Lingnan University

Digital Commons @ Lingnan University

Theses & Dissertations

Department of English

2010

Phonological features of Hong Kong English : patterns of variation and effects on local acceptability

Andrew John SEWELL

Follow this and additional works at: https://commons.ln.edu.hk/eng_etd

Part of the [English Language and Literature Commons](#)

Recommended Citation

Sewell, A. J. (2010). Phonological features of Hong Kong English: Patterns of variation and effects on local acceptability (Doctor's thesis, Lingnan University, Hong Kong). Retrieved from http://dx.doi.org/10.14793/eng_etd.1

This Thesis is brought to you for free and open access by the Department of English at Digital Commons @ Lingnan University. It has been accepted for inclusion in Theses & Dissertations by an authorized administrator of Digital Commons @ Lingnan University.

Terms of Use

The copyright of this thesis is owned by its author. Any reproduction, adaptation, distribution or dissemination of this thesis without express authorization is strictly prohibited.

All rights reserved.

**PHONOLOGICAL FEATURES OF HONG KONG ENGLISH:
PATTERNS OF VARIATION AND EFFECTS ON LOCAL
ACCEPTABILITY**

ANDREW JOHN SEWELL

PhD

LINGNAN UNIVERSITY

2010

**PHONOLOGICAL FEATURES OF HONG KONG ENGLISH:
PATTERNS OF VARIATION AND EFFECTS ON LOCAL
ACCEPTABILITY**

by

SEWELL, Andrew John

**A thesis submitted in partial fulfillment
of the requirements for the Degree of
Doctor of Philosophy in English**

Lingnan University

2010

ABSTRACT

Phonological Features of Hong Kong English: Patterns of Variation and Effects on Local Acceptability

by

SEWELL, Andrew John

Doctor of Philosophy

The changing dynamics of international communication in English have led to a intense questioning of the relevance of native-speaker pronunciation models in language teaching and testing. In addition, the World Englishes approach to local varieties has increased their level of recognition. Both of these developments suggest that English pronunciation models need to be reviewed, and Hong Kong represents an interesting case study. Although it has been claimed that Hong Kong English is at the ‘nativization’ stage, the existence of exonormative attitudes towards English is also well known. Two important questions arise from this inherent tension, neither of which has been intensively addressed in previous studies. Firstly, although many of the features of Hong Kong English pronunciation have been described, patterns of inter-speaker variation have not been investigated in detail. Secondly, the attitudes of Hong Kong English users towards the phonological features of their own variety have not been studied in ways that take account of such variation.

This dissertation addresses both of these questions by being features-based in approach and using local listeners to evaluate accent samples. After an initial review of the features of Hong Kong English pronunciation, a preliminary study surveys the occurrence of consonantal phonological features within a mini-corpus of speech samples taken from local television programmes. Its findings are presented in the form of an implicational scale, which not only shows the relative frequencies with which different features occurred, but also indicates the existence of implicational patterns of co-occurrence. In the main study, twelve authentic accent samples (eleven Hong Kong speakers and one British speaker) were presented to 52 first-year undergraduate students for evaluation as to their acceptability, defined here as acceptability for pedagogical purposes.

Multivariate statistical analysis discovered firstly that phonological ‘errors’, as marked by the student listeners, were the most important measured factor in determining the acceptability scores, and secondly that only certain types of ‘error’ or ‘feature’ had significant effects. These features were either related to L1 transfer or involved other salient phenomena such as idiosyncratic alterations to syllable structure. The explanatory part of the study includes acceptability as one of the factors determining feature persistence, in an ‘ecological’ or ‘evolutionary’ model of

L2 phonology acquisition and development that combines the findings of the preliminary and main studies. Among the other factors that determine feature persistence or disappearance, salience, intelligibility and markedness are invoked as important influences.

The acceptability data also has pedagogical implications, in that local listeners did not give the British accent the highest acceptability rating. This contrasts with the findings of previous studies regarding the pedagogical acceptability of the Hong Kong English accent. However, the features-based approach indicates that only certain types of local accent were acceptable to these listeners, and that these accents were more, rather than less, 'native-like'. In various ways, the study contributes to an understanding of accent variation and acceptability within a new variety of English.

DECLARATION

I declare that this is an original work based primarily on my own research, and I warrant that all citations of previous research, published or unpublished, have been duly acknowledged.

(Andrew John Sewell)
Date:

CONTENTS

	List of Tables	vi
	List of Figures	viii
	Acknowledgements	xi
1	Introduction	1
1.1	Topic and context	1
	1.1.1 Justification and general research orientation	4
1.2	Aims and methods of the study	5
	1.2.1 Aims	5
	1.2.2 Methodological approach	6
	1.2.3 Explanatory approach	7
1.3	Terms and concepts employed in the study	8
	1.3.1 General	8
	1.3.2 Notation, symbols and variables	13
1.4	Structure	14
2	Literature Review	16
2.1	Introduction	16
2.2	The current situation in pronunciation teaching	16
	2.2.1 Native speaker models	16
	2.2.2 Criticisms of NS models: RP	18
	2.2.3 General criticisms of NS models	20
2.3	Possible solutions and alternatives	21
	2.3.1 English as a Lingua Franca (ELF)	22
	2.3.2 The Lingua Franca Core (LFC)	23
	2.3.3 The LFC and its relationship with ELF and World Englishes	25
2.4	The evaluation framework	27
2.5	Quadrant 1: linguistic, user-internal factors	29
2.6	Quadrant 2: linguistic, user-external factors	32
	2.6.1 Intelligibility	32
	2.6.2 Sources of intelligibility problems	34
	2.6.3 Features contributing to intelligibility	36
	2.6.3.1 The avoidance of vowel reduction and word stress	36
	2.6.3.2 The avoidance of vowel reduction in function words	39
	2.6.3.4 The avoidance of vowel reduction and prosodic factors	40

2.6.4	Other studies of intelligibility	44
2.6.5	Other linguistic factors: functional load and frequency	47
2.6.6	Functional factors, synchronic variation and diachronic change	48
2.6.7	Summary: the predictability of intelligibility	50
2.7	Quadrant 3: non-linguistic, user-internal factors	54
2.8	Quadrant 4: non-linguistic, user-external factors	57
2.8.1	Stigmatisation	60
2.9	A features-based evaluation of the LFC proposals	62
2.9.1	The dental fricatives	63
2.9.2	L vocalisation	69
2.9.3	Consonant clusters	71
2.9.4	Other features: consonant substitutions	72
2.9.5	Other features: phonetic characteristics	73
2.9.6	Vowel quality	75
2.10	General pedagogical factors	77
2.10.1	Models and goals	77
2.11	Ideology and pedagogy	79
2.11.1	Contested terminology	81
2.11.2	Understanding the debate: ELF and globalization	82
2.12	Summary	83
3	The Hong Kong Background and the Preliminary Study	84
3.1	Introduction	84
3.2	Language in education	84
3.2.1	Other issues in language education: declining standards?	86
3.3	The role of pronunciation teaching in Hong Kong	86
3.3.1	Pronunciation models	87
3.4	Hong Kong English: status and attitudes	90
3.4.1	Assessing variety status	90
3.4.2	Attitudes and identity	92
3.4.3	HKE accent studies	94
3.5	The phonology of Hong Kong English	96
3.5.1	Two previous studies compared	96
3.5.2	The variation problem	101
3.6	The preliminary study	104
3.6.1	The HKE mini-corpus	105
3.6.2	Consonantal features of HKE	106

