaluations of speaker accent. The use of the PMI should serve as a complement to other ongoing developments in the field, such as the DyViS project at Cambridge University (Chapter 2, this thesis, p.96). I believe the data gathered across the two stages have demonstrated that it has not compromised access to non-linguists' attitudes and evaluations and but rather, it has been instrumental in eliciting the wide range of patterned responses observed.

The study has also brought together dimensions of speaker evaluation, which have been the mainstay of language attitudes work, with the research on attributions of guilt, bringing together evaluations along the dimensions of prestige and social attractiveness Zahn and Hopper (1985) and exploring their interactions with evaluations of speaker guilt, based on recorded vocal percepts only.

It is believed that the methodology of this study achieves a cross-fertilisation of fields resulting in possible new approaches and new directions for the fields of language attitudes, social psychology and forensic linguistics.

7.3 Non-linguist perceptions and conceptions

With regards to non-linguists' conception and recognition of accents, the earlier sections of this research addressed Preston's (1996) criticism that much language attitudes research has engaged non-linguists in the evaluation of accents without investigating where the respondents believe the speakers to be from, potentially compromising the validity of the research. Given the need for the police and legal professionals to be able to elicit accent descriptions from

non-linguists, the present research accommodated this criticism by presenting to the respondents a sound-only condition utilising pre-recorded vocal percepts rather than working from conceptual labels (compare with Bishop et al. 2005 and Coupland and Bishop 2007). The data from this study demonstrated that in overt tasks, the majority of non-linguist respondents from across the south of the UK were capable of perceptual discrimination between the accents of the speakers presented, these perceptions being evidenced in their accurate estimations as to the origin of the accents used by the speakers.

However, variation was observed in the accuracy and focus of the estimations of speaker origin, with those speakers using traditional regional accents (Speakers 1, 2 and 3) receiving the highest frequency of accurate estimations of origin. Those speakers who were not strongly associated with any one region (Speakers 4 and 5) received more disparate estimations of origin.

The introduction in this study of guided perceptual mapping as an elaboration of Preston's (1996) and Niedzielski and Preston's (1999) approach to folk linguistics has shown itself to have applications in the forensic field. This approach has demonstrated that the non-linguists' areal taxonomy of regional language variation in the UK is well-developed - in open questions, respondents were able to invoke some kind of perceptual differentiation between the regions and their accents. Moreover, data from the second Stage suggest that the respondents' ability to elaborate on their areal taxonomy of regional variation and accent strength is likely to be enhanced also by the provision of simple maps and region listings, as a simple way of expressing their social associations with an accent that has no overt geographical origin. Researchers and legal professionals may wish to draw upon and further develop such a technique.

With regard to respondent origin and the identification of speaker accent, this study has provided evidence that social and perceptual distance and not just Euclidian measures between a respondent and a speaker's origin have a role to play in the accuracy of the identification. The evidence of perceptual asymmetry is consistent with other data from the UK context (Williams et al. 1996; Kerswill

and Williams 2002; Garrett et al. 2004). Perceptual distance is an essential area for further research, as any audiofit that follows as a result of this or future research will be hobbled by a lack of understanding of which accents are community salient, how they are indexed by each respective community. There is for example, the possibility that the higher rates of recognition of the RP speaker in the South East may be related to social meaning and that variety having greater status in that area. Likewise, in the Welsh context, such an accent may carry less prestige and is not 'claimed' in the same way, explaining the lowest rates of identification. Community-level variation must also be a consideration in the selection of future vocal percepts for experimentation, as should percepts from country wide, non-adjacent regions. It is essential to acknowledge that media representations of the accent, the social and geographical mobility of the respondents, in and outgrouping as a function of social attractiveness and social stereotyping, and the focused or diffuse (Le Page 1980) nature of the linguistic community observed all have a role to play in perceptual distance and community salience and these must be considerations for the audiofit and for further research in the area of forensic speaker description.

Turning to the recognition and identification of accents not traditionally associated with any specific British region, it appears that perceptions of the speaker and their origin may be mediated by the task put before the respondents. If asked in open questions to identify an RP-like speaker's regional origin, respondents are more likely to rely on identification based on socio-economic stereotypes, perceiving RP as the accent of the statusful, the wealthy and the powerful, associating the speaker with the British capital city of London. For the linguist or legal professional eliciting the descriptions, the semantic duality of labels such as 'London' is potentially misleading as it may be interpreted as referring to a 'Cockney' accent. The findings in this research indicate that other strategies may be more effective in eliciting a description of the RP speaker's accent, namely a composite 'strength of accent' scale which provides an alternative resource for eliciting intuitions regarding non-regional accents without forcing respondents to provide accent information through a geographical medium. If a regional identification approach is nevertheless

preferred, the data in this research suggest it is helpful to include on any map of other towns that can serve as repositories for respondents' associations with well-spokenness, such as Oxford or Cambridge.

With regards to Speaker 5, this study has given us access to conceptions of an accent which have been largely absent from previous research. Although identified as a Midlands speaker by a few respondents (usually from the Midlands themselves), the regional aspect of the speaker's accent was something of an admixture to the overall accent perception. More commonly, Speaker 5 was identified as having 'no accent', neither regionally nor socially marked. As linguists we know this not to be the case, as it is impossible to speak a language without communicating some regional or social information about oneself (compare, for example, with Giles and Powesland (1975) *inherent value* versus *imposed norm* hypotheses). However, the data from the 'strength of accent' scale lend support to the conclusion that, in the UK, the concept of 'no accent' is a very real non-linguist alternative to, on the one hand, the socially marked RP/well-spoken/BBC/Oxford/posh English and on the other hand, regional/local/vernacular/accented English. What is clear from the current research is that non-linguist ideologies of correct, standard and non-standard variation are culturally embedded and fed, and that for an audiofit to be generated, each culture and language community needs to survey how language variation is conceptualised by its non-linguist communities.

Turning to perceptions of voice, in this study, the matched guise technique was rejected as part of the methodology, on the basis that when investigating forensic speaker description, we need firstly a variety of *physical voices*, each of which could be presented to the non-linguist respondents with the aim of eliciting perceptions, conceptions and descriptions of a range of distinctive features. The use of recorded vocal percepts was responsible for eliciting a corpus of non-linguist perceptions regarding voices, which when analysed provided new data regarding the salience and distinctiveness of vocal features and formed the basis of a vocal audiofit. This audiofit shared many features in common with those categories suggested by professional linguists from Laver's (1980) phonetic approach to Hollien's (2002) forensic linguist treatment

demonstrating that non-linguist perceptions and conceptions may be realised as reliable descriptions of unknown speakers' voices. This is important to highlight given the scepticism in the existing literature towards non-linguists' abilities to describe voice characteristics (see Chapter 1), and it must inform future research in the area. Furthermore, non-linguists' identification of characteristics and the operationalisation of voice descriptors derived from their data provides evidence that it is not only accent that may mediate speaker evaluations, but physical voice features also. This is an area in great need of future exploration.

Finally, the findings regarding non-linguist perceptions lend themselves to comparison with the *modes of linguistic awareness* suggested by Preston (1996). Where Preston suggests that not all facets of language have equal availability to the folk, ranking them as 'unavailable', 'available', 'suggestible' and 'common', the findings of this research lend considerable support to the claim that non-linguist perceptions regarding the accent and vocal features of a speaker are usually 'available' and with the aid of guided perceptual mapping and strength of accent scales, certainly 'suggestible'. The key to accessing them is the choice of system variables and elicitation framework to be utilised by the researcher.

7.4 Non linguist descriptions of unknown speakers

Turning first to the description of the speaker's origins, the data demonstrated that, when invited to provide their own labels to identify the origins of the speakers, the non-linguists were able to provide differing levels of specificity, with global and specific (Preston 1996) estimations of speaker origin: by offering more regional level than urban level identifications - Speaker 1, Wales, Speaker 2, the South West of England, and Speaker 3, the South East of England - respondents demonstrated a clear tendency towards a more global level of description. Within this process of finding a suitable label for the accents encountered, there is evidence to suggest that respondents are likely to engage in perceptual dislocation (Kerswill and Williams 2002) and perceptual upscaling, by way of ensuring a comfortable (for the non-linguist) level of accuracy. To

invite non-linguists to rely on their own cognitive representations and processes when locating and describing an accent therefore endows resultant descriptions with a level of community authenticity essential to the field of forensic speaker description, in which non-linguist police and legal professionals are eliciting linguistic descriptions from non-linguist members of the public. Researcher-prescribed conceptual labels based on a technical understanding of language variation may be all too opaque for the non-linguist stakeholders involved in forensic speaker description, compromising the validity of the questioning process, whereas community-authentic labels are bound to heighten semantic and conceptual transparency. Engaging with respondents' own identifications and descriptions of accents and bringing into focus non-linguists' perceptual scales should inform practice amongst forensic linguists and legal professionals, and mediate the expected outcomes of eliciting speaker descriptions from non-linguists, in terms of what may be expected and the limitations of speaker description.

The 'strength of accent' scale, mentioned above, has been an important innovation in this study. Loosely based on van Bezooijen and van Hout's (1985) 'degree of accentedness scale', the scale appears also to dovetail neatly with British non-linguists' conception of accentedness, facilitating description. The consistency between respondents' estimations in their plotting of the speakers' accents sheds light on how, at the conscious and possibly unconscious level, English native speaker accents in the UK are organised in the minds of the non-linguist. The non-linguists appear to have engaged with this tripartite model, utilising the framework to map their intuitions. What is required is further work presenting the model to non-linguists in re-organised formats, moving the labels around the scale or presenting the framework three-dimensionally. What we need to establish is whether the strength of accent scale in its current format is the closest representation of non-linguists' conception of accent type and whether local variants need to be developed.

One of the most important findings of this research has been the identification of the three channels of linguistic description, geo-social referencing, physical speech characteristics and comparative social associations. Elements of these

channels have been explored in previous research, with perceptual dialectology being explored through non-linguist maps in the United States (Preston 1996; Niedzielski and Preston 1999) and in Wales (Garrett et al. 2003). The present research shifted the geographical focus to the south of the United Kingdom, including Wales, and further explored the estimator variables that may mediate the accurate identifications of a speaker's regional origin (see above). On the basis of those accents that were readily identifiable, this research established that none of the regions from which the non-linguists were drawn was sufficiently unfamiliar with an accent to perform badly at the identification task, although when identifying a regional accent, it is usually those non-linguists from the same region that enjoy the highest rates of identification accuracy. It is therefore observed that within the southern half of the UK, regional accents remain sufficiently salient, albeit not symmetrically, to serve as a basis for identifying the origin of an unknown speaker in a forensic context.

However, and with regards to the non-regional accents, it was also noted that a lack of regional accuracy was not equivalent to providing no description. Non-linguists were able to supply, using their own terms or through guided perceptual mapping, descriptions as to how the speaker sounded, be they based on region or not. The key point is for researchers, the police and non-linguists to be judicious in the materials selected to elicit such descriptions.

The second channel of speaker description draws upon respondents' descriptions of the physical characteristics of a speaker's voice. Yarmey (2001) gives one of the only accounts of an attempt to elicit non-linguist descriptions, and reports that between one and thirteen voice characteristics have been provided by earwitnesses. However, in previous research, non-linguists have generally been provided with researcher-derived frameworks (see Section 2.3.2) within which to plot their perceptions. The innovation in the present research was to formulate an audiofit based entirely on non-linguists' own descriptions. The subsequent non-linguist derived audiofit was seen to be used with a large degree of consistency by a second cohort of non-linguists; using bipolar semantic differential scales, the non-linguists appear to readily rate as high or low those features that are distinctive to a speaker's voice. The non-

linguists' ratings were largely free from inter-rater disagreement, indicating that bi-polar semantic differential scales using non-linguists terms serve as helpful frameworks to assist the articulations of non-linguist descriptions.

Moreover, the data suggest that many of the rating scales derived from the non-linguist responses, and also used by professional linguists may conceptually load onto a smaller number of factors. This is a possible avenue for further research, as the implications are that an audiofit for non-linguists need not be as complex as professional linguists may have envisaged, although this will of course depend on the details of the task within the forensic setting. With regard to non-polar descriptive items, these were used less frequently, and it may be that the ideal format for presentation has yet to be established. The next step in terms of the progression of non-polar and bi-polar descriptors is to present a wider variety of vocal percepts than the five voices used in this study to establish a more complete set of audiofit descriptors, including female speakers, younger and older speakers and non-native speakers. It may well be that alternative descriptors apply to the female voice, such as the description of pitch and height of voice, and for non-native speakers, it would be of great use to investigate the relationship between the salience of the accent and social stereotypes attached to it, and the voice characteristics that non-linguists perceive.

The third channel of speaker description is comparative social associations. Kerswill (2002) has suggested that the recognition of non-community accents may be mediated by whether the voice sounds like someone the judge happens to know. This study has employed soundalikes as an instrument in the description of speakers rather than the recognition of accents, and the data in this study have shown that there is a clear potential for non-linguists to articulate their perceptions by describing a person known to them. The remarkably high rate of 'David Beckham' soundalike descriptions served as an overwhelmingly powerful alternative to description through other the more technically-oriented channels of accent and voice description. Even when the soundalike invoked is a simple terms such as 'Londony' or 'posh', learnt value-judgements that reflect the way that society is perceived to be organised and

the roles that people are seen to play, all of these features are, broadly speaking, uniform in meaning across a linguistic community and are the results of socially-determined learning, but as such, show fairly universal patterns in respect of certain phonetic or linguistic features. Linguists may baulk at the idea of introducing such non-technical street assessment into research but it is non-linguist police and non-linguist earwitnesses that we are focusing on, and so long as such terms are universally salient and have wide currency, they remain viable ways of communicating from one non-linguist to another what has been heard.