3.6.3	Data analysis	110
3.6.4	Implicational scaling	113
3.6.5	Explaining the implicational patterns	117
3.7	The research space	120
3.7.1	Research questions	121
4	Methodology	122
4.1	Introduction	122
4.2	Previous studies	123
4.2.1	Standardised reading passages	123
4.2.2	The matched-guise technique	124
4.2.3	Authentic data	126
4.3	Sample selection	126
4.3.1	Advantages of broadcast material	127
4.3.2	Disadvantages of broadcast material	128
4.4	Selecting the survey samples	130
4.4.1	Linguistic preselection	135
4.5	Other speaker variables	137
4.5.1	Lexical and syntactic complexity	138
4.5.2	Fluency and speech rate	142
4.5.3	Prosodic factors: pitch span	144
4.6	The survey respondents	148
4.7	Survey design	149
4.7.1	The concept of acceptability revisited	149
4.7.2	Part 1 of the survey form	150
4.7.3	Part 2 of the survey form	152
4.8	Administering the survey	154
5	Data Analysis and Findings	155
5.1	Introduction	155
5.2	Part 1 data	155
5.2.1	Internal consistency	156
5.2.2	Inter-rater reliability	158
5.2.3	Inter-item correlations	156
5.2.4	The acceptability ratings	162
5.3	The effects of speaker variables	164
5.3.1	Phonological accuracy	165
5.3.2	Correlations between the variables	166

5.3.3	Regression analysis	167
5.4	Student error codings	173
5.4.1	Error codings and categories	173
5.4.2	Error categories	177
5.4.2.1	Vowel-related errors: SYLL, VOWEL	180
5.4.2.2	Consonantal features: PHONSUB, OTHER C-SUB	184
5.4.2.3	Other consonantal features: TH-STOP, TH-FRONT, L-VOCAL	187
5.4.2.4	Consonant clusters: CCRF, CCM-I	190
5.4.2.5	Connected speech: CS-LINK	192
5.4.3	The effects of the features on the acceptability ratings	193
5.5	Student comments and individual speaker analysis	200
5.5.1	Speaker 1	201
5.5.2	Speaker 6	204
5.5.3	Speaker 2	207
5.5.4	Speaker 3	209
5.5.5	Speaker 4	212
5.5.6	Speaker 5	214
5.5.7	Speaker 7	216
5.5.8	Speaker 8	218
5.5.9	Speaker 9	220
5.5.10	Speaker 10	222
5.5.11	Speaker 11	225
5.5.12	Speaker 12	227
5.6	Summary of the findings	229
6	Discussion	232
6.2	Towards an explanatory model	233
6.2.1	The intelligibility-acceptability correspondence	233
6.2.2	Synchronic perspectives: the salience of features	238
6.2.3	Salience and stigmatisation	242
6.3	A combined approach: a dynamic or evolutionary model	244
6.4	The phonological features from an evolutionary perspective	248
6.4.1	SYLL	249
6.4.2	CCRF-PV and CCRF-PC	249
6.4.3	PHONSUB-V	250
6.4.4	VOWELSUB	252
6.4.5	OTHER C-SUB	254
6.4.6	Non-significant features	254
6.5	General summary	258

7	Conclusions	264
7.1	Introduction	264
7.2	Research questions	264
	7.2.1 Research question 1	264
	7.2.2 Research question 2	266
	7.2.3 Research question 3	267
7.3	Possible candidates for acceptance	269
	7.3.1 TH stopping	269
	7.3.2 L vocalisation	272
	7.3.3 Consonant cluster simplification	274
	7.3.4 Full vowels and the avoidance of schwa	275
	7.3.5 Other possibilities	275
7.4	Recommendations and implications	276
	7.4.1 Teaching pronunciation	276
	7.4.2 Testing pronunciation	279
	7.4.3 The models debate	280
	7.4.4 Implications for the description of NVEs	285
	7.4.5 Implications for studies of ELF (English as a Lingua Franca)	288
7.5	Overall review of the study	290
	7.5.1 Limitations of the study	290
	7.5.2 Achievements of the study	295
	7.5.3 Directions for future research	297
	7.5.4 Final words	298
	References	300
	Appendix 1: details of speakers and speech samples in the Hong Kong English mini-corpus	322
	Appendix 2a: Part 1 of the survey form used in the study	323
	Appendix 2b: Part 2 of the survey form used in the study	324
	Appendix 3: correlation matrix for the thirteen speaker variables	325

LIST OF TABLES

<i>Table</i>	<i>Title</i>	<i>Page</i>
1.1	Lexical sets (Wells 1982) referred to in the study	14
3.1	‘At the benchmark’ descriptors for pronunciation from two tests	88
3.2	The percentage of word tokens containing variants of /θ/	99
3.3	Similarities and differences in two descriptions of HKE phonology	101
3.4	HKE consonantal features considered in the preliminary study	107
3.5	Number of word tokens and average number of tokens per speaker, distribution of features across speakers, and rate of agreement for each feature category	112
3.6	The implicational scale of HKE consonantal features	114
4.1	Speaker biodata and source of recording	132
4.2	Transcripts of the twelve accent samples	134
4.3	Implicational scale for the twelve accent samples	135
4.4	Lexical complexity measures	139
4.5	Syntactic complexity measures	141
4.6	Fluency factors	144
4.7	Pitch span measurements for the ten male speakers	145
4.8	Details of the three groups of respondents	148
5.1	SPSS output table for Cronbach’s alpha coefficient (six items)	157
5.2	SPSS output table for Cronbach’s alpha coefficient (five items)	158
5.3	Intraclass correlation coefficients according to gender, major and self-rating of pronunciation skills	159
5.4	Inter-item Pearson correlation matrix	160

5.5	The twelve speakers ranked by overall acceptability scores (ranking 1) and direct acceptability scores (ranking 2)	163
5.6	Measures of phonological accuracy	166
5.7	The thirteen speaker variables selected for analysis	166
5.8	Correlation matrix for the six speaker variables	169
5.9	Model summary and coefficients for regression analysis of the six speaker variables	171
5.10	Error categories used in the study	178
5.11	The SPSS model summary output table	196
5.12	The SPSS regression coefficient output table	196
5.13	Students' Part 2 comments regarding Speaker 1	202
5.14	Students' Part 2 comments regarding Speaker 6	205
5.15	Students' Part 2 comments regarding Speaker 2	208
5.16	Students' Part 2 comments regarding Speaker 3	210
5.17	Students' Part 2 comments regarding Speaker 4	212
5.18	Students' Part 2 comments regarding Speaker 5	214
5.19	Students' Part 2 comments regarding Speaker 7	216
5.20	Students' Part 2 comments regarding Speaker 8	218
5.21	Students' Part 2 comments regarding Speaker 9	220
5.22	Students' Part 2 comments regarding Speaker 10	223
5.23	Students' Part 2 comments regarding Speaker 11	226
5.24	Students' Part 2 comments regarding Speaker 12	228
5.25	The distribution of three significant errors according to acceptability rankings	231