A simple note of caution remains, however: we all, at some point confuse the identities or names of individuals. Phoneticians Ladefoged and Ladefoged (1980) give the example from their own trials in which one of the authors was not able to recognise his own mother. Given this, further research should play back examples of the persons that are cited as soundalikes, in order to confirm the identity, and also to explore exactly what it is that is similar between the speakers, and to elaborate on any differences. We must also remember the risk, mentioned in Chapter 6, of verbal overshadowing (Schooler and Engstler-Schooler 1990). Should comparative social associations be utilised as part of an audiofit, there is the danger that once a person or social stereotype is invoked, the non-linguist may proceed to describe the voice of the soundalike rather than the voice of the suspect. Soundalikes should be suggested only after any accent or voice description has been elicited, in which case they may offer a powerful level of details beyond accent and voice description.

What we are seeing here addresses the criticisms that Niedzielski and Preston (1999) identified as common to folk linguistic work: that folk-beliefs are unscientific and worthy only of disdain, that there is an impoverishment of data and that whatever data may exist are inaccessible. What the comparative social associations and soundalikes achieve is to overcome these impediments, by providing non-linguists with the option of utilising a non-technical, non-linguist derived shorthand based on real world examples in order to articulate their perceptions.

7.5 Non-linguist evaluations of unknown speakers

Garrett et al. (2004) have suggested that recognition of accents is bound up with evaluative processes (Garrett et al. 2004:198).The present study accommodates this view and the need for a greater understanding of the effects that language attitudes have upon the process of speaker evaluation, particularly in the forensic context.

The present study has explored these evaluations by presenting a content neutral, sound-only condition, reducing the likelihood of any 'lexical, syntactic or morphological information contained in each speaker's recording being linked to their social, occupational or regional backgrounds' (van Bezooijen and van Hout 1985:138). All identifications of accents should be made on the basis of the pronunciation and suprasegmental features (Kerswill 2004), such as prosody and stress. Underpinning the evaluations is the fact that by asking non-linguists to provide their own descriptions of speaker voice and accent, we know which accent the respondents believed they were evaluating and we can relate their evaluations back to their own perceptions of each speaker's recorded text. This is a departure from previous research in the field of forensic linguistics: the evaluations of the speakers have been based on five real speakers rather than on a matched guise technique or the use of conceptual labels; the voice donors and the respondents were recruited from across four regions of the south of the United Kingdom rather than the United States (Lind and O'Barr 1979), Australia (Seggie 1983), South Africa (Dixon et al. 1994) or Birmingham/Midlands of England (Dixon et al. 2002; Dixon and Mahoney 2004). In terms of location, range and format of vocal percepts, and coincidence of respondent and speaker regions, this study marks a new direction in research into forensic speaker description and evaluation.

In terms of the more general evaluation scales (intelligence, honesty, aggressiveness) the approach employed in this research has highlighted remarkably similar findings to those seen in research from thirty-five years ago (Giles 1970) and more recent research employing conceptual labels as stimuli (Bishop et al. 2005; Coupland and Bishop 2007). The first conclusion we can draw from this is that, contrary to the suggestion by Bishop et al., attitudinal

consistency is not necessarily ingrained in the conceptual approach that both Giles and Bishop et al. adopted, as an unlabelled voice percept approach has peeled away one conscious, ideological layer of evaluation associated with conceptual approach and engaged respondents with a cognitively more complex evaluation process, yet still arrived at broadly the same results: regional accents, especially urban ones, are downgraded in non-linguists' evaluations, accents resembling a more standard English generally enjoy more positive evaluations.

Secondly, the evaluations observed in the present study closely match those seen in previous studies and studies using conceptual stimuli. Van Bezooijen and Gooskens (1999) have argued that in those cases where language attitudes are (mainly) triggered by stereotypes associated with particular language varieties, we may come to understand whether those identifications have been based on the correct identification of the varieties. That is, if the evaluations of the speakers match those in previous studies we can conclude firstly that the non-linguists have identified those same varieties 'correctly' and secondly, that from the vocal percept approach used in this study, we have seen further evidence we can elicit those stereotypical responses from non-linguists without overtly presenting any conceptual labels. Moreover, the data suggest that the methodology and statistical analysis employed in the present study have a sound basis.

Thirdly, the evaluations of the traditional regional and RP accents and the similarities with previous research act as a control against the evaluations of an accent that have been missing in previous evaluative studies, that of the speaker evaluated as having 'no accent' (Speaker 5). Comparison of the evaluations of the regional speakers and RP-like speaker with the 'no accent' speaker indicate that it is 'no accent' that receives the most consistently positive evaluations. We must remember that the methodological approach of the present study allowed for community authenticity in the identification of accents, and it appears from the evaluations of Speaker 5, that non-linguists have identified a variety of a generic southern English, which, although it may contain mild examples of regional or RP-like markers, is seen as largely free from social

or specific regional cues. This variety appears more acceptable and better favoured by the majority of the non-linguists and its arrival on the linguistic radar of young non-linguists may betoken the advent of an emerging shift in language ideology amongst non-linguists. Whereas RP has conventionally been identified as a national standard, and seen as the 'voice of success' (Garrett et al. 2003), the current research confirms that it is not necessarily claimed as an *ingroup voice* by a younger generation of non-linguists. By contrast, having allowed non-linguists to identify this variety of English, the current research concludes that future studies should expand the range of vocal percepts past the traditional regional versus RP paradigm to include a variety that non-linguist themselves have defined as 'no accent'.

Turning to the attributions of speaker guilt, the study has broadened our understanding of the processes behind these evaluations. Previous studies have sought to investigate the strength of the relationship between speaker accent, and crime or crime type. For example, Seggie (1983), pointed to a relationship between accent type (RP, Australian and Asian Englishes), and crime types - financial and violent crime. However, whilst both Stages 1 and 2 have shown that the different accents were significantly differently evaluated, as a function of crime type. Dixon and Mahoney (2004) have suggested that it is not in fact the crime type that mediates attributions of guilt, but rather, the more general association between speaker accent and criminality, and the results of Stage 1 would seem to corroborate this finding. Moreover, although Stage 2 used a different questionnaire and cohort of respondents, we can observe that a romantic scenario did not lead respondents to differentiate significantly between the speakers, giving further possible support to the theory that it is the suggestion of any criminal activity that leads respondents to evaluate the speakers differently rather than the specific type of crime they are suspected of. To further explore this notion, future work should consider a repeated measures design with a wider selection of crime types both physical, and intellectual in nature, and with a selection of other non-criminal scenarios.

With regard to the motivations behind attributions of guilt, one of the limitations of previous studies is that the focus has tended to be solely on the link between

accent and assessments of guilt, implying that the relationship is an exclusively direct one. What previous research has not attempted is to unpack the complex relationship between the *connotations* of speaker accents along general evaluative dimensions, and attributions of guilt. An innovation within the current study was to gain access to the possible unconscious and indirect associations behind accents and speaker voice characteristics, such as connotations of intelligence, honesty and aggressiveness and examine their relationship with attributions of guilt. What the data in this research have indicated is that the accents themselves are not directly associated with attributions of guilt or broader stereotypes of criminality (Dixon and Mahoney 2004), without also being mediated by other evaluative dimensions. The closeness of the speaker evaluations for both sexual assault and assault and robbery lends at least partial support to this idea. However, rather than being an association with general criminality, the contrasting patterns of guilt attribution for the synthesised 'violent personal attack' and the crime of fraud and the lack of significant differences between the evaluations for speakers in the 'got the girl' scenario suggest it is neither criminality nor specific crime type that mediate guilt attributions.

In exploring the mechanisms behind these attributions we may utilise two forms of social categorisation (Smith and Mackie 2000) that might explain aspects of this process. Kerswill (2002) has suggested the process of *mediated attribution*, akin to, 'the speaker sounds to me as if he has the following social and personal characteristics. Somebody with these characteristics speaks language variety X' (Kerswill 2002:168). In an opposite direction, Irvine (1996:17) proposes the process of *iconicity*, whereby the presumed social attributes of a group are transferred to the linguistic features associated within that group, and an occurrence of those features may directly trigger recognition of those attributes. Both examples illustrate the indirect ways in which attributions and identifications of speaker accents may be made. It is the contention of this research that the relationship between speaker accent and speaker guilt, and the separate motivations with respect to each accent and each attribution of guilt are similarly mediated through both linguistic and social group features. At

what level of consciousness this occurs and its directionality is a matter for further research.

7.6 Limitations and other future work

Inevitably with any study, one can see with hindsight where alternative choices could have been made or identify missed opportunities, some of which have been mentioned in their contexts during this thesis; others will be discussed in this section. The first of these points is to consider the selection of respondents. Based on an opportunity sampling approach, the present research engaged with young students from 18-21, all embarking on a university degree. This is of course a narrow cohort and there are inevitably issues over how generalisable the findings are to the wider population. Future work could concentrate on an older section of the population, or recruit a wider cross section of the public (and a larger number of respondents), perhaps by exploring the use of internet-based materials as was demonstrated by Coupland and Bishop (2007) and Bishop et al. (2005), who recruited over 5000 respondents electronically.

With regards to the attitude objects, the present study did not cross check the salience of the accents by consulting other professional linguists prior to playing the voices to the respondents, a decision not free from controversy. As mentioned in Chapter 6, there were explicit reasons for not selecting this option and I believe the final data and the insight that has come from relying only on the non-linguists' judgements justify this decision. However, it is acknowledged that as a general rule it would be advisable to consider carefully whether this course of action is the wisest, and in much linguistic research it will not be.

When selecting the number of attitude objects, it may also be worth considering the use of more than one token from each language variety, in order to establish how consistently each accent is evaluated, and to what extent between-speaker variations in voice characteristics are responsible for variations in speaker evaluations. Moreover, further studies could also consider the geographical range of the language varieties used. By investigating non-linguist perceptual accuracy on a local community basis, we may establish how

finely tuned non-linguists are to the shades of variation within that community. If the study were replicated using varieties of accent found within and close to one city, how well could local residents discriminate between the community-salient varieties in their local area and would we see evidence of a local language ideology in the speaker evaluations and attributions of guilt? Continuing with the area of guilt, the potential flaw in the design of the research concerns the comparisons of speaker intelligence/honesty and aggression with guilt. As the respondents were only able to identify one guilty person, they may not all have used the scale range for all speakers. Future designs should consider either not indicating the number of potentially guilty parties or allowing for respondents the possibility to indicate that they thought the speaker was 'not present'.

With regards to the voice descriptors, one of the most pressing areas for future research is to investigate the currency of voice descriptions elicited; how meaningful are they for other non-linguists and how useful are they in the identification of suspects from a larger pool of candidate authors? Furthermore, future work should also explore the voice description of females. We do not yet know if the same voice descriptors elicited would apply across both male and female sexes and this has certainly been an under-researched area.

Turning to the description of ages, one of the weaknesses of this study is the use of the age descriptor bands. They have at times been used accurately; however, they may also be misleading. Those respondents that indicated an age that was one year away but in the next band were deemed statistically to be inaccurate when this is not strictly true. Future work should either abandon age bands or investigate further how non-linguists conceptualise age ranges.

With regard to the issue of salience and availability, we need to know more in terms of exactly what it is that the non-linguists are identifying as distinctive, and how they are indexing those features. Further work may consider providing respondents at a final stage with transcripts of the recordings, asking them to indicate exactly what it is that is distinctive to them, as has been done by this researcher on the transcripts in the Appendices of the thesis. It is anticipated that if married with speaker evaluations, much may be revealed with regards to

the extent to which respondents are relying on a global identification of clusters of features or basing their decisions regarding speaker backgrounds on a handful of shibboleths.

The 'strength of accent' scale is an instrument that appears to have been utilised with a large degree of inter-rater consistency by the non-linguists. This is not a technique without controversy, however, as it is based firstly on nominal data plotted on an interval scale, and secondly arranged so that RP and regional accents are mutually exclusive. I believe the data collected from the non-linguists are quite compelling, suggesting that this is close to the non-linguist conceptualisation of accent strength. However, it is essential in future work to present the scale in a re-arranged format in order to evaluate its strengths or weaknesses against the current configuration. If a purely nominal approach is preferred, the concepts could be presented as separate words to be circled such as: 'local accent, regional accent, no-accent, well-spoken, other: _____'

Finally, a word regarding ecological validity: One of the limitations of studies such as these is that ethical constraints upon researchers in the field of forensic speaker description prevent them from recruiting real earwitnesses. Without this, we are not in a position, in ecologically valid terms, to mitigate for the effects of memory degradation of voices, or stress encountered by the earwitness. However, although we are not able to practise real crimes upon respondents in the name of research, one compromise in future work may be the use of reconstructed crimes, such as the soundtracks of the UK-based television programme Crimewatch. They may not be authentic experiences, but they may be the closest researchers in forensic speaker description can come without crossing the ethical line.

7.7 A prototype audiofit

The need for a prototype audiofit arises from the question whether the police and legal professionals have been working with an informed understanding of the abilities and limitations of non-linguists. From the outset the answer was a clear negative, and the breadth of evidence reported in the current research, covering a new area of forensic speaker description, has illustrated the gap between current practice and potential benefits in adopting a systemic typology, based on research evidence. This penultimate section of this thesis provides a prototype of an audiofit, based on the conclusion of the research.

Before providing a summary of the proposed contents of an audiofit, a cautionary statement is required regarding how or when an audiofit may be utilised. First of all, the position stated at the beginning of this thesis remains unchanged as a result of the research: an audiofit is, due to lack of resources for electronically replicating the voices encountered by the earwitness, not comparable to a videofit, and is unlikely ever to play as significant a role in the investigative process as its visually-oriented sibling. We will need to conduct further research into the psychological processes behind the recall of voices and the negotiation of description before comparisons between the audiofit and videofit can reliably be drawn. A second and more practical point regarding the prototype suggested in this section is that the description of accent and voice characteristics is an activity that is likely to require a degree of cognitive elaboration on the part of the earwitness, and such tasks will not be appropriate for all earwitnesses; there will inevitably be some individuals who will find the task too taxing or abstract. An audiofit should therefore be an activity that the earwitness elects into because the suspect's voice may be of importance to the investigation, rather than an activity that the earwitness is required to do. A third point is the recommendation that when earwitnesses elect into an audiofit task, their responses may represent their best guess or association, and as such may be technically inaccurate. However, responses should be interpreted with community salience in mind and earwitnesses should be encouraged to elaborate on what features of a speaker's voice are meaningful to them.

Fourthly, audiofit descriptions should only be pursued in the knowledge that an individual's speech will be subject to internal variation determined by emotional, interlocutor, acoustic context and physical condition.

Turning to the format of an audiofit itself, in this prototypic stage of development, and based on the evidence available in this and other research, it seems wise to encourage respondents to retrieve information by all possible means rather than focus on one specific area of voice description. To this end, the use of all the three descriptive channels outlined in Chapter 6 is recommended, and with a specific order of use. The proposal is based on the evidence that the order and framework of elicitation are of paramount importance and the prototype is organised to reflect this.

Stage 1 - Geo-social referencing should seek to extract from the earwitness their perceptions regarding the geographical or social associations stimulated upon encountering the unfamiliar speaker. Questions such as 'did he have an accent', which when handled by non-linguists are likely to be inseparable from the underlying assumption that accents are linked to regional origin, are disfavoured as they may divert respondents' attention to regional accents exclusively. Such questions are also ambiguous and have the potential for earwitnesses to respond in the negative if the speaker has an accent that they do not recognise, an accent similar to their own, or in the UK context, a high status accent such as RP or an accent that is regarded as 'no accent', i.e., too neutral for description. Such responses, handled by non-linguist police officers, are likely to be misinterpreted.

Alternative enquiries that are likely to return greater detail regarding social as well as regional accents should be employed, such as:

- Tell me about his accent.
- Was it familiar to you?
- Did you recognise the accent?
- Was it similar to yours? If so, how would someone describe this accent?
- Did you feel it was a British or a foreign accent?
- Was there anything distinctive to the accent?

It is worth noting the importance of including a question regarding 'an accent similar to mine'. Some respondents do not recognise the distinctiveness of their own speech and hence harbour preconceived notions that their speech is unmarked for dialect features (Thomas 2002:124). If earwitnesses respond that the accent is similar to their own, it would be worth encouraging them to elaborate on how another person would describe their own accent.

The current research has shown that the description of social accents present a problem for non-linguists, especially when presented with questions on the geographical origin of the speaker. The deployment of a 'no accent scale' may be useful here by offering earwitnesses the opportunity to articulate the differences between three accent types, the regional, the RP and the 'no accent' without having to necessarily attach a location to the voice. On the other hand, should the earwitness identify that the accent is a regional one, the respondent should be encouraged to elaborate on whether it is a regional accent that is local to where they live, or from a different part of the UK.

When asking the earwitness specifically about possible regional origins of the speaker, or associations that the respondent has with the accent they encountered, we cannot assume that all of the population use the same dimensions (Garrett et al. 2003:64) when indexing accents. Evidence for this was provided in Study 1, where open questions were used and respondents variably provided list of towns, cities, regions and countries. Also, there is evidence that the ability to identify accents may be enhanced by the provision of a perceptual map or list, including the names of towns, cities and regions. Further, as non-linguists have not demonstrated a great understanding of social variation and the origins of accents like RP, any future audiofit should include the locations such as Oxford and Cambridge on the map and list which may be associated with social varieties. To avoid earwitnesses attempting to make geographical estimations beyond their abilities, the question attached to the perceptual mapping exercise should enquire whether respondents *associate* the speaker's accent with any of the places marked.

Perceptual maps should not be prescriptive as we do not know what people's salient markers are or how global or detailed their areal taxonomy may be. The audiofit should allow them to use their own terms of reference, labelling perceptual maps, or adding locations to perceptual lists if it assists them. Earwitnesses should be allowed to provide as much or as little comment as they want and should not be pressed or compelled to estimate where someone is from or give any greater detail than they feel comfortable with as to do so would simply invoke response biases. Moreover, pressing respondents into answering that they don't know where someone is from may distract them from what they do know, such as the speaker not having a strong regional accent or speaking 'a bit posh', or an accent 'like mine', all of which valid, salient non-linguist terms.

The current research has also reported evidence that perceptual asymmetry is an issue that may mediate non-linguists' ability to recognise accents from other regions and the amount of global or specific detail they may be able to provide. This presents the opportunity to obtain more detailed accent description perhaps than at national level, and what must be developed is a typology of community-salient accents for any area where the audiofit is employed. We need to know how non-linguists perceive local accent/dialect variation, where the perceptual isoglosses are, and how these varieties are indexed. If a localised perceptual map/list can be presented to earwitnesses, the salience of the descriptions for the local non-linguist police officers it is likely to be greatly enhanced. Without this, the audiofit is likely to be operating with one-hand tied behind its back.

The second stage of the audiofit should focus on physical speech characteristics. It seems logical given evidence in other description areas, that an audiofit design should again start with general questions, beginning by asking if the earwitness noticed anything distinctive about the voice or way of speaking, thereafter moving to discuss possible voice-characterising labels, in order to avoid false suggestion. The current research has demonstrated that there are a number of salient voice descriptors to be employed in an audiofit and such voice features as tone/pitch, speed, confidence and loudness may be presented, either in polar or non-polar designs, where appropriate. Finally,

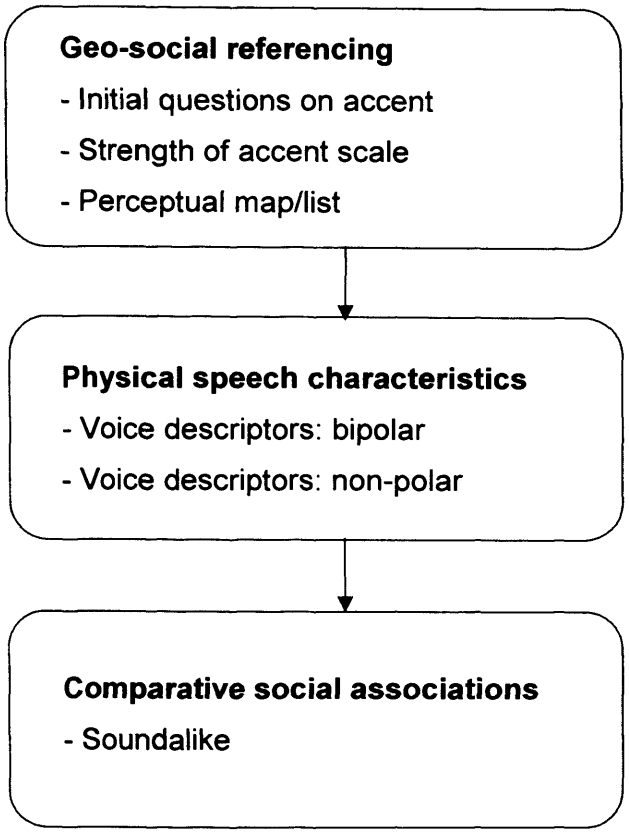
other, non-polar features may be presented, such as speech impediments, lisps or stammers, voice quality, rough, whispery and effeminate, plus the opportunity for earwitness to make their own remarks on any distinctive features. As part of the ongoing development of the audiofit these should be supplemented in future use by descriptions elicited from non-linguists in relation to a wider range of speakers, in terms of age and also female voices.

We cannot be sure that the speaker's accent will be salient to the earwitness or that their areal taxonomy or knowledge of geo-linguistic variation is sufficiently robust to provide accent information. Neither can we be sure that the voice descriptors identified in the audiofit research will be meaningful to all earwitnesses. Thus, the third and final stage, comparative social associations, offers the earwitness the opportunity to consider voices that the speaker might resemble, either someone known personally or through the media or a social stereotype, offering the opportunity to provide their description via another channel. As we have seen in this research, the provision of a soundalike may prove richer in salient detail to non-linguists than an itemised description. Moreover, providing a soundalike is in many ways an easier task than identifying an accent, as we can say who someone sounds like, without needing to know where our referent is from; we can give detail about voice qualities and manner of speech without knowing technical phonetic detail. To exploit the social associations channel, the earwitness should be presented with questions such as:

- Does the speaker sound like anyone you know - personally or famous?
- Do you associate the speaker's voice with anyone or anything?

A last word of warning should be flagged here. As has been discussed before, questions of comparative social judgements should not be administered in the first stage of the audiofit, but should be left until the final point in order to prevent verbal overshadowing. We must avoid the earwitness describing the voice of the soundalike, rather than the suspect, and also avoid the referent serving as an attitude-triggering cue. The prototype for the whole audiofit is presented in Figure 7.7 below, in diagrammatic form.

Figure 7.7 Diagrammatic representation of an audiofit prototype



Finally, comment should be made regarding the administration of an audiofit. In the interviews with officers from South Wales Police, it was revealed that there are no current guidelines for handling voice descriptions. The use of questions such as 'did he have an accent?' is common practice, a practice which, on the evidence reported in the current research, should be reviewed. This research has highlighted the many difficulties that non-linguists have as a result of inaccurate and ill-formed non-linguist ideologies, and the police as a body of non-linguists, are no exception. Yet around the world, non-linguist police officers routinely take statements from non-linguist earwitnesses in relation to linguistic issues, meaning that as a matter of state-condoned practice, witnesses to and victims of crime are being questioned by almost exclusively untrained people. If an audiofit is to be developed and successfully integrated into the task of evidence gathering, the key interlocutors handling the interview should themselves be given training. The police should be aware of the impact that lack of informed linguistic awareness may have on the questioning of earwitnesses, how they may project their own assumptions, accent prejudices and ideologies of language variation onto the statement taking and questioning

processes, how lack of awareness may lead to the neglect of possibly rich veins of enquiry, and how an appreciation of perceptual asymmetry, community-salient labels and social associations may enhance the evidence gathering process and effect the course of social justice.

7.8 Closing remarks

As a final assessment of this study, I believe that the methodology and findings set out in the core of the work represent a significant step forward in the contribution of linguistics and social psychology to the field of social science research. The application of techniques drawn from previously unconnected fields has expanded our understanding of the social-psychological processes at work in the minds of non-linguists, endowed the field with a fresh respect for the abilities of non-linguists, and opened up new lines of investigation towards a clear social purpose.

This research does not of course represent a mission accomplished, as there is undoubtedly more to be explored with regards to elicitation techniques, statistical analyses and underlying relationships between the factors outlined. It is hoped that whatever shortcomings may be identified in this research, they will serve as an inspiration to move the field further forward. The field of forensic speaker description may have its limitations, but it may also be unethical not to pursue it (Braun and Künzel 1998).

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

Allport, G. (1935). 'Attitudes', in C. Murchison (ed.) *A Handbook of Social Psychology.* Vol. 2. Worcester MA: Clark University Press, pp. 798-844.

Andersson, L.-G. and Trudgill, P. (1990). *Bad Language.* London: Penguin.

Baker, C. (1992). *Attitudes and Language.* Clevedon: Multilingual Matters.

Baldwin, J. and French, J.P. (1990). *Forensic Phonetics.* London: Pinter.

Bauer, L. and Trudgill, P. eds. (1998). *Language Myths.* London: Penguin.

Bayard, D. (1995). *Kiwitalk: Sociolinguistics and New Zealand Society.* Palmerston North: The Dunmore Printing Company Ltd.

Bell, A. (1984). 'Language style as audience design'. *Language in Society* 13(2), pp. 145-204.

Bell, A. (1997). 'Language style as audience design', in N. Coupland and A. Jaworski (eds.) *Sociolinguistics: A reader.* St. Martin's Press, pp. 240-250.

Bell, B.E. and Loftus, E.F. (1985). 'Vivid persuasion in the courtroom'. *Journal of Personality Assessment* 49(6), pp. 659-664.

Bishop, H., Coupland, N. and Garrett, P. (2005). 'Conceptual accent evaluation: 30 years of accent prejudice in the UK'. *Acta Linguistica Hafniensia* 37, pp. 131-154.

Blatchford, H. and Foulkes, P. (2006). 'Identification of voices in shouting'. *International Journal of Speech, Language and the Law* 13(2), pp. 241-254.

Bortz, J. (1970). 'On the reciprocal relationships between rating of speakers' voices and the voices of the rater '. *Archiv für Psychologie* (122), pp. 231-248.

Bourhis, R.Y. and Giles, H. (1976). 'The language of co-operation in Wales: A field study'. *Language Sciences* 42, pp. 13-16.

Braun, A. (1996). 'Age estimation by different listener groups '. *Forensic Linguistics: The International Journal of Speech, Language and Law* 3(1), pp. 65-73.

Braun, A. and Künzel, H.J. (1998). 'Is forensic speaker identification unethical - or can it be unethical not to do it?' *Forensic Linguistics: The International Journal of Speech, Language and Law* 5(1), pp. 10-21.