LIST OF FIGURES

<i>Figure</i>	<i>Title</i>	<i>Page</i>
2.1	The four-quadrant evaluation model used in the study	28
3.1	The distribution of Hong Kong English consonantal features according to the percentage of speakers using them in at least one context	113
4.1	Maximum and minimum frequencies in Hz for the ten male speakers	146
4.2	Part 1 of the survey form	151
4.3	Part 2 of the survey form	153
5.1	Overall acceptability scores for the twelve speakers and three groups of raters	162
5.2	Acceptability scores using the two measurements of ‘overall’ acceptability and ‘direct’ acceptability	164
5.3	The distribution of identified errors across error categories, ranked according to frequency	179
5.4	Identified word tokens for SYLL	183
5.5	Identified word tokens for VOWEL SUB (subcategory of VOWEL)	183
5.6	Identified word tokens for FULL VOWEL (subcategory of VOWEL)	184
5.7	Identified word tokens for PHONSUB-R	186
5.8	Identified word tokens for PHONSUB-V	187
5.9	Identified word tokens for OTHER C-SUB	187
5.10	Identified word tokens for TH-STOP	189
5.11	Identified word tokens for prevocalic or prepausal final CCR (CCRF-PV)	191
5.12	Identified word tokens for preconsonantal final CCR (CCRF-PC)	192

5.13	Identified word tokens for connected speech linking phenomena (CS-LINK)	193
5.14	Transcript of Speaker 1's sample showing the location of errors and the number of identifications	201
5.15	The distribution of Part 2 positive and negative comments for Speaker 1	202
5.16	Transcript of Speaker 6's sample showing the location of errors and the number of identifications	204
5.17	The distribution of Part 2 positive and negative comments for Speaker 6	204
5.18	Transcript of Speaker 2's sample showing the location of errors and the number of identifications	207
5.19	The distribution of Part 2 positive and negative comments for Speaker 2	207
5.20	Transcript of Speaker 3's sample showing the location of errors and the number of identifications	209
5.21	The distribution of Part 2 positive and negative comments for Speaker 3	210
5.22	Transcript of Speaker 4's sample showing the location of errors and the number of identifications	212
5.23	The distribution of Part 2 positive and negative comments for Speaker 4	212
5.24	Transcript of Speaker 5's sample showing the location of errors and the number of identifications	214
5.25	The distribution of Part 2 positive and negative comments for Speaker 5	214
5.26	Transcript of Speaker 7's sample showing the location of errors and the number of identifications	216
5.27	The distribution of Part 2 positive and negative comments for Speaker 7	216
5.28	Transcript of Speaker 8's sample showing the location of errors and the number of identifications	218
5.29	The distribution of Part 2 positive and negative comments for Speaker 8	218

5.30	Transcript of Speaker 9's sample showing the location of errors and the number of identifications	220
5.31	The distribution of Part 2 positive and negative comments for Speaker 9	220
5.32	Transcript of Speaker 10's sample showing the location of errors and the number of identifications	222
5.33	The distribution of Part 2 positive and negative comments for Speaker 10	222
5.34	Transcript of Speaker 11's sample showing the location of errors and the number of identifications	225
5.35	The distribution of Part 2 positive and negative comments for Speaker 11	225
5.36	Transcript of Speaker 12's sample showing the location of errors and the number of identifications	227
5.37	The distribution of Part 2 positive and negative comments for Speaker 12	228
6.1	Factors affecting individual feature development and use in L2 phonology acquisition	261
7.1	The four-quadrant evaluation model (from Chapter 2, section 2.4)	268
7.2	The concentric circles of English users	284

ACKNOWLEDGEMENTS

A dissertation evolves over a considerable period of time and involves a myriad of contributions from many different people. I shall do my best to acknowledge them here. First in line should be my supervisor, Dr. Roger Berry, and co-supervisor, Dr. Judy Ho. Their comments and corrections, as well as their overall guidance and encouragement, played a major role in shaping the dissertation. Also within the Department of English at Lingnan University, I am indebted to all of my colleagues and the administrative staff for their contributions and support. My students at Lingnan have provided extremely valuable insights into the subject matter, and they were also the patient subjects who completed the survey forms.

Looking further afield, I am grateful to the anonymous reviewers from two journals – *World Englishes* and *English World-Wide* – who gave comments on articles that were based on chapters and sections of this dissertation. The editor of *English World-Wide*, Professor Edgar Schneider, also provided feedback on an article based on the preliminary study. I would also like to thank Professors Jennifer Jenkins and Andy Kirkpatrick, whose prompt and generous replies helped to clarify my understanding of their work. Thanks are also due to Professor Lau Tak Shing, who gave valuable advice regarding statistical methods.

CHAPTER 1

INTRODUCTION

1.1 Topic and context

The topic of this dissertation is at once local and global. Its starting point is the contention, made by some, that existing models for the teaching of English pronunciation are inappropriate. This contention is based on the view that such models are based on native speaker norms, which are seen as increasingly irrelevant in a globalising world where the majority of English users are actually non-native speakers (Crystal 1997; see also Jenkins 2000). Not all observers agree with this contention, and the ensuing ‘models debate’ (see Kirkpatrick 2006) has been one of the most heated controversies in applied linguistics during the last decade. The wider, global origin of the debate lies in phenomena such as globalisation, and indeed the topic of this dissertation, stated in its broadest terms, is the effect of globalisation on the use of English and on the attitudes people hold towards it.

The phenomenon of globalisation has many dimensions, and one explanation of the models debate is that it inevitably touches on social, political and economic issues. To fully understand the debate, it is necessary to frame it within wider currents of thought and historical trends; however, one of the aims of the dissertation is to provide data, rather than debating points, and to concentrate on linguistic, rather than political issues. At the same time, it also attempts to maintain an appropriate awareness of the origin and context of both the input, in terms of previous research and commentary, and the output, in terms of the study’s findings.

In more specific terms the dissertation is concerned with the question of whether existing pronunciation models in Hong Kong are in need of modification, for example in line with the proposals put forward by Jennifer Jenkins in her 2000 book *The Phonology of English as an International Language*. This work has become

something of a debate in itself, with at least one volume devoted to considering its implications for English teaching (Dziubalska-Kořaczyk and Przedlacka 2005). Contributions from eminent linguists such as John Wells, Peter Trudgill and Peter Roach confirm the importance of the topic, although most of the chapters are sceptical about some or all of Jenkins' claims; her own chapter is entitled 'Misinterpretation, Bias and Resistance to Change: the Case of the Lingua Franca Core' (Jenkins 2005). This dissertation includes, as part of a general evaluative orientation, a thorough assessment of the Lingua Franca Core or LFC (Jenkins 2000). This is a list of the phonological features of English that are thought to preserve intelligibility in international communication. Certain features outside this 'core' form promising candidates for omission from teaching syllabi, assuming that international intelligibility is seen as being an important criterion.