Britain, D. (2002). 'The difference that space makes: an evaluation of the application of human geographic thought in sociolinguistic dialectology', in J. K. Chambers, P. Trudgill and N. Schilling-Estes (eds.) *Handbook of Language Variation and Change.* Oxford: Blackwell, pp. 603-737.

Broeders, A.P. (1996). 'Earwitness identification: common ground, disputed territory and uncharted areas'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 3(1), pp. 3-15.

Brown, B.L., Giles, H. and Thakerar, J.N. (1985). 'Speaker evaluations as a function of speech rate, accent and context'. *Language and Communication* 5(1), pp. 207-220.

Bull, R. and Clifford, B. (1984). 'Earwitness voice recognition accuracy', in G. L. Wells and E. F. Loftus (eds.) *Eyewitness Testimony: Psychological Perspectives.* Cambridge: Cambridge University Press, pp. 92-123.

Bull, R. and Clifford, B. (1999). 'Earwitness Testimony'. *Medicine, Science and Law* 39, pp. 120-127.

Burgoon, M. (1995). 'Language expectancy theory: Elaboration, explication and extension', in C. Berger and M. Burgoon (eds.) *Communication and Social Influence Processes.* East Lansing: Michigan State University Press, pp. 29-52.

Cameron, D. (1995). *Verbal Hygiene.* London: Routledge.

Cargile, C., Giles, H., Ryan, E. and Bradac, J. (1994). 'Language attitudes as a social process: a conceptual model and new directions'. *Language & Communication* 14 (3), pp. 211-236.

Chambers, J.K. and Trudgill, P. (1980). *Dialectology.* Cambridge: Cambridge University Press.

Chen, R. and Hanson, J. (2004). 'Categorically biased: the influence of knowledge structures on law and legal theory'. *Southern California Law Review* 77(4), pp. 1106-1252.

Cook, S.W. and Sellitz, C. (1964). 'A multiple indicator approach to attitude measurement'. *Psychological Bulletin* 62, pp. 36-55.

Coulthard, M. (1997). 'A failed appeal'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 4(2), pp. 287-302.

Coulthard, M. (2000). 'Whose text is it? On the linguistic investigation of authorship ', in S. Sarangi and M. Coulthard (eds.) *Discourse and Social Life.* London: Longman, p. 270–287.

Coulthard, M. (2004). 'Author Identification, Idiolect and Linguistic Uniqueness'. *Applied Linguistics* 25(4), pp. 431-447.

Coulthard, R.M. (1994). 'Powerful evidence for the defence: an exercise in forensic discourse analysis', in J. Gibbons (ed.) *Language and the Law.* London: Pearson Education, pp. 414-442.

Coupland, N. and Bishop, H. (2007). 'Ideologised values for British accents'. *Journal of Sociolinguistics* 11(1), pp. 74-93.

Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language.* Cambridge: Cambridge University Press.

Darbyshire, P. (2005). 'A Recent History of Discrimination in the Legal Profession', *Darbyshire on the English Legal System.* 8th revised edition ed. London: Sweet & Maxwell.

Davies, W. (2007). 'Book Review - Bad Language: Are some words better than others, edited by E. Battistella'. *Journal of Sociolinguistics* 11(1), pp. 136-140.

Dawes, J. (2008). 'Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales'. *International Journal of Market Research* 50 (1), pp. 61-77.

Devine, P. (1989). 'Stereotypes and prejudice: Their automatic and controlled components'. *Journal of Personality and Social Psychology* 56(1), pp. 5-18.

Dixon, J.A. and Mahoney, B. (2004). 'The effect of accent evaluation and evidence on a suspect's perceived guilt and criminality'. *Journal of Social Psychology* 144(1), pp. 63-73.

Dixon, J.A., Mahoney, B. and Cocks, R. (2002). 'Accents of guilt? Effects of regional accent, race and crime type on attributions of guilt'. *Journal of Language and Social Psychology* 21(2), pp. 162-168.

Dixon, J.A., Tredoux, C.G., Durrheim, K. and Foster, D.H. (1994). 'The role of speech accommodation and crime type in attribution of guilt'. *Journal of Social Psychology* 134(4), pp. 465-473.

Eagleson, R. (1994). 'Forensic analysis of personal written texts: a case study', in J. Gibbons (ed.) *Language and the Law.* London: Pearson Education, pp. 362-373.

Eagly, S. and Chaiken, A.H. (1998). 'Attitude structure and function', in T. Gilbert, S. T. Fiske and G. Lindzey (eds.) *The Handbook of Social Psychology.* New York: Oxford University Press, pp. 269-322.

Edwards, J. (1982). 'Language attitudes and their implications among English speakers', in E. Ryan and H. Giles (eds.) *Attitudes towards Language Variation.* London: Edward Arnold, pp. 20-33.

Edwards, J. (1999). 'Refining our understanding of language attitudes'. *Journal of Language and Social Psychology* 18(1), pp. 101-110.

Ellis, S. (1994). 'The Yorkshire Ripper Enquiry: Part 1'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 1(2), pp. 197-206.

Elwell, C.M., Brown, R.J. and Rutter, D.R. (1984). 'Effects of accent and visual information on impression formation'. *Journal of Language and Social Psychology* 3(4), pp. 297-299.

Fazio, R. (1989). 'On the power and functionality of attitudes: the role of attitude accessibility', in A. R. Pratkanis, S. J. Breckler and A. G. Greenwald (eds.) *Attitude Structure and Function.* Hillsdale, NJ: Erlbaum, pp. 153-179.

French, J.P. and Harrison, P. (2004). 'Case Report. R -v- Ingram, C., Ingram, D.and Whittock, T. The Who Wants to be a Millionaire? Fraud Trial'. *International Journal of Speech, Language and the Law* 11(1), pp. 131-145.

French, J.P. and Harrison, P. (in press). "Investigative and evidential applications of forensic speech science", in E. S. A. Heaton-Armstrong, G.Gudjonsson and D. Wolchover (ed.) *Witness Testimony: Psychological, Investigative and Evidential Perspectives.* Oxford: Oxford University Press.

French, J.P., Harrison, P. and Windsor Lewis, J. (2006). 'Case Report. R -v- Humble, J.S. The Yorkshire Ripper Hoaxer Trial'. *International Journal of Speech, Language and the Law* 13(2), pp. 255-273.

Gallois, C. and Callan, V. (1981). 'Personality impression elicited by English accented speech'. *Journal of Cross-Cultural Psychology* 12, pp. 347-359.

Garrett, P. (1992). 'Accommodation and hyperaccommodation in foreign language learners: contrasting responses to French and Spanish English speakers by native and non-native recipients'. *Language & Communication* 12, pp. 295-315.

Garrett, P., Coupland, N. and Williams, A. (1999). 'Evaluating dialect in discourse: Teachers' and teenagers' responses to young English speakers in Wales'. *Language in Society* (28), pp. 321-354.

Garrett, P., Coupland, N. and Williams, A. (2003). *Investigating Language Attitudes.* Cardiff: University of Wales Press.

Garrett, P., Coupland, N. and Williams, A. (2004). 'Adolescents' lexical repertoires of peer evaluation: boring prats and English snobs', in A. Jaworski, N. Coupland and D. Galasinski (eds.) *The Sociolinguistics of Metalanguage.* Berlin: Mouton de Gruyer, pp. 193-226.

Garrett, P., Williams, A. and Evans, B. (2005). 'Attitudinal data from New Zealand, Australia and UK about each other's Englishes: Recent changes or consequences of methodologies'. *Multilingua* 24, pp. 211-235.

Giles, H. (1970). 'Evaluative reactions to accents'. *Educational review*, pp. 211-227.

Giles, H., Bourhis, R.Y. and Davies, R.A. (1979). 'Prestige speech styles: the imposed norm and inherent value hypotheses', in W. C. McCormack and S. Wurm (eds.) *Language in Anthropology IV: Language in Many Ways.* The Hague: Mouton, pp. 589-596.

Giles, H. and Coupland, N. (1991). *Language: Contexts and Consequences.* Buckingham: Open University Press.

Giles, H. and Coupland, N. (1992). *Language: Contexts and Consequences.* Pacific Grove: Brooks/Cole.

Giles, H. and Powesland, P.F. (1975). *Speech Style and Social Evaluation.* London: Academic Press.

Gimson, A.C. (1980). *An Introduction to the Pronunciation of English.* London: Edward Arnold.

Gould, P. and White, R. (1986). *Mental Maps.* Boston, MA: Allen and Unwin.

Grice, P. (1975). 'Logic and conversation', in P. Cole and J. L. Morgan (eds.) *Syntax and Semantics 3: Speech Acts.* New York: Academic Press, pp. 41-58.

Guilford, J.P. (1956). *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill.

Hollien, H. (1990). *The Acoustics of Crime.* New York: Plenum Press.

Hollien, H. (2002). *Forensic Speaker Identification.* San Diego: Academic Press.

Hollien, H. and Schwartz, R. (2000). 'Aural-perceptual speaker identification: problems with non-contemporary samples'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 7(2), pp. 199-211.

Huntley, R., Hollien, H. and Shipp, T. (1987). 'Influences of listener characteristics on perceived age estimations'. *Journal of Voice* 1, pp. 49-52.

Inoue, F. (1999). 'Subjective dialect division in Great Britain', in D. Preston (ed.) *Handbook of Perceptual Dialectology.* Vol. 1. Amsterdam: Benjamins, pp. 161-176.

Irvine, J. (1996). 'Shadow conversations: The indeterminacy of participant roles', in M. Silverstein and G. Urban (eds.) *Natural History of Discourse.* Chicago: University of Chicago Press, pp. 131-159.

Kalin, R. (1982). 'The social significance of speech in medical, legal and occupational settings', in E. Ryan and H. Giles (eds.) *Attitudes Towards Language Variation: Social and Applied Contexts.* London: Arnold, pp. 148-163.

Kalven, H. and Zeisel, H. (1966). *The American Jury.* Boston: Little, Brown and Company.

Katz, D. (1960). 'The functional approach to the study of attitudes'. *Public Opinion Quarterly* 24, pp. 163-204.

Katz, I. (1981). *Stigma: A social psychological analysis.* Hillsdale, NJ: Earlbaum.

Kerswill, P. (1996). 'Children, adolescents and language change'. *Language Variation and Change 8* 8, pp. 177-202.

Kerswill, P. (2002). 'A dialect with 'great inner strength'? The perception of nativeness in the Bergen speech community', in D. Long and D. Preston (eds.) *A handbook of perceptual dialectology.* Vol. 2. Amsterdam: Benjamins, pp. 155-175.

Kerswill, P. (2004). 'Social dialectology/Sozialdialektologie', in K. Mattheier, U. Ammon and P. Trudgill (eds.) *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society.* Vol. 1. Berlin: De Gruyter, pp. 22-33.

Kerswill, P. and Williams, A. (2002). 'Dialect Recognition and Speech Community Focussing in New and Old Towns in England', in D. Long and D. Preston (eds.) *Handbook of Perceptual Dialectology.* Vol. 2. Michigan: John Benjamins Publishing Company, pp. 173-204.

Kimble, C. and Seidel, S. (2005). 'Vocal signs of confidence '. *Journal of Nonverbal Behavior* 15 (2), pp. 99-105.

Komulainen, E.K. (1988). 'Subjective voice identification: the literal meaning of 'Talking yourself behind bars''. *Alberta Law Review* XXVI, pp. 521-547.

Köster, O., Hess, M.M., Schiller, N. and Künzel, H.J. (1998). 'The correlation between auditory speech sensitivity and speaker recognition ability'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 5(1), pp. 22-32.

Kredens, K. and Góralewska-Łach, G. (1998). 'Language as sole incriminating evidence: the Augustynek case'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 5(2), pp. 193-202.

Kristiansen, T. (1997). 'Language attitudes in a Danish cinema', in N. Coupland and A. Jaworski (eds.) *Sociolinguistics: A Reader and a Coursebook.* Basingstoke: Macmillan, pp. 291-305.

Künzel, H.J. (1994). 'On the problem of speaker identification by victims and witnesses'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 1(1), pp. 45-57.

Künzel, H.J. (1996). 'Identifying Dr. Schneider's voice: an adventure in forensic speaker identification'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 3(1), pp. 146-154.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Ladefoged, P. and Ladefoged, J. (1980). 'The ability of listeners to identify voices'. *UCLA Working Papers in Phonetics* 49, p. 43–51.

Ladegaard, H. (2001). 'Popular perceptions of Standard Language: Attitudes to 'Regional Standards' in Denmark'. *Language Awareness* 10(1), pp. 25-40.

Lambert, W.E., Hodgson, R.C., Gardner, R.C. and Fillenbaum, S. (1960). 'Evaluational reactions to spoken languages'. *Journal of Abnormal and Social Psychology* 60(1), pp. 44-51.

Laver, J. (1980). *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Laver, J. (1991). *The Gift of Speech.* Edinburgh: Edinburgh University Press.

Laver, J. (1994). *Principles of phonetics.* Cambridge: Cambridge University Press.

Le Page, R.B. (1980). 'Projection, Focussing, Diffusion, or, Steps towards a Sociolinguistic Theory of Language, illustrated from the Sociolinguistic Survey of Multilingual Communities. Stages I: Cayo District, Belize (formerly British Honduras), and II: St. Lucia. ' *Society for Caribbean Linguistics Occasional Paper No.9.*

Lee, R. (1971). 'Dialect perception: a critical review and re-evaluation'. *Quarterly Journal of Speech* 57, pp. 410-417.

Levin, L., Giles, H. and Garrett, P. (1994). 'The effects of lexical formality and accent on trait attributions'. *Language & Communication* 14(3), pp. 265-274.