Jenkins (2000, 2007) argues that in today's world, 'international communication' in English is likely to feature the absence of native speakers. There are numerical arguments in support of this. Crystal (1997: 54) believes there may be up to a billion English speakers in the so-called 'expanding circle' of countries such as China, Japan and Germany, with an additional 300 million in the 'outer circle' made up of former colonies such as India and Singapore. A total of 1.3 billion non-native speakers compares with 380 million native speakers in the 'inner circle'. While there may be great differences in the frequency and proficiency with which these people use English, the numerical evidence is persuasive: why should native speaker norms and inner circle pronunciation models form the basis for English teaching worldwide?

The numerical argument finds political expression in discussions of the 'ownership' of English. Jenkins (2000) gives a prominent position to the views of Widdowson (1994: 385), who believes that native speakers are 'irrelevant' and that English is an international language over which no nation can have custody. The problem with discussions of ownership in linguistic terms is that there are no criteria by which it can be assessed; in many ways it appears that English belongs to everybody and nobody. Despite this, researchers within the English as a Lingua Franca (ELF) movement or paradigm (e.g. Seidlhofer 2004; Seidlhofer, Breiteneder and Pitzl 2006) appear to have the aim of 'uncoupling' the language from its native speakers and

their norms (Phillipson 2008: 250). These researchers frequently use Jenkins' work on international intelligibility in support of their aim, although the amount of empirical evidence is limited.

The models debate has intensified during the last decade, but it is hardly a new topic in applied linguistics. The celebrated exchanges between Braj Kachru and Randolph Quirk in the 1980s began the debate, and led to the creation of epithets such as 'liberation linguistics' (Quirk 1990) to describe, somewhat disparagingly, the attempts to achieve self-determination for other varieties of English. This period also saw the emergence of another research paradigm, the World Englishes movement. This has been particularly influenced by the work of Kachru, whose cogent arguments for the linguistic systematicity and functional range of non-native Englishes has helped to 'shift negative perceptions in some quarters and bolster linguistic self-confidence in others' (Bruthiaux 2003: 172).

The former British colony of Hong Kong has often been characterised as a pragmatic society, one that would seem to be far removed from concerns about 'linguistic human rights' (Phillipson 1992: 93). Nevertheless, both the ELF and World Englishes movements have left their mark on scholarly discussion about English. While the term 'Hong Kong English' appears to have little meaning for the general public, there have been detailed descriptions of its distinctive linguistic features, including the Hong Kong English accent (e.g. Bolton and Kwok 1990; Hung 2000; Deterding, Wong and Kirkpatrick 2008). These accounts of phonological features have followed the imperative to describe 'new varieties of English' or NVEs on their own terms, and imply the existence – at least at the formal level – of a local variety of English that shows 'autonomy and creativity' (Bolton 2000). The local scholar Andy Kirkpatrick has argued that Hong Kong needs 'local bilingual models' for English teaching (Kirkpatrick 2007b: 376). His identification of the value of a 'lingua franca model' (Kirkpatrick 2006) shows that the World Englishes and ELF paradigms converge to some extent, and also suggests that a review of pronunciation models is needed.

1.1.1 Justification and general research orientation

Despite the evident need, there are several obstacles to such a review. The first is the lack of a detailed description of the local variety, including at the phonological level. The need to base this description on proficient speakers has been identified by Kirkpatrick (2007b: 387), who sees a need for a ‘codified description of the local bilingual variety of English...as exemplified by highly proficient users of English who are mother-tongue speakers of Cantonese’. A limitation of earlier descriptions (e.g. Hung 2000; Stibbard 2004; Deterding *et al.* 2008) is that they have tended to focus on university students, who are often far from being highly proficient. Also, while providing useful data about phonological features, there is little indication in these studies of how they are actually distributed across the population. A variationist perspective on language data is required, and this is the approach taken by this dissertation.

Another obstacle is the lack of research data about how local people react to the phonological features of the local variety. In fact there is little data available anywhere on how non-native speakers process the speech of other non-native speakers, although Jenkins (2000) and Munro, Derwing and Morton (2006) have looked at intelligibility in these situations. As well as intelligibility, it is vital that a description of the local variety possesses local acceptability, especially if it is intended to be used for pedagogical purposes. This dissertation approaches the concept of acceptability from this angle, and intends to investigate the question of whether a local model would be acceptable to students, as well as identifying the implications for the teaching of English pronunciation in Hong Kong.

The second major characteristic of the research orientation is that it is features-based. In conjunction with its variationist outlook, this will avoid the ‘generic’ or ‘varieties-based’ approach of earlier Hong Kong English accent studies such as Forde (1995), Luk (1998) and Candler (2001). These studies investigated the acceptability of the local accent by comparing it with other native and non-native accents; however, they did not control the accent samples for their phonological features, and generally only provided one sample of each accent. The orientation of the present study resembles

that of Bolton and Kwok (1990), who presented ‘mild’ and ‘broad’ accents to student listeners. However, it will try to ensure that a range of accent samples and phonological features is included, and it will use more detailed statistical analysis to assess the effects of different features. It is not a primary intention of the study to compare the local accent with other English accents, as it is taken for granted that the model most local students are likely to encounter in the classroom is the local accent.

A third characteristic of the study is that it will adopt a multidimensional approach to the evaluation of phonological features. Intelligibility and acceptability form two of the dimensions, but others, such as the markedness of features and their relationship with features of the first language, will also be considered. The evaluation procedure will be informed by the study’s own data on the distribution of features and their acceptability, and will draw on indirect evidence from other studies of intelligibility. It will thus identify possible adjustments to pronunciation teaching syllabi.

The practical aims of the study locate it within the field of applied linguistics, in particular within language and education. The terms ‘variationist’ and ‘acceptability’ indicate that it is concerned with the social dimensions of language. It is hoped that there will also be some relevance for research areas such as World Englishes and the description of new varieties.

1.2 Aims and methods of the study

1.2.1 Aims

The study therefore aims to provide more detail about variation and acceptability in Hong Kong English phonology, with an overall orientation towards pedagogical applications. More specifically, the aims of the dissertation are:

- to critically examine the ‘models debate’;
- to conduct an initial evaluation of certain features in order to identify possible departures from standard models;

- to describe features and variation patterns in Hong Kong English phonology;
- to investigate the concept of ‘acceptability’;
- to examine listener reactions to accents and features in terms of ‘acceptability’;
- to provide a principled explanation for the rating differences, and for the differences in areas such as intelligibility;
- to further evaluate features in order to make recommendations for language teaching and testing in Hong Kong; and
- to identify any significant implications, for example those related to the teaching of pronunciation in general, and to the description of local varieties of English.

1.2.2 Methodological approach

The study contains two complementary stages of research. The first part, called the preliminary study, uses a mini-corpus of spoken Hong Kong English to describe feature use and variational patterns within a sample of proficient speakers. The mini-corpus was derived from broadcast material and thus prioritises authenticity, focusing mainly on spontaneous speech samples. The considerable differences between speakers led to a focus on inter-speaker, as opposed to intra-speaker, variation. The technique employed to depict the patterns of variation found in the data is implicational scaling. This is a representation of hierarchical co-occurrence patterns that originated in sociolinguistic studies (e.g. DeCamp 1971; Bickerton 1973), but has been used only rarely in the study of new varieties of English (cf. Ho and Platt 1993).