Likert, R. (1932). *A Technique for the Measurement of Attitudes*. New York: Columbia University Press.

Lind, E.A. and O'Barr, W.M. (1979). 'The social significance of speech in the courtroom', in H. Giles and R. S. Clair (eds.) *Language and Social Psychology*. Oxford: Basil Blackwell, pp. 66-87.

Lippi-Green, R. (1994). 'Accent, standard language ideology and discriminatory pretext in courts'. *Language in Society* 23, pp. 163-198.

McMenamin, G. (2002). *Forensic Linguistics: Advances in Forensic Stylistics*. Boca Raton: CRC Press.

Milroy, J. (1998). 'Children can't speak or write properly anymore', in L. Bauer and P. Trudgill (eds.) *Language Myths*. London: Penguin, pp. 58-65.

Milroy, L. and McClenaghan, P. (1977). 'Stereotyped reactions to four educated accents in Ulster'. *Belfast Working Papers in Language and Linguistics* 2(4), pp. 1-11.

Niedzielski, N. (1999). 'The effect of social information on the perception of sociolinguistic variables'. *Journal of Language and Social Psychology* 18(1), pp. 62-85.

Niedzielski, N. and Preston, D. (1999). *Folk Linguistics*. Berlin: Mouton de Gruyer.

Nolan, F. (1991). 'Forensic phonetics'. *Journal of Linguistics* 27, pp. 483-493.

Nolan, F. ed. (2005). *Forensic speaker identification and the phonetic description of voice quality*. Mahwah NJ: Lawrence Erlbaum, pp. 385-411.

Nolan, F., de Jong, G., McDougall, K. and Hudson, T. (2008). *DyViS project*. Available at: <URL: http://www.ling.cam.ac.uk/dyvis/database/database.htm> [Accessed: 10/9/08].

Nolan, F. and Grabe, E. (1996). 'Preparing a voice lineup'. *Forensic Linguistics: The International Journal of Speech, Language and Law* 3(1), pp. 74-94.

O'Doherty. (2002). *R v O'Doherty*. Criminal Cases Review Commission 2002. Available at: <URL: www.ccrc.gov.uk/CCRC_Uploads/ODOHERTY_ANTHONY_-_19_4_02.DOC > [Accessed: 8/5/2008].

Orton, H., Halliday, W.J. and Barry, M.V. (1962-1971). *Survey of English Dialects*. Leeds: E.J. Arnold.

Orton, H., Sanderson, S. and Widdowson, J. (1978). *The Linguistic Atlas of England*. London: Croom Helm.

Oskamp, S. (1977). *Attitudes and Opinions*. Englewood Cliffs, NJ: Prentice-Hall.

Ostrom, T.M., Bond, C.F., Krosnick, J.A. and Sedikides, C. (1994). 'Attitude scales: How we measure the unmeasurable', in S. Shavitt and T. Brock (eds.) *Persuasion: Psychological Insights and Perspectives*. Needham Heights: Allyn & Bacon, pp. 15-42.

Pahn, J. and Pahn, E. (1991). 'Formblatt, Eigenschaften, Ablauf und Bedeutung des Tests der Sensibilität formaler sprachlicher elemente im Hinblick auf Perzeption und Produktion'. *Sprache, Stimme, Gehör* 15, pp. 19-23.

Perfect, T.J., Hunt, L.J. and Harris, C.M. (2002). 'Verbal overshadowing in voice recognition '. *Applied Cognitive Psychology* (16), pp. 973-980.

Philippon, A.C., Cherryman, J., Bull, R. and Vrij, A. (2007). 'Lay people's and police officers' attitudes towards the usefulness of perpetrator voice identification'. *Applied Cognitive Psychology* Volume 21(1), pp. 103 - 115.

Plichta, B. (2001). 'Hearing faces: the effects of ethnicity on speech perception'. *30th Conference on New Ways of Analysing Variation*. Raleigh, N.C., USA: 11-14 October 2001.

Potter, J. and Wetherell, M. (1987). *Discourse and Social Psychology: Beyond Attitudes and Behaviours*. London: Sage.

Preston, D. (1989). *Perceptual Dialectology: Non-linguists' Views of Areal Linguistics*. Dordrecht: Foris.

Preston, D. (1994). 'Content-oriented discourse analysis and folk linguistics'. *Language Sciences* 16, pp. 285-331.

Preston, D. (1996). 'Whaddayaknow? The modes of linguistic awareness'. *Language Awareness* 5(1), pp. 40-74.

Preston, D. (1999). 'A language attitude approach to the perception of regional variety', in D. Preston (ed.) *Handbook of Perceptual Dialectology*. Vol. 1. Amsterdam: John Benjamins, pp. 359-374.

Preston, D. (2005). 'What has folk linguistics done for you lately?' *Invited guest speaker*. Cardiff University: 21/4/05.

Riches, P. and Foddy, M. (1989). 'Ethnic accent as a status cue'. *Social Psychology Quarterly* 52(3), pp. 197-206.

Rodman, R. (2002). 'How knowledge of, or ignorance of, elementary linguistics may affect the dispensing of justice'. *Forensic Linguistics* 9(1), pp. 92-101.

Rose, P. (2002). *Forensic Speaker Identification*. London: Taylor & Francis.

Rosenberg, M.J. and Hovland, C.I. (1960). 'Cognitive, affective and behavioural components of attitudes', in M. J. Rosenberg and C. I. Hovland (eds.) *Attitude Organisation and Change*. New Haven: Yale University Press, pp. 1-14.

Roskos-Ewoldsen, D., Arpan-Ralstin, L. and St.Pierre, J. (2002). 'Attitudes, accessibility and persuasion', in J. P. Dillard and M. Pfau (eds.) *The Persuasion Handbook*. Thousand Oaks: Sage, pp: 39-62.

Schooler, J.W. and Engstler-Schooler, T.Y. (1990). 'Verbal overshadowing of visual memories: some things are better left unsaid'. *Cognitive psychology* (22), pp. 36-71.

Sears, D. (1983). 'The persistence of early political predispositions: the role of attitude object and life stage', in L. Wheeler and P.Shaver (eds.) *Review of Personality and Social Psychology*. Vol. 4. Beverly Hills, CA: Sage, pp. 79-116.

Sears, D. and Kosterman, R. (1994). 'Mass media and political persuasion', in S. Shavitt and T. Brock (eds.) *Persuasion: Psychological Insights and Perspectives.* Boston MA: Allyn and Bacon, pp. 251-277.

Seggie, I. (1983). 'Attribution of guilt as a function of ethnic accent and type of crime'. *Journal of Multilingual and Multicultural Development* 4(2), pp. 197-206.

Sherif, M. and Sherif, C. (1967). *Attitude, Ego Involvement and Change.* New York: Wiley.

Shuy, R. (1993). *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom.* Cambridge MA: Blackwell.

Smith, E. and Mackie, D. (2000). *Social Psychology.* 2nd ed. Philadelphia: Taylor & Francis.

Solan, L. and Tiersma, P. (2003). 'Hearing voices: Speaker identification in court'. *Hastings Law Journal* 54(2), pp. 373-435.

Stalans, L. (2002). 'Measuring attitudes to sentencing', in J. Roberts and M. Hough (eds.) *Changing Attitudes to Punishment.* Devon: Willan, pp. 15-32.

Stephan, C. (1997). 'The unknown Englishes? Testing German students' ability to identify varieties of English', in E. W. Schneider (ed.) *Englishes Around the World: Studies in Honour of Manfred Görlach.* Vol. 1. Amsterdam: Benjamins, pp. 93-108.

Street, R.L., Brady, R.M. and Lee, R. (1984). 'Evaluative response to communicators: The effects of speech rate, sex and interaction context'. *Western Journal of Speech Communication* 48, pp. 14-27.

Svartvik, J. (1968). *The Evans Statements: A Case for Forensic Linguistics.* Göteborg: University of Gothenburg Press.

Tanner, D. and Tanner, M. (2004). *Forensic Aspects of Speech Patterns: Voice Prints, Speakers Profiling, Lie and Intoxication Detection.* Tucson: Lawyers and Judges Publishing Company.

Thomas, E.R. (2002). 'Sociophonetic applications of speech perception experiments'. *American Speech* 77(2), pp. 113-147.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich.* Cambridge: Cambridge University Press.

Trudgill, P. and Giles, H. (1978). 'Sociolinguistics and linguistic value judgements: correctness, adequacy and aesthetics', in F. Coppierters and D. L. Goyvaerts (eds.) *Functional Studies in Language and Literature.* Ghent: Story-Scientia, pp. 167-190.

van Bezooijen, R. (1988). 'The relative importance of pronunciation, prosody and voice quality for the attribution of social status and personality characteristics', in R. v. Hout and U. Knops (eds.) *Language Attitudes in the Dutch Area.* Dordrecht: Foris, pp. 85-103.

van Bezooijen, R. (1994). 'Aesthetic evaluations of Dutch language varieties'. *Language and Communication* 14(3), pp. 253-263.

van Bezooijen, R. (2002). 'Aesthetic evaluations of Dutch: Comparisons across dialects, accents and languages', in D. Long and D. R. Preston (eds.) *Handbook of Perceptual Dialectology*. Vol. 2. Michigan State University, pp. 13-31.

van Bezooijen, R. and Gooskens, C. (1999). 'Identification of language varieties: The contribution of different linguistic levels'. *Journal of Language and Social Psychology* 18(1), pp. 31-48.

van Bezooijen, R. and van Hout, R. (1985). 'Accentedness ratings and phonological variables as measures of variation in pronunciation'. *Language and Speech* 28(2), pp. 129-142.

van Bezooijen, R. and Ytsma, J. (1999). 'Accents of Dutch: Personality impression, divergence and identifiability'. *Belgian Journal of Linguistics* 13(1), pp. 105-129.

Wells, J.C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Wikipedia. (2007). *Farmer Palmer.* Available at: <URL: http://en.wikipedia.org/wiki/Farmer_Palmer> [Accessed: 10 April 2007].

Williams, A. and Garrett, P. (2002). 'Communication evaluations across the life span: From adolescent storm and stress to elder aches and pains'. *Journal of Language and Social Psychology* 21(2), pp. 101-126.

Williams, A., Garrett, P. and Coupland, N. (1996). 'Perceptual dialectology, folk linguistics and regional stereotypes: Teachers' perceptions of variation in Welsh English'. *Multilingua* 15(2), pp. 171-199.

Williams, A., Garrett, P. and Coupland, N. (1999). 'Dialect recognition', in D. Preston (ed.) *Handbook of Perceptual Dialectology*. Amsterdam: Benjamins, p. 345–358.

Wolfram, W. and Schilling-Estes, N. (1998). *American English*. Oxford: Blackwell.

Yarmey, D. (2001). 'Earwitness description and speaker identification'. *Forensic Linguistics* 8(1), pp. 113-122.

Yarmey, D. and Matthys, E. (1992). 'Voice identification of an abductor'. *Applied Cognitive Psychology* 6, pp. 366-367.

Zahn, C.J. and Hopper, R. (1985). 'Measuring language attitudes: The speech evaluation instrument'. *Journal of Language and Social Psychology* 4, pp. 113-123.

APPENDICES

Appendix A: Picture Memory Interview **1**
Questions to the voice donors: 1
Pictures for the voice donors: 1
Appendix B: Transcripts of the five voice samples **3**
Speaker 1 3
Speaker 2 4
Speaker 3 5
Speaker 4 6
Speaker 5 6
Appendix C: Questionnaires **7**
Stage 1 Questionnaire 7
Stage 2 Questionnaire – Crime version 9
Stage 2 Questionnaire – 'Got the girl version' version 15
Appendix D: Qualitative voice descriptions from Stage 1 ..... **16**
Speaker 1 16
Speaker 2 18
Speaker 3 20
Speaker 4 22
Speaker 5 24

Appendix A: Picture Memory Interview

Questions to the voice donors:

1/. Why were you celebrating? *(Look at Picture A)*

2/. Where did you go and what time did you get there? *(Look at Picture B)*

3/. Who were you with and where did you stand? *(Look at Picture C)*

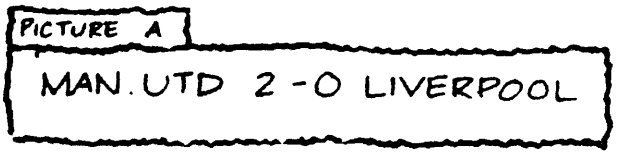
4/. Where was the girl and who was she with? *(Look at Picture D)*

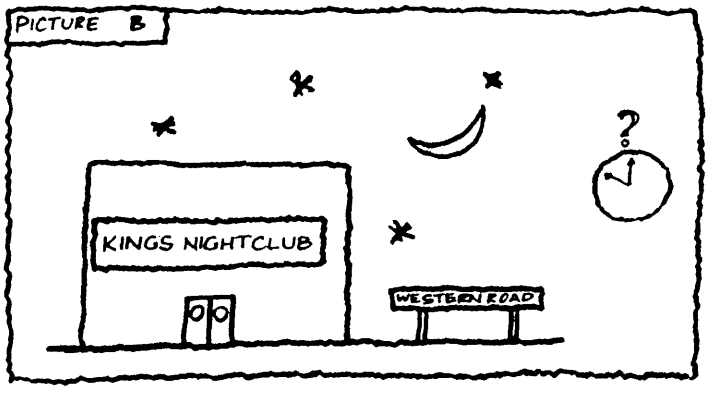
5/. What was she wearing and what was she drinking? *(Look at Picture E)*