The second part, called the main study, also makes use of the mini-corpus by selecting twelve accent samples (eleven Hong Kong English speakers and one British English speaker). These accent samples were played to three groups of university students in intact classes, who rated the samples for their acceptability (using a six-item questionnaire) and marked salient phonological features on an accompanying transcript. In order to investigate the relationship between the acceptability ratings

and the occurrence of phonological features, the multivariate procedure of linear regression was used.

One of the principal methodological problems for the main study was maximising its internal validity by differentiating between listener factors and stimulus properties (Gass and Varonis 1984, in Munro 2008: 205). Listener factors refer to differences between the raters, such as the amount of experience with accented speech. The usual statistical verifications of internal consistency and inter-rater reliability were employed, but a more serious problem remained in the area of stimulus properties. Certain linguistic variables within the samples, such as the degree of lexical and syntactic complexity, the speech rate, and the range of voice pitch used by the speaker, were measured in order to see how these compared with the effects of phonological accuracy (as rated by the student listeners). However, this approach was unable to take account of all the possible variables that might affect listener ratings, whether paralinguistic (such as aspects of intonation) or extralinguistic (such as tone of voice). The findings regarding accuracy in general, and the effects of particular phonological features, must therefore be interpreted with some caution.

1.2.3 Explanatory approach

As a consequence of the above limitations, which are largely unavoidable with the use of authentic speech data, the explanatory approach is an ‘ecological’ one in that it accepts the notion of ‘multiple causation’ (Thomason and Kaufman 1988: 57, in Farrar and Jones 2002: 1). The explanatory model is ambitious in that it attempts to synthesise the findings and interpretations of the preliminary study, regarding variation and development, with those of the main study, regarding acceptability. The explanatory model is in fact a refinement of the evaluation model used elsewhere in the study. It includes both linguistic and non-linguistic factors, further subdivided into two subcategories, so that linguistic factors include ‘user-internal’ factors, such as contrasts between the first and second languages, and ‘user-external’ factors, such as intelligibility. Non-linguistic factors also include user-internal factors such as issues of personal identity, as well as user-external, more socially-oriented factors such as language attitudes. In both the evaluation model and the explanatory

model, synchronic evidence from variation in other varieties of English and diachronic evidence from language change are also used to inform the study's conclusions. A fuller explanation and diagrammatic representation of the evaluation model is given in Chapter 2, section 2.4, Figure 2.1, while the explanatory model is shown in Chapter 6, section 6.5, Figure 6.1.

1.3 Terms and concepts employed in the study

1.3.1 General

One of the early indications of the study is that terms such as 'native speaker' and 'variety' are highly problematic, and may even have caused the models debate to become unnecessarily polarised. The existing terminology does not appear to be able to keep up with and reflect change. As English is frequently the language spoken at home in countries such as Singapore (Brown 1991; Wee 2002), there is no longer any reason to restrict the 'native speaker' label to the inner circle. Even the 'concentric circles' model (Kachru 1992b) seems unable to depict the heterogeneity of language use, and more recent formulations by Kachru depict the 'inner circle' as the group of highly proficient speakers of English, those who have achieved 'functional nativeness' regardless of their nationality or first language background (Graddol 2006: 110). It is the intention of the study to avoid making assumptions about the nature of 'native' and 'non-native' Englishes, and to observe them, as far as possible, on their own terms and without automatic reference to 'native' or 'standard' varieties. This requires a thorough consideration of both terminology and methodology.

First of all, the artificial dichotomy between 'native speaker' and 'non-native speaker' is one of the hardest problems to resolve. The solution of Jenkins (2000) is to adopt the terms 'MES' (monolingual English speaker) for the majority of inner circle users, 'BES' (bilingual English speaker) for both native speakers and fluent bilingual speakers of English, and 'NBES' (non-bilingual English speaker) for non-English speaking bilinguals. This terminology is somewhat unwieldy, and Jenkins

herself relies mainly on the native/non-native distinction in her later work (for example, Jenkins 2007). This study will also employ the terms ‘native speaker’ and ‘non-native speaker’, but it is acknowledged that the latter term reflects bilingual, rather than monolingual, competence. An assumption of much second language acquisition (SLA) research is that native-speaker competence is the goal of language learning, leading Kachru (1994) to warn of a ‘monolingual bias’ amongst researchers. While this study will investigate the extent to which local speakers of English approximate native speaker norms, and will assess the acceptability of various phonological features for a local audience, it will avoid the assumption that all speakers need, or want, to achieve native-like competence.

The terms ‘inner circle’ and ‘outer circle’ will be used from time to time in this study, to signify the broad commonalities of both groups. Hong Kong will be assumed to belong to the outer circle, although clearly there are differences between it and other outer circle countries such as Singapore and India, resulting from it being less multicultural and multilingual in nature. The essential characteristics of the outer circle are that they are ESL (English as a second language) environments that have their own spoken norms, but which rely on the inner circle for norms of written English (Mesthrie and Bhatt 2008: 29). The historical or dynamic model of variety development proposed by Schneider (2003) places Hong Kong English as being at the stage of ‘nativization’, in which there is some tension between local forms of English and more prestigious norms, for example in the guise of a ‘complaint tradition’ (Mesthrie and Bhatt 2008: 33). These two aspects of variety development further suggest that there is a need for more detailed investigation of attitudes towards the local accent, and an assessment of acceptability can be seen as an indication of the nature of local norms.

Despite the fact that research has demonstrated the systematic and sociolinguistically complex nature of English in many outer circle contexts, the associated varieties still tend to be distinguished from their inner circle counterparts by the use of terms such as ‘indigenized varieties’ and ‘local forms of English’ (Moag and Moag, 1977; Strevens 1992; in Higgins 2003: 118). Schneider (2007) adopts the term Postcolonial Englishes (PCEs) to denote the qualitatively different environments of the Englishes

found in many countries in Africa and Asia. The cover term New Varieties of English (NVEs) is adopted in this study, to include Hong Kong English and other regional varieties such as Singapore English. A problem with the term ‘new’ is that some inner circle varieties, such as Australian English, are actually newer than some outer circle varieties, such as Indian English. The study may also refer to ‘L2 varieties’ and ‘L1 varieties’ if some aspect of variation appears to be related to language background, for example the transfer of features from the L1 (the first language) to the L2 (the second language).

Given the study’s variationist orientation, it will also avoid making the assumption that varieties of English are used in the same way by all their speakers. Variation is a fact of all languages, dialects and varieties (Wardhaugh 2009: 23), and this variation often has social significance (Chambers 2002: 3). This study will distinguish between ‘subvarieties’ of the local variety of English on the basis of phonological feature use, but only as part of a more general model-building exercise; it does not assume that these subvarieties have any sociolinguistic basis, and the possible social correlations are not explored. In addition, the study does not concern itself in detail with the question of whether a local variety exists in Hong Kong. As will be made clear, the existence of common features at the phonological level (see also Hung 2000) is taken as evidence of existence at a formal level, whether or not varietal status is supported by evidence from the functional or attitudinal levels (Mollin 2006).