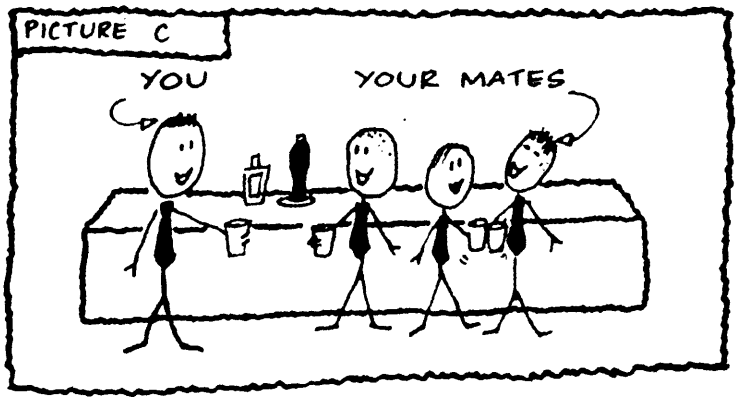
6/. How much had you had to drink? *(Look at Picture F)*

7/. What did you do next? *(Look at Picture G)*

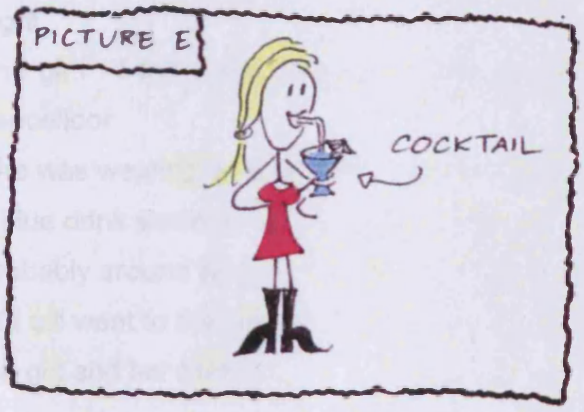
Pictures for the voice donors:



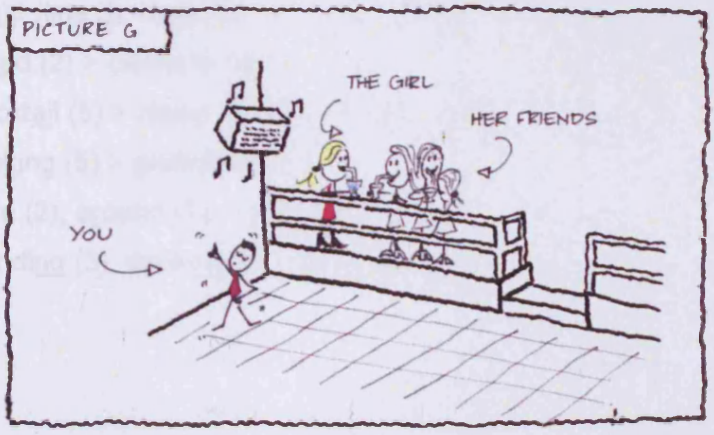





PICTURE D
THE GIRL
HER FRIENDS


PICTURE E
COCKTAIL


PICTURE F
STELLA
+
?


PICTURE G
THE GIRL
HER FRIENDS
YOU

Appendix B: Transcripts of the five voice samples

The transcripts have been annotated to include markers indicating an elongated word (:) or pausing (.). Words which in the researcher's judgement appear particularly distinct for their accent and are possibly salient also to the non-linguists have been highlighted. This highlighting and description of features has been applied as a guide for the reader only and is not intended to indicate an example of a detailed phonetic analysis.

Speaker 1

1 Man U had just beaten Liverpoo:l . two nil
2 Err we went to the Kings nightclub on . Western Road . . we got there around tenish
3 Errm . I was with err: few of my mates and we were standing by the bar for most of the night
4 The gir:l . I think she was with a few of her friends . and she was standing by the dancefloor
5 She was wearing a:: short pink dress .. black knee-high boots .. and she was drinking a blue drink some so it's probably a cocktail
6 Probably around four: or five pints of Stella
7 We a:ll went to the dancefloor and we were stand er dancing around near the:. near to the girl and her friends

Pronunciation features:

[ɑ:] bar (3) dance (4, 7) **[æ:]**

[e] Western (2), friends (4, 7), dress (5) > closer to **[æ]**

[ɜ:] girl (4, 7) > closer to **[ə:]**

[ʌ] club (2) > closer to **[ə]**

[u:] Liverpool (1) > closer to two syllables **[uʷə]**

[aɪ] night (line 2), high (5), by (3, 4), five (6), pints (6) > closer to **[ɔɪ]**

[əʊ] road (2) > closer to **[u:]**

[eɪ] cocktail (5) > closer to two syllables **[eɪə]**

[k] drinking (5) > glottalised **[ʔ]**

[r] there (2), around (2) > alveolar flap **[ɾ]**

[ŋ] standing (3), drinking (5), dancing > **[n]**

Speaker 2

1 Umm err we were celebrating 'cause err Man United beat Liverpool two nil
2 It werr it must've been about ten o'clock and er we went to Kings nightclub umm down on the Western Road
3 Umm, it was just me it was a few of my mates and we were just stood at the bar, having a drink
4 Umm sh she was over on the podium err next to the dancefloor and she was umm with with a couple of her mates
5 Umm I think she was wearing umm like a pink dress and some knee-length boots . umm she must've been drinking a cocktail 'cause it had err an umbrella in it
6 Umm well I must've had umm quite a few pints of Stella by then I think it was about umm four or four or maybe five pints of Stella
7 Umm well we we went across the dancefloor and we having a bit of a dance umm just in front of the girls umm just on the dancefloor

Pronunciation features:

[ɒ] clock (2) > closer to **[ɑ]**

[e] Western (2), friends (4, 7), dress (5) > closer to **[æ]**

[ɑ:] dance (4, 7) **[a:]**

[əʊ] road (2), over (4), podium (4) > closer to **[əu:]**

[aɪ] United (1), nightclub (2), by (6) > closer to **[əɪ]**

[ŋ] length (5) > **[n]**

[ŋ] celebrating (1) having (7) > **[n]**

[h] having (3), her (4), had (5) > not articulated

[r] dancefloor (4), her (4) > rhotic

Speaker 3

1 Well we celebrating cause United beat Liverpool two nil
2 Well. Went to: err Kings nightclub on Western Road umm we got there about ten
3 Well it was just me and some mates umm we were we were standing by the bar
4 She was . umm over the other side . of the dancefloor behind the railings they got there and . she was with some of her mates err three or four of them
5 Well, she'd got she'd got blonde hair . long blonde hair and she was wearing this short pink dress. with long. dark black boots . and she was drinking one of those cocktails
6 I had about four or five pints of Stella . can't really remember
7 Mmwell we went. on the dancefloor..umm we were dancing over near near her and her mates

Pronunciation features:

[aɪ] United (1), nightclub (2) side (4)> closer to **[ʌɪ]**

[eɪ] railings (4) mates, (4, 7) > closer to **[ɑɪ]**

[əʊ] road (2), over (4) > closer to **[ʌʊ]**

[h] her (4), (7), hair (5) > not articulated

[h] in mid position, behind (4) > **[j]**

[ŋ] celebrating (1) standing (3), drinking (5), dancing (7) > **[n]**

[l] postvocalic in final position, Liverpool (1), nil (1) cocktail (5) > labialised **[ö]**

[d] followed by **[g]** she'd got (5) > glottalised and released as **[ʔg]**

[s] followed by **[ð]** railings they (4) > **[ð]** not articulated

[t] followed by **[ð]** got there (2) > glottalised and **[ð]** not articulated

[t] in middle and final position, United (1), got (2, 4, 5) > glottalised **[ʔ]**

[Θ] in initial position, three > **[f]**

[v] followed by **[ð]** of them (4) > **[ð]** not articulated

Speaker 4

1 Because Man U won two nil against Liverpool
2 We went to Kings on Western Road. at about ten
3 I was with three mates and we stood by the bar
4 She was on the balcony..by the dancefloor with three of her mates I think
5 She was wearing a short pink dress..black boots..and drinking a. blue cocktail
6 I don't know. four five pints of Stella
7 We we went up and danced on the dancefloor next to her

Pronunciation features

No RP-divergent features noted

Speaker 5

1 Manchester United had beaten Liverpool two nil and um: we were out celebrating the result
2 Err we went down to the the Kings nightclub on the on the Western Road got to there..probably about ten o'clock
3 Umm with.. three mates..umm we stood at the bar . mainly
4 Err she was over the the other side of the dancefloor on on the sort of a step errm with a few of her friends as well
5 Ummm. she was wearing a: a shortish pink dress.ummm knee-high black boots umm drinking cocktails
6 Err four or five pints of Stella something like that
7 Ummm we went on to the dancefloor for for a dance sort of over in the: corner

Pronunciation features

[aʊ] out (1), down (2), about (2) > closer to **[eʊ]**

[ʌ] result (1) > closer to **[ɒ]**

[ʌ] club (2), other (4) > closer to **[ə]**

[ɪ] mainly (3) > closer to **[eɪ]**

[ɑ:] dance (4, 7) **[a:]**

Appendix C: Questionnaires

Stage 1 Questionnaire

Are you male/female? ***What is your first language?*** ***Age?***

Which town/city/county/region are you from?

1/. Which city or region do you think they are from?

Speaker 1 _____ / Don't know
Speaker 2 _____ / Don't know
Speaker 3 _____ / Don't know
Speaker 4 _____ / Don't know
Speaker 5 _____ / Don't know

2/. Does this person come from your region/city? Please circle your answer:

Speaker 1 _____ Yes / No / Don't know
Speaker 2 _____ Yes / No / Don't know
Speaker 3 _____ Yes / No / Don't know
Speaker 4 _____ Yes / No / Don't know
Speaker 5 _____ Yes / No / Don't know

3/. How old is the speaker? Please circle one age:

Speaker 1 _____ Under 18 / 18-25 / 26-35 / 36-45 / 46-55 / 56-65 / Over 65
Speaker 2 _____ Under 18 / 18-25 / 26-35 / 36-45 / 46-55 / 56-65 / Over 65
Speaker 3 _____ Under 18 / 18-25 / 26-35 / 36-45 / 46-55 / 56-65 / Over 65
Speaker 4 _____ Under 18 / 18-25 / 26-35 / 36-45 / 46-55 / 56-65 / Over 65
Speaker 5 _____ Under 18 / 18-25 / 26-35 / 36-45 / 46-55 / 56-65 / Over 65

4/. How would you describe each speaker's voice?

Speaker 1 _____
Speaker 2 _____
Speaker 3 _____
Speaker 4 _____
Speaker 5 _____

5/. How would you describe each speaker's personality?

Speaker 1 _____
Speaker 2 _____
Speaker 3 _____
Speaker 4 _____
Speaker 5 _____

6/. How do you feel about each speaker? Please circle a number:

Speaker 1	**unintelligent**	1	2	3	**4**	5	6	7	**intelligent**
	dishonest	1	2	3	**4**	5	6	7	**honest**
	aggressive	1	2	3	**4**	5	6	7	**unaggressive**
Speaker 2	**unintelligent**	1	2	3	**4**	5	6	7	**intelligent**
	dishonest	1	2	3	**4**	5	6	7	**honest**
	aggressive	1	2	3	**4**	5	6	7	**unaggressive**
Speaker 3	**unintelligent**	1	2	3	**4**	5	6	7	**intelligent**
	dishonest	1	2	3	**4**	5	6	7	**honest**
	aggressive	1	2	3	**4**	5	6	7	**unaggressive**
Speaker 4	**unintelligent**	1	2	3	**4**	5	6	7	**intelligent**
	dishonest	1	2	3	**4**	5	6	7	**honest**
	aggressive	1	2	3	**4**	5	6	7	**unaggressive**
Speaker 5	**unintelligent**	1	2	3	**4**	5	6	7	**intelligent**
	dishonest	1	2	3	**4**	5	6	7	**honest**
	aggressive	1	2	3	**4**	5	6	7	**unaggressive**

7/. Based on the voices that you've heard, which speaker/s is/are most probably guilty of the crime of assault and robbery, and which speaker/s is/are probably not guilty?

Please circle a number:

Speaker 1	**Probably not guilty**	1	2	3	**4**	5	6	7	**Most probably guilty**
Speaker 2	**Probably not guilty**	1	2	3	**4**	5	6	7	**Most probably guilty**
Speaker 3	**Probably not guilty**	1	2	3	**4**	5	6	7	**Most probably guilty**
Speaker 4	**Probably not guilty**	1	2	3	**4**	5	6	7	**Most probably guilty**
Speaker 5	**Probably not guilty**	1	2	3	**4**	5	6	7	**Most probably guilty**

8/. Please give reasons for your choice:

Most probably guilty because: _____

_____

_____

Probably not guilty because: _____

_____

_____

Stage 2 Questionnaire – Crime version

Section 1 *About you:*

1/. ***Gender*** male /female (please circle)

2/. ***Your age:***

3/. ***Your 1st language:***

4/. ***Which city/region/country are you from?***

5/. ***If you are from the UK, please mark on the map which part of the country you're from by writing "me" in one of the regions:***



6/.If you're from Great Britain, write in the appropriate box where you are from:

For example, if you're from Scotland and from Glasgow, write "me"

Scotland	me
Edinburgh	
Glasgow	me
Other (please say:)	

North Wales	
Bangor	
Wrexham	
Other (please say:)	

South Wales	
Cardiff	
Newport	
Valleys	
Swansea	
Other (please say:)	

North East England	
Newcastle	
Middlesbrough	
Other (please say:)	

North England	
Leeds	
Sheffield	
Other (please say:)	

North West England	
Manchester	
Liverpool	
Other (please say:)	

Midlands	
Birmingham	
Nottingham	
Other (please say:)	

East Anglia	
Norwich	
Ipswich	
Other (please say:)	

South England	
Oxford	
Cambridge	
Other (please say:)	

South East England	
London	
Brighton	
Other (please say:)	

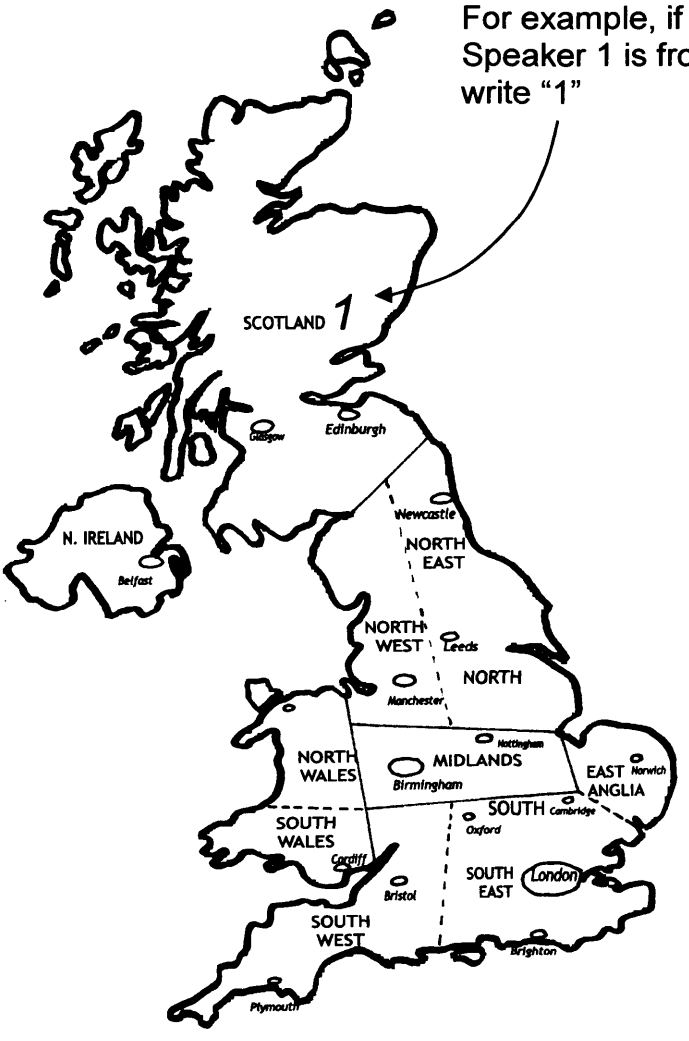
South West England	
Bristol	
Plymouth	
Other (please say:)	

Section 2 *5 Speakers and their accents:*

7/. How would you describe each speaker's accent? Put a circle around the number

		no accent
Speaker 1	**Strong regional accent** | 1 ¦ 2 ¦ 3 | **4** | 5 ¦ 6 ¦ 7 | **Well-spoken**	
	Anything distinctive about the accent?	
		no accent
Speaker 2	**Strong regional accent** | 1 ¦ 2 ¦ 3 | **4** | 5 ¦ 6 ¦ 7 | **Well-spoken**	
	Anything distinctive about the accent?	
		no accent
Speaker 3	**Strong regional accent** | 1 ¦ 2 ¦ 3 | **4** | 5 ¦ 6 ¦ 7 | **Well-spoken**	
	Anything distinctive about the accent?	
		no accent
Speaker 4	**Strong regional accent** | 1 ¦ 2 ¦ 3 | **4** | 5 ¦ 6 ¦ 7 | **Well-spoken**	
	Anything distinctive about the accent?	
		no accent
Speaker 5	**Strong regional accent** | 1 ¦ 2 ¦ 3 | **4** | 5 ¦ 6 ¦ 7 | **Well-spoken**	
	Anything distinctive about the accent?	

8/. If you think you know where the speakers are from, write their speaker number 1,2,3,4, or 5 on the map in the place they might be from.



9/. Can you be more specific? If you think you can be more exact, write the speakers' numbers in the appropriate boxes below

For example, if you think Speaker 1 is from Scotland and from Glasgow, write "1"

Scotland	1
Edinburgh	
Glasgow	1
Other (please say:)	

North Wales	
Bangor	
Wrexham	
Other (please say:)	

South Wales	
Cardiff	
Newport	
Valleys	
Swansea	
Other (please say:)	

North East England	
Newcastle	
Middlesbrough	
Other (please say:)	

North England	
Leeds	
Sheffield	
Other (please say:)	

North West England	
Manchester	
Liverpool	
Other (please say:)	

Midlands	
Birmingham	
Nottingham	
Other (please say:)	

East Anglia	
Norwich	
Ipswich	
Other (please say:)	

South England	
Oxford	
Cambridge	
Other (please say:)	

South East England	
London	
Brighton	
Other (please say:)	

South West England	
Bristol	
Plymouth	
Other (please say:)	

Section 3 *The speakers' voices – listen again:*

11/. Choose any of the following categories to describe the voice- you don't have to choose them all! Put a circle round the number

Speaker 1

Slow	1	2	3	**4**	5	6	7	**Fast**
Hesitant	1	2	3	**4**	5	6	7	**Confident**
Deep voice	1	2	3	**4**	5	6	7	**High pitched**
Monotone	1	2	3	**4**	5	6	7	**Melodious**
Soft	1	2	3	**4**	5	6	7	**Loud**
Shaky	1	2	3	**4**	5	6	7	**Steady**

Anything distinctive about the voice? ❑ dry ❑ whispery ❑ sharp ❑ shrill ❑ effeminate ❑ rough ❑ croaky ❑ raspy ❑ anything else?

Any speech defects? ❑ lisp ❑ stammer ❑ stutter ❑ anything else?

Does it sound like anyone famous/anyone that you know?

Speaker 2

Slow	1	2	3	**4**	5	6	7	**Fast**
Hesitant	1	2	3	**4**	5	6	7	**Confident**
Deep voice	1	2	3	**4**	5	6	7	**High pitched**
Monotone	1	2	3	**4**	5	6	7	**Melodious**
Soft	1	2	3	**4**	5	6	7	**Loud**
Shaky	1	2	3	**4**	5	6	7	**Steady**

Anything distinctive about the voice? ❑ dry ❑ whispery ❑ sharp ❑ shrill ❑ effeminate ❑ rough ❑ croaky ❑ raspy ❑ anything else?

Any speech defects? ❑ lisp ❑ stammer ❑ stutter ❑ anything else?

Does it sound like anyone famous/anyone that you know?

Speaker 3

Slow	1	2	3	**4**	5	6	7	**Fast**
Hesitant	1	2	3	**4**	5	6	7	**Confident**
Deep voice	1	2	3	**4**	5	6	7	**High pitched**
Monotone	1	2	3	**4**	5	6	7	**Melodious**
Soft	1	2	3	**4**	5	6	7	**Loud**
Shaky	1	2	3	**4**	5	6	7	**Steady**

Anything distinctive about the voice? ❑ dry ❑ whispery ❑ sharp ❑ shrill ❑ effeminate ❑ rough ❑ croaky ❑ raspy ❑ anything else?

Any speech defects? ❑ lisp ❑ stammer ❑ stutter ❑ anything else?

Does it sound like anyone famous/anyone that you know?

Speaker 4

Slow	1	2	3	**4**	5	6	7	**Fast**
Hesitant	1	2	3	**4**	5	6	7	**Confident**
Deep voice	1	2	3	**4**	5	6	7	**High pitched**
Monotone	1	2	3	**4**	5	6	7	**Melodious**
Soft	1	2	3	**4**	5	6	7	**Loud**
Shaky	1	2	3	**4**	5	6	7	**Steady**

Anything distinctive about the voice? ❑ dry ❑ whispery ❑ sharp ❑ shrill ❑ effeminate ❑ rough ❑ croaky ❑ raspy ❑ anything else?

Any speech defects? ❑ lisp ❑ stammer ❑ stutter ❑ anything else?

Does it sound like anyone famous/anyone that you know?

Speaker 5

Slow	1	2	3	**4**	5	6	7	**Fast**
Hesitant	1	2	3	**4**	5	6	7	**Confident**
Deep voice	1	2	3	**4**	5	6	7	**High pitched**
Monotone	1	2	3	**4**	5	6	7	**Melodious**
Soft	1	2	3	**4**	5	6	7	**Loud**
Shaky	1	2	3	**4**	5	6	7	**Steady**

Anything distinctive about the voice? ❑ dry ❑ whispery ❑ sharp ❑ shrill ❑ effeminate ❑ rough ❑ croaky ❑ raspy ❑ anything else?

Any speech defects? ❑ lisp ❑ stammer ❑ stutter ❑ anything else?

Does it sound like anyone famous/anyone that you know?

Section 4 *Your feelings towards the speakers*

12/. Finally, how do you feel towards each speaker? Put a circle around the appropriate number

Speaker 1	**probably guilty**	1	2	3	**4**	5	6	7	**probably not guilty**
Speaker 2	**probably guilty**	1	2	3	**4**	5	6	7	**probably not guilty**
Speaker 3	**probably guilty**	1	2	3	**4**	5	6	7	**probably not guilty**
Speaker 4	**probably guilty**	1	2	3	**4**	5	6	7	**probably not guilty**
Speaker 5	**probably guilty**	1	2	3	**4**	5	6	7	**probably not guilty**

13/. Please give reasons for your choice:

Most probably guilty because: _____

_____

Probably not guilty because: _____

_____

Stage 2 Questionnaire – 'Got the girl version' version

As the 'Crime version', but with a changed Section 4:

Section 4 *Your feelings towards the speakers*

Speaker 1	**Probably got the girl**	1	2	3	**4**	5	6	7	**Probably didn't get the girl**
Speaker 2	**Probably got the girl**	1	2	3	**4**	5	6	7	**Probably didn't get the girl**
Speaker 3	**Probably got the girl**	1	2	3	**4**	5	6	7	**Probably didn't get the girl**
Speaker 4	**Probably got the girl**	1	2	3	**4**	5	6	7	**Probably didn't get the girl**
Speaker 5	**Probably got the girl**	1	2	3	**4**	5	6	7	**Probably didn't get the girl**

13/. Please give reasons for your choice:

Most probably got the girl because: _____

_____

Probably didn't get the girl because: _____

_____

Appendix D: Qualitative voice descriptions from Stage 1

Speaker 1

Speaker No.	Comment	Speaker No.	Comment
1	accent	1	deep
1	aggressive	1	deep
1	aggressive	1	deep
1	aggressive	1	deep
1	apparent accent	1	deep
1	assertive	1	deep
1	assertive tone	1	deep
1	assured	1	deep
1	bit boring	1	deep
1	bit common	1	deep
1	bloke/lad	1	deep
1	bold	1	deep
1	boring	1	deep
1	boring	1	deep
1	boring, uninteresting	1	deep
1	Bristol	1	deep
1	brusque and husky	1	deep
1	calm	1	deep
1	calm	1	deep
1	calm	1	deep
1	Cardiff	1	deep
1	Cardiff	1	deep
1	Cardiff	1	deep
1	Cardiff dialect	1	deep
1	careful spoken	1	deep
1	casual	1	deep
1	clear	1	deep
1	clear	1	deep and husky
1	clear	1	deep and low
1	clear and loud	1	deep, masculine
1	clear voice	1	definite
1	confidence	1	detailed
1	confident	1	dialect
1	confident	1	dialect
1	confident	1	different tones
1	confident	1	docile
1	confident	1	drawn out
1	confident	1	drones on and on at same pitch
1	confident	1	dull
1	confident	1	dull
1	confident	1	dull
1	confident	1	dullish
1	confident and closed	1	easy to listen to
1	confident tone	1	expressionless
1	confident tone	1	fairly monotone
1	confident voice	1	flat, monotone
1	confident, calm	1	flippant
1	contradicts himself	1	harsh

1	has to think what he's saying
1	hesitant
1	hint of accent
1	hoarse
1	honest
1	lack of tone
1	laddish
1	laid back
1	lengthened emphasis on words
1	long
1	lots of pauses between words
1	low
1	low
1	low, manly
1	lower tone
1	manly
1	masculine
1	mature
1	middle-aged
1	matter of fact
1	mature
1	monotone
1	monotonous
1	monotonous
1	monotonous
1	monotonous
1	monotonous
1	monotonous
1	more mature
1	nasty
1	nervous, husky
1	Newport
1	nice
1	no distinct features
1	non-standard
1	non-standard, slow, Welsh
1	not much change in pitch
1	older
1	older
1	older
1	older
1	one dimensional
1	overly masculine
1	precise detail
1	quite deep
1	quite monotonous
1	quite well-spoken
1	regional accent
1	regional accent
1	regional 'twang'
1	relaxed
1	relaxed
1	relaxed
1	loud
1	loud
1	loud
1	loud and assertive
1	loud and confident
1	loud and confident
1	low
1	relaxed, slow
1	rises at end of sentences
1	rough
1	roughish
1	self-assured
1	sharp, brief
1	short bursts
1	similar to 1 but stronger
1	slight accent
1	slow
1	slow
1	slow
1	slow
1	slow
1	slow
1	slow
1	slow
1	slow
1	slow & clear
1	slow and boring
1	slow and deliberate
1	slow, thinking about answer
1	small range of intonation
1	soft
1	stable
1	Standard
1	standardised speech
1	stresses the vowels (a)
1	strong
1	strong accent
1	strong accent
1	strong accent
1	strong accent
1	strong and forceful
1	strong regional accent
1	strong regional accent
1	strong, confident
1	sure of what he's saying
1	trying to sound convincing
1	unsure
1	very confident speech
1	very confident, no shakiness