An initial problem with this differentiation between varieties and subvarieties is again one of labelling. The term ‘high proficiency’ has already been used to characterise the language data in the mini-corpus, and will be used again in the study to denote a subset of the accent samples used in the research. However, ‘proficiency’ does not seem to be satisfactory, as it is usually associated with learning and assessment. Many of the speakers in the study have been using English for several decades, and they are unlikely to be still ‘learning’ in any substantive way; rather, their English represents ‘ultimate attainment’ (cf. Moyer 2004) and is the result of adaptation within past and present environments of incentives and constraints. The

term ‘fossilisation’ is rejected by this study, as it has negative connotations of frozenness and limitation.

In this study, subvarieties are related to the concept of an ‘accent continuum’ (for example, see Altendorf 2003). First of all, ‘accent’ is taken to mean ‘the features of pronunciation which identify where a person is from, regionally or socially’ (Crystal 2003: 3). There may also be accent features which are associated with processes of language learning, as in the models of Major (2001) and Hansen (2006). The design of the preliminary study is intended to investigate how some of these features are distributed across speakers in Hong Kong English, and whether there are any patterns of co-occurrence. As the study has some pedagogical aims, criteria such as intelligibility and acceptability will also be used to evaluate accent features. Intelligibility is assessed using a range of theoretical considerations and perspectives, with reference to relevant research evidence. Acceptability is explored by the empirical research of the main study.

This brings the discussion of problematic terms to those that are more specific to the study. Although intelligibility is not addressed directly by the research, it plays a prominent role in the evaluation of features and the explanation of their patterns of use. Intelligibility can be narrowly defined as ‘word and utterance recognition’ and is usually associated with segmental features, for example as in Jenkins (2000). Segmental features are the focus of Jenkins’ study, but this does not imply that other phonological features have no part in maintaining or reducing intelligibility. Similarly, the study does not deny the importance of the higher-level qualities of comprehensibility and interpretability (Smith and Nelson 1985). These terms are explained in more detail in Chapter 2, but the general nature of ‘intelligibility’ as used in this study can be stated from the outset: in accordance with the research orientation of Jenkins (2000) and Deterding and Kirkpatrick (2006), it is seen as primarily involving the intelligibility of non-native speakers for other non-native speakers.

Turning to acceptability, which is addressed by the main study, the term is slightly more difficult to define. Previous studies have tended to associate it with

acceptability for native speaker listeners, but this is incompatible with the orientation and aims of the present study. In essence, the concept denotes a range of evaluative judgments regarding the use of language features. Jenkins (2007) uses the four components of correctness, acceptability, pleasantness and familiarity to assess non-native English accents, but the concept of acceptability is not clearly defined. Acceptability depends on the communicative context, as features that are acceptable for some contexts and audiences may not be acceptable in others. Again, the pedagogical orientation of the present study led to the inclusion of a questionnaire item assessing 'direct' acceptability, or acceptability for pedagogical purposes, as one of several components of 'overall' acceptability. These include the areas addressed by Jenkins (2007), and statistical analysis is used to investigate their degree of interrelatedness.

The concept of acceptability is central to this study. While there can be little doubt that the phenomenon of Hong Kong English exists in terms of its language forms, including its phonological features, the attitudes of its users towards these forms remains an under-researched area. There is a general sense of a strongly exonormative orientation, suggested by the Hong Kong English accent studies of Forde (1995), Luk (1998) and Candler (2001). However, none of these studies took a features-based approach, adopting instead a varieties-based approach with other varieties serving as comparisons. More detail about the acceptability of different types of Hong Kong accent is thus needed. The general importance of acceptability relates to the importance of language attitudes in general; without the existence of positive attitudes towards a variety or subvariety, it will be impossible to use it as a pedagogical model.

Finally, although the neutral term 'feature' is generally used in preference to 'error', to avoid prejudging the nature of phonological features (or 'innovations', depending on the research perspective), the former term is used as part of a consideration of phonological accuracy. In the main study, student listeners were required to mark significant features on transcripts of accent samples. They were asked to focus on 'negative' features, and these error codings formed the input for both overall measurements of accuracy and an assessment of the effects of particular features.

While later discussion may revert to the use of ‘feature’, in some parts of the study the term ‘error’ will also be used. The study avoids native speaker assessments of error and accuracy, although for the purposes of internal consistency the student error codings were verified by repeated listening and a consideration of the coding patterns among the raters.

1.3.2 Notation, symbols and variables

In terms of phonetic symbols, as in most studies slant brackets (/ /) are used to enclose symbols when they are being discussed from a phonological point of view, as an ‘abstract underlying identity’ (Altendorf 2003: xii). These symbols therefore represent the phonemes of a language. Square brackets ([]), on the other hand, are used when sounds are being considered from a phonetic perspective, as physical, surface realisations. Where phonetic, as opposed to phonemic, transcription is used, it may only be ‘selectively narrow’ in that it omits phonetic features that are not strictly relevant to the discussion (Altendorf 2003: xii).

The principles of feature naming employed in this study generally follow those in previous studies. The phonological features of Hong Kong English are seen as being variable features, and in most cases the variants involved consist of one that is associated with Hong Kong English, and another that is associated with ‘standard’ varieties. This inevitably involves comparison, and any study of frequency is obliged to make such comparisons. Some researchers have modified the way they refer to certain sounds to avoid implications of how they ‘should’ be pronounced; for example, Deterding *et al.* (2008) refer to ‘TH sounds’ instead of ‘dental fricatives’. In this study, while the naming of features is based on a comparison with the features of standard varieties, this is largely because of the desirability of achieving comparability and consistency with other studies. Most of the features are therefore familiar variables that appear in many varieties of English, such as TH stopping and TH fronting, and L vocalisation. Another category of features includes those that appear to be more specific to Hong Kong English, such as the conflation of [n] and [ɲ] and what are referred to in this study as /v/ substitution and /r/ substitution. The analysis of error codings also generated two features or variables that have not

appeared in previous studies, namely syllabic modification and initial consonant cluster modification (abbreviated to ‘initial CCM’, consistent with the abbreviation of final consonant cluster reduction to ‘final CCR’). A list of the consonantal features considered in the preliminary study is provided in Chapter 3, section 3.6.2, Table 3.4.

While the preliminary study focuses on these consonantal features, in the main study students were allowed to mark any kind of ‘error’ that they thought was significant. A list of the features included in the main study can be found in Chapter 5, section 5.4.2, Table 5.10. Vowel features are referred to using the ‘lexical set’ convention developed by Wells (1982). This enables comparisons to be made without referring to the actual vowels of ‘standard’ varieties; in any case, the variability within vowel systems is perhaps greater than within consonant systems, making any specification of a ‘norm’ problematic. For example, the vowel in the DRESS lexical set is usually given as [e] in RP (Received Pronunciation), while other varieties or speakers may have [ɛ]. Similarly, distinctions between lexical sets may be made in different ways by speakers of different varieties. The TRAP/DRESS contrast is achieved by the use of [æ] and [e] in RP, while other varieties may use vowels such as [a] and [ɛ] (Schneider 2004: 1114-1115). For reference, the lexical sets referred to in this dissertation are listed in Table 1.1 below, along with their RP vowel phonemes.

Table 1.1. Lexical sets (Wells 1982) referred to in the study.