1	Welsh
1	Welsh
1	Welsh
1	Welsh
1	Welsh
1	Welsh
1	Welsh
1	Welsh accent (Cardiff?)
1	Welsh/West accent
1	young

Speaker 2

Speaker No.	Comment	Speaker No.	Comment
2	average	2	genuine
2	average sounding	2	genuine
2	bit slow	2	girly voice
2	Bristol	2	grating
2	broad accent (Plymouth?)	2	hesitant
2	calm, unaggressive	2	hesitant
2	chirpy	2	hesitant
2	clear	2	hesitant
2	clear	2	hesitant
2	clear	2	hesitant
2	coarse	2	hesitant
2	colloquialised	2	hesitant
2	common	2	hesitant
2	common	2	hesitant and nervous
2	confident	2	hesitant and nervous
2	confident	2	high
2	creepy	2	high
2	croaky	2	high pitched
2	deep	2	high pitched
2	different tones	2	high pitched
2	doesn't seem bothered	2	hint of accent
2	drinkin'	2	innocent voice
2	dry	2	insecure
2	dry and low	2	less certain (intonation)
2	dull	2	light
2	easy going tone	2	light
2	easy to listen to	2	like a boy's voice
2	emotional	2	little nervous
2	expressive	2	looking for assurance
2	fairly quiet	2	lots of intonation
2	farmer	2	lots of pauses (hmms)
2	farmer accent	2	medium pitch
2	farmer accent - Somerset	2	medium pitch
2	farmer twang	2	medium pitch
2	feminine	2	monotonous
2	friendly	2	monotonous
2	friendly and casual	2	monotonous
2	frightened	2	most facts well described
2	full of pauses	2	nervous
2	genuine	2	nervous

2	non-standard, saying 'in for 'ing'
2	not city based
2	not clear to understand
2	not clear, maybe a lisp
2	not flowing, very hesitant
2	not much intonation
2	not RP
2	not very calm
2	not very clear
2	on the same level
2	polite
2	questions himself
2	quick speech
2	quiet
2	quiet
2	quirky
2	quite accented
2	quite calm
2	quite common
2	quite deep
2	quite fast
2	quite slow placed
2	regional
2	regional accent
2	regional accent
2	regional accent
2	regional accent
2	repeats himself
2	respectful
2	rising intonation
2	rising intonation
2	rising intonation at end of sentences
2	rural
2	rural
2	rural
2	same tone
2	shaky
2	shaky and scared
2	slightly animated
2	slightly hesitant
2	slow
2	slow
2	slow
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	nervous
2	non-aggressive
2	non-standard
2	soft
2	soft
2	soft
2	soft
2	softer, nervous
2	soft-spoken
2	sounds more comfortable
2	sounds young
2	South quite well spoken
2	South West
2	South West
2	Southern
2	stammer
2	strong accent
2	strong accent
2	stumbling
2	stutter
2	stuttering
2	stuttering
2	stuttery
2	telling the truth
2	tentative
2	tentative
2	tentative
2	timid
2	truthful
2	truthful
2	typical farmer accent
2	unaggressive
2	uneasy, nervous, emotional
2	unsteady
2	unsteady
2	unsure
2	unsure
2	unsure
2	unsure
2	unsure
2	unsure
2	unsure
2	unsure
2	unsure, bit nervous
2	urban
2	urban accent
2	varied tone
2	very unsteady
2	wary
2	wary
2	West Country
2	West Country
2	West Country

2	western tints
2	worried
2	young
2	young
2	young
2	young

Speaker 3

Speaker No.	Comment	Speaker No.	Comment
3	average	3	distinct accent
3	bit husky	3	doesn't pronounce 'h's
3	blokish, but not deep	3	doesn't pronounce words to Standard English
3	calm and complacent	3	drawling
3	calm on the surface, but nervous underneath	3	drops letters
3	casual	3	drops letters
3	chirpy	3	dull
3	chirpy	3	dull
3	clear	3	dull and monotonous
3	clear	3	edgy
3	clear	3	edgy
3	clear and loud	3	emotional
3	clear sounding	3	Estuary, clipped
3	cockney	3	excitable
3	cockney	3	expressive
3	cockney	3	fairly monotone
3	cockney	3	fast speech
3	cockney	3	fast speech
3	cockney	3	fast, alert, detailed
3	cockney	3	faster
3	cockney	3	faster-paced
3	cocky	3	friendly and informal
3	colloquial	3	gentle
3	common	3	guilty?
3	common	3	harsh
3	common	3	hesitant
3	common accent	3	hesitant
3	Common, 'alright mate?' accent	3	hesitant
3	confident	3	hesitant
3	confident	3	hesitant and shaky
3	confident	3	high pitched
3	confident	3	high pitched
3	confident	3	higher
3	deep	3	higher pitch
3	deep	3	higher than norm
3	deep	3	high-pitched
3	deep	3	immature
3	deep accent	3	jumpy
3	delayed	3	knows order of events + order well
3	dialect of a group of lads	3	laddish
3	disrupted	3	laddish

3	laddish
3	laddish, vernacular
3	laid back
3	less certain
3	light
3	light
3	lively
3	London
3	loud
3	low
3	low
3	masculine
3	masculine
3	medium pitch
3	misses 'g' off -ing
3	misses 'h' in 'hair'
3	monotone
3	monotone
3	monotone and boring
3	more innocent
3	nervous
3	nervous
3	nervous
3	nervous
3	nervous
3	nervous
3	nervous
3	nervous
3	nervous
3	nervous of the situation
3	no stuttering
3	non-standard
3	non-standard
3	non-standard, nervous
3	normal
3	not confident, worried
3	not posh
3	not so pleasant
3	not sure about his story
3	not very articulate
3	not very calm
3	not well spoken
3	not well spoken
3	not well spoken 'er' instead of 'her'
3	on edge
3	pauses a few times
3	pikey
3	quick
3	quick tempo
3	quick, jumpy
3	quiet
3	quiet
3	quite cockney and common
3	quite slow
3	regional
3	regional
3	regional accent
3	regional accent
3	regional accent
3	relaxed
3	repetitive
3	rough around the edges, 'common'
3	rough sounding
3	roughish
3	shaky, sounds nervous
3	slang
3	slang, cockney
3	slangy
3	slightly nervous
3	nervous but reasonably confident
3	slightly unsure
3	interested in the conversation
3	sounds young
3	Southern
3	strong accent
3	strong accent
3	strong accent
3	strong accent
3	stupid
3	thick
3	touch of Cockney
3	unemotional
3	unsure
3	unsure
3	unsure
3	unsure
3	unsure
3	unsure at first
3	unsure, pauses
3	upbeat
3	uses a lot of fillers
3	uses colloquial language
3	varied pitch
3	vernacular
3	very accented
3	very common
3	very regional
3	young
3	young
3	young
3	young
3	young
3	young
3	young
3	young London
3	young -sounding
3	younger sounding
3	youngest of all

Speaker 4

Speaker No.	Comment	Speaker No.	Comment
4	aggressive tone	4	deep
4	almost as if he's reading off paper	4	deep
4	assertive	4	deep
4	assured	4	deep
4	bit husky	4	deep
4	bit intimidating	4	deep
4	boring	4	deep
4	boring	4	deep
4	boring	4	deep
4	boring	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm	4	deep
4	calm and cool	4	deep and husky
4	calm, unaggressive	4	deep and rough
4	careful	4	deep tone
4	clear	4	deep tone
4	clear	4	deep voice
4	clear	4	deliberately slow
4	clear	4	dull
4	clear	4	dull
4	clear	4	dull
4	clear	4	dull
4	clear	4	dull
4	clear	4	dull and boring tone
4	clearly spoken	4	edgy
4	complacent	4	eloquent
4	concise	4	emphasising
4	confident	4	even tone
4	confident	4	expressionless
4	confident	4	fairly quiet
4	confident	4	few changes in pitch
4	confident with answers, calm	4	flowing
4	consistent	4	fluent
4	creepy	4	genuine
4	dark	4	guilty voice
4	deep	4	hard to hear
4	deep	4	hard to understand
4	deep	4	harsh
4	deep	4	harsh, rough voice
4	deep	4	hesitant
4	deep	4	knows his story
4	deep	4	lack of intonation
4	deep	4	laid back voice
4	deep	4	less regional than the others
4	deep	4	like he's reading off a paper

4	low
4	low
4	low
4	low
4	low tone
4	low tone
4	low voice
4	low, sexy, mature
4	manly
4	middle class
4	Middle Class, RP
4	monotone
4	monotone
4	monotone
4	monotone
4	monotone
4	monotone
4	monotone
4	monotone
4	monotone, robotic
4	monotonous
4	monotonous
4	monotonous
4	monotonous
4	muffled
4	muffled
4	mumbles
4	no accent
4	no accent
4	no clear accent
4	no regional accent
4	no strong accent
4	no strong accent
4	non-descriptive
4	not confident
4	not confident
4	not expressive
4	not hesitant
4	not nervous
4	not nervous
4	not rough sounding
4	not so confident, worrying
4	plain
4	plain
4	plain
4	plain
4	posh
4	posh
4	posh
4	posh
4	posh
4	posh public school boy
4	poshest
4	prepared, badly delivered, too confident
4	pronounces well
4	proper
4	quiet
4	quiet
4	quiet
4	quiet
4	quiet
4	quite creepy
4	quite posh
4	relaxed
4	rises at end of sentences
4	scared, but sure
4	serious
4	serious
4	serious
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow
4	slow but decisive
4	slow speaking
4	slowed down version of number 5
4	slurred
4	soft
4	soft voice
4	sombre, no variation
4	sounded unsure of the details
4	standard accent and dialect
4	Standard English
4	Standard English
4	strong
4	sure
4	sure
4	taking his time
4	tired
4	toneless
4	unassertive
4	unclear
4	unemotional

4	trying to come across as unworried
4	uninspiring
4	unsure
4	unsure
4	unsure
4	vague
4	very deep voice
4	very descriptive
4	very slow
4	well-educated
4	well-educated
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	well-spoken
4	whistling voice
4	young

Speaker 5

Speaker No.	Comment	Speaker No.	Comment
5	animated	5	changes in pitch of voice
5	arrogant	5	clear
5	average	5	clear
5	average sounding	5	clear
5	bit slimey	5	clear
5	boring and monotone	5	clear
5	boring voice	5	clear
5	calm	5	clear
5	calm	5	clear and confident
5	calm	5	clear and confident
5	calm	5	clear speaker
5	calm	5	coherent
5	calm	5	common
5	calm and thoughtful	5	confident
5	calm and unflustered	5	confident
5	calm, assertive	5	confident
5	careful	5	confident
5	careful speech	5	confident
5	casual	5	confident
5	casual speech	5	confident
5	certain	5	confident tone

5	confident tone
5	confident, calm
5	confident, knows what he's talking about
5	concise, but lots of 'umms'
5	controlled
5	deep
5	deep
5	disjointed
5	doesn't seem bothered
5	drowsy
5	dull
5	easy to listen to
5	easy to listen to
5	easy to listen to
5	easy to listen to
5	easy to understand
5	expressive
5	fairly dull
5	farmer-like
5	flowing
5	fluid, quite understandable
5	friendly
5	friendly
5	gentle
5	goes up and down
5	gravely
5	hesitant
5	hesitant
5	hesitant
5	hesitant
5	hesitant
5	hesitant
5	hesitant but out of choice - unbothered
5	hesitant, lots of 'umms'
5	higher class
5	honest
5	laddish
5	not aggressive
5	not mean
5	not particularly deep
5	not too sure, pauses
5	not very strong accent
5	not very worried
5	polite voice
5	posh + perfect
5	posher
5	questioning
5	quick
5	quite calm
5	quite confident
5	quite confident, but unsure
5	quite fragmented
5	quite pleasant
5	quite posh, no accent
5	quite RP
5	quite soothing
5	quite well-spoken
5	quite well-spoken
5	regional accent
5	relatively standard
5	relaxed
5	same tone
5	self-assured
5	sinister
5	slightly well-spoken
5	slow
5	slow speaking
5	smooth, middle class, eloquent
5	sounds vague
5	lively
5	Londoner
5	lots of umms
5	loud and confident
5	matter of fact
5	mature
5	mature gentleman
5	medium pitch
5	medium tone
5	middle class
5	middle class
5	Midlands, but not strong
5	moderately well-spoken
5	more like RP
5	more standard
5	nervous
5	nervous
5	nervous
5	nervous, due to speed
5	nice
5	nice
5	nice pronunciation
5	no accent
5	no accent
5	no accent
5	no accent
5	no accent
5	no accent
5	no apparent accent
5	non-hesitant
5	non-standard
5	normal
5	normal

5	Southern
5	Southern English
5	Standard
5	standard - middle class
5	Standard English
5	standard variety
5	strange intonation
5	stuttering
5	stutters
5	sure of himself
5	thoughtful and clear
5	trying hard to be unrehearsed
5	uncertain
5	uncertain
5	uncertain and nervous
5	undisturbed
5	uninterested
5	unoffensive
5	unsure
5	unthreatening
5	vague
5	vague
5	weak memory
5	well-projected
5	well-pronounced
5	well-spoken
5	well-spoken
5	well-spoken
5	normal
5	normal pitch
5	well-spoken
5	well-spoken
5	well-spoken, standard
5	young