Lexical set	RP vowel
TRAP	/æ/
DRESS	/e/
FACE	/eɪ/
GOAT	/əʊ/
LOT	/ɒ/
THOUGHT	/ɔ:/

1.4 Structure

This dissertation follows a conventional structure. In Chapter 2, the literature regarding the models debate is surveyed, and the approach to evaluation is described. The operation of the evaluation criteria is illustrated by a review of some of the

phonological features identified by Jenkins (2000) as being possible candidates for removal from pronunciation teaching syllabi. Chapter 3 then introduces the Hong Kong background, particularly in terms of its educational culture and the exonormative attitudes associated with it. Previous studies of Hong Kong English phonology are also examined in this chapter, and are used as the basis for the preliminary study of feature variation. The methodological approach of the main study is described in Chapter 4. Chapter 5 presents the study's findings, firstly those relating to the effects of various speaker variables, such as phonological accuracy, on the acceptability scores, and secondly those pertaining to the relationships between these scores and the phonological features used by speakers. Chapter 6 attempts to provide a principled explanation of the results, combining the findings of the preliminary and main studies and building a general explanatory model of feature use with reference to development, variation, intelligibility and acceptability. Finally, in Chapter 7 the study's research questions are addressed and the focus returns to the local, pedagogical context, before the wider implications of the study are considered along with its overall achievements and limitations.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter surveys the literature on pronunciation models, and uses the findings of intelligibility studies to identify possible areas for change in the traditional model. As there are more considerations than intelligibility alone, by introducing and examining other areas of concern an evaluation model will be constructed. This four-quadrant model includes both linguistic and non-linguistic factors, as well as those defined by the study as being either internal or external to the language user. Adopting this model and a features-based approach, some of the possibilities for change in Hong Kong will then be subjected to an initial evaluation. These possibilities are informed by the Lingua Franca Core (LFC) of Jenkins (2000, 2007), and the chapter also serves as a critical evaluation of this proposal. Finally, some of the pedagogical factors that need to be considered when making recommendations will also be reviewed, and the chapter closes with a brief account of the ideological positions that underlie the debate about pronunciation models. This chapter provides a general, rather than a context-specific, orientation to the subject; the Hong Kong background, including the phonological features of the local accent, will be presented in Chapter 3.

2.2 The current situation in pronunciation teaching

2.2.1 Native speaker models

Although the attention given to pronunciation teaching in the world's ELT classrooms varies enormously, there is little doubt that the models most commonly used are based on native speaker models, whether British English, often in the form

of RP (Received Pronunciation), or American English, in the form of General American or GA. Referring to these varieties as ‘older varieties of English’, Setter and Jenkins (2006) note some of the reasons for this, among them the availability of reference materials such as pronunciation dictionaries (e.g. Roach, Hartmann and Setter 2003) and introductory texts on phonetics and phonology (e.g. Roach 2000, who substitutes the label ‘BBC English’ while retaining the features of RP used in earlier editions of the same book).

Some recent texts employ slightly modified models. Collins and Mees (2003) make use of ‘NRP’ (Non-Regional Pronunciation), but this is practically synonymous with RP, the only difference being that the diphthong of SQUARE, normally /eə/, is replaced by the /ɛ:/ monophthong. Despite the ‘change in attitudes’ towards non-native varieties noted more than thirty years ago by Kachru (1976), and although today there is fairly wide agreement that learners would benefit from exposure to different accents (e.g. Phillipson 2002), there is little evidence of this being implemented at the level of commercially available pronunciation materials. The survey carried out by Wrembel (2005: 430) found that while alternatives to RP and GA were becoming more common, these were ‘less popular varieties’ such as Australian and Canadian English, rather than NVEs (New Varieties of English, such as Singapore English).

Another factor explaining the dominance of native speaker models is the innate conservatism of national governments, who perhaps fear that their international competitiveness or prestige may be compromised by the adoption of alternative models (Kirkpatrick 2006). At school and university, then, it can be assumed that native speaker models still dominate the explicit curriculum, although of course the actual variety of English that most learners are exposed to is likely to be a local model, that of the teacher. Crystal (2004: 531) notes that: ‘[t]he first dialect learned by most English-speaking children is a non-standard variety’, and it is clear that this applies to both ‘native-speaker’ and ‘non-native speaker’ children. It is left to the increasing number of college-level courses in ‘World Englishes’ (see, for example, Sakai and d’Angelo 2005) to provide the kind of intentional exposure to new

varieties of English that may well reflect the realities of English usage in today's world.

2.2.2 Criticisms of NS models: RP

The debate about the 'what' of pronunciation teaching has been a long-running one; as Prator (1968: 23) realised, pronunciation is 'almost always a sensitive point'. Widdowson (2004: 361) notes that in recent years there have been signs of 'an increasing recognition that the nature of English as an international language calls for a reconsideration of the assumption that learner objectives must necessarily be predicated on native-speaker norms'. This is also the starting point for Jenkins (2000: 16), who begins her critique of RP by claiming that English is now 'learnt and spoken most frequently to serve international functions among L2 speakers in international contexts', and that its L1 speakers have therefore 'forfeited the right to dictate standards of pronunciation for L2 use'. The notions of frequency of use and ownership are highly problematic in themselves, and positing a relationship between them is a political act. The political and polemical nature of Jenkins' claims will be left aside for the time being, and the rest of this section will examine some of her more specific criticisms of NS models.

Firstly, a familiar criticism of RP is that very few people actually speak it (Jenkins 2000: 14). Crystal (2004: 472) estimated that the proportion of the UK population employing a 'non-regionally tinged' RP accent was under 2 per cent, and falling. Of course, one must be careful to define exactly what is meant by RP. As mentioned above, by broadening the definition slightly to 'NRP', 'BBC English' or the even more diffuse concept of 'Standard English' (see Bex and Watts, 1999), it is likely that the percentage would be substantially increased.

Jenkins' second criticism of RP is that it is 'by no means the easiest accent for learners to acquire' (2000: 15). The reasons given for this include the fact that it lacks a close relationship with orthography, contains a large number of diphthongs and elides postvocalic /r/. The lack of systematic sound-spelling correspondence in English is largely a feature of all its varieties, although RP's non-rhoticity is indeed

noteworthy in this respect. Jenkins also identifies the three centring diphthongs of RP (/ɪə, eə, ʊə/) as being ‘redundant’ insofar as they can be replaced with monophthongal variants (for example, *poor* as [pɔ:] or [pɔ:r], and *here* as [hɪr]), in closer accordance with Scottish English or GA.

A third argument against the use of existing descriptions of RP, according to Jenkins (2000: 15), is that it has altered over time and that the version of it presented in teaching materials is something of an anachronism. She points out that teaching materials often do not reflect widespread changes such as the loss of the /ʊə/ diphthong in *poor*, the so-called ‘hapPY tensing’ (Wells 1982) observable in words such as *happy* and *very*, and more controversial innovations such as L vocalisation. According to Crystal (2004: 417), some ‘controversial’ features, such as the use of glottal stops, are present in recordings of early twentieth century speakers, including the phonetician Daniel Jones.

There are other, more ‘political’ criticisms of RP that relate more directly to questions of ownership and identity. The insistence on conformity with target language pronunciation norms can be seen as forcing learners to reject their own identity (Dalton and Seidlhofer 1994a, in Jenkins 2000: 16). The question of identity is complex and such debates tend to be difficult to resolve, not least because the concept of ‘identity’ is also subject to the pressures of globalization and internationalisation. Arguably, any pedagogical norm is controversial in that it may pose a threat to identity.

Questions of identity and attitudes will be considered in more detail later on, but at this stage the arguments against the use of RP appear to merit further attention. It is uncertain to what extent RP really serves as a ‘model’ or a ‘goal’ in actual classrooms, and it may well be the case that in the debate about models, RP is being set up as a ‘straw man’. It may be more accurate to say that it is English pronunciation in general that is difficult to learn, not its specific manifestations in the forms of standard varieties. It seems unlikely that many learners are forced to reproduce RP in the way that some of its critics appear to assume. But given its continuing influence in written teaching materials, even the problems mentioned

above suggest there are grounds for a reconsideration and review of pronunciation teaching models for second language learners of English.

2.2.3 General criticisms of NS models

The critique of Kirkpatrick (2006: 71) applies to NS models in general, including pronunciation models, and is framed within a global perspective on power relations and structures. He claims that publishers and international English language teaching institutions have strong commercial reasons for promoting an exonormative, NS model, and they will lobby for its adoption. Other factors explaining the domination of such models include their codification in the form of grammars, dictionaries and evaluation criteria; there is also a 'prestigious corpus of literature written in these varieties' (Kirkpatrick 2006: 72). Kirkpatrick implicates the innate conservatism of 'politicians and bureaucrats', who fear that international intelligibility will be compromised if non-native varieties are adopted. In terms of the models debate, it is argued that NS models represent powerful interests and are more likely to be represented and promoted. The overall effect of codification, historical authority, bureaucratic inertia, fears about 'intelligibility' and vested interests is to make choosing an NS model 'the easy or safe option' (Kirkpatrick 2006: 72).

This does not mean, however, that NS models are optimal for the world's learners and teachers. Kirkpatrick (2006: 73) follows Jenkins in identifying the underlying reason why NS models are no longer relevant: the vast majority of learners are learning English in order to be able to communicate with fellow non-native speakers. The issue of intelligibility will be considered in more detail in a later section, but another key argument against NS models is that they are not necessarily any more intelligible than NNS varieties. Another is that their use has the negative effects on the confidence of NNS teachers of English, who are 'being required to teach a model that they do not themselves control' (Kirkpatrick 2006: 74).

2.3 Possible solutions and alternatives

Despite the perceived need to move beyond NS models, there are few concrete proposals for what should replace them. Some commentators have taken the perspectives of local teachers into account; Kirkpatrick (2002: 222) examines curriculum alternatives in an Asian context and lists several advantages of NNS teachers, including the fact that ‘they provide an appropriate and attainable model of the language’. It should be noted, however, that there is little consideration of what kinds of model would be acceptable, or evidence in favour of them being more attainable. There is an unstated implication that the preferred model is a local, NNS model. Phillipson (2002: 22) makes this explicit, while not denying the possibility of a role for other varieties: ‘[l]earners need to develop receptive competence in many Englishes, beginning of course with local variants’.

Kirkpatrick (2006) examines three possible candidates for teaching models in outer and expanding circle countries: native-speaker models, ‘nativized’ or local models, and lingua franca models. His conclusion is that a lingua franca model offers the most appropriate solution for contexts where the learners’ main need is to communicate with other non-native speakers (Kirkpatrick 2006: 81). According to Kirkpatrick, in outer circle contexts the advantages of nativized models stem from their empowering nature. This is so at the classroom level, where local teachers gain ownership and expertise, and at a societal level, where ‘the choice of a nativized model over a native-speaker model is the choice of democracy over imperialism’ (Kirkpatrick 2006: 76). However, in expanding circle contexts, Kirkpatrick notes that there may be problems with gaining acceptance for local varieties; for example, in China there is a ‘traditional and strongly held attachment to standards and correctness’ (Kirkpatrick 2006: 77). But attitudes can change very quickly, and there may be a time when ‘China English’ becomes an acceptable choice. In fact, there are already some signs of change; Xu (2002: 230) asserts that when recruiting teaching staff, Chinese universities should ‘pay more attention to the educational backgrounds and English proficiency of the candidates than whether they are native speakers or not’. These views may mark the beginning of a more decentralized attitude towards pedagogical norms, but there is a need for more detailed research

into what ‘China English’ actually means, and into how far existing concepts of ‘proficiency’ can be applied.

The main reason for preferring a lingua franca model, however, is that it is seen as being liberating for both teachers (who are free to focus on communication, rather than conforming to an externally-imposed norm) and for learners (for whom such a model may be more culturally appropriate). Kirkpatrick (2006: 80) agrees with Jenkins in not seeing English as a Lingua Franca (ELF) as a ‘single standard’ or a new norm; it will be possible for users to adapt Lingua Franca English, for example by using local pragmatic norms. However, while the statistical, educational and cultural justifications for a lingua franca model have been cogently argued, the main problem is the lack of linguistic description and codification. The research of Jenkins (2000) in the area of phonology is one attempt, and the next section will examine her proposals.

2.3.1 ELF (English as a Lingua Franca)

Some of the most controversial proposals for specific changes to pronunciation models have come from the work of Jennifer Jenkins. Her 2000 work *The Phonology of English as an International Language* offers an indication of ‘how things could be done differently’ (Saraceni 2008: 21). The central premise of Jenkins’ ELF proposals is that NNSs of English now outnumber NSs, thus prioritising the needs of the former. Maintaining international intelligibility is increasingly important, and Jenkins’ empirical research has identified some of the causes of communication breakdown in various situations. These situations included multilingual classrooms with pairs or groups of students from different L1 backgrounds engaged in communication tasks, as well as social settings. Thus, optimal mutual intelligibility between non-native speakers is the goal of the Lingua Franca Core (LFC), a list of the phonetic and phonological features that are seen as indispensable for the maintenance of international intelligibility.

The purpose of the LFC is not to provide a new ‘monolithic variety’ (Jenkins 2007: 20), but rather to identify the features of accents that interfere with intelligibility. If

the findings are accepted, non-native speakers will not necessarily have to attempt to abandon their accents or copy a native-speaker model, if intelligibility is the goal. The other, more radical side of the coin is that modifications will be required from native speakers, who will have to ‘make phonological adjustments to render themselves intelligible on the international circuit’ (Jenkins 2000: 32). The LFC has a linguistic basis, but its implications are far-reaching in educational, cultural and political terms. The model outlined by Jenkins (2000) offers some indications of what an alternative, ELF-based pronunciation teaching syllabus might be like, and it will be described and evaluated in some detail in the following sections. Although it may appear to have limited relevance for Hong Kong, the preservation of international intelligibility is likely to be an important consideration in Asia’s ‘world city’.

2.3.2 The Lingua Franca Core (LFC)

The areas identified by Jenkins (2000: 159) as forming the Lingua Franca Core (LFC) and safeguarding international intelligibility are listed below. All of the features are seen as being equally important.

1. Consonant sounds: the inventory of permissible sounds resembles existing NS models, with the following differences:
 - Rhotic /r/ is preferred, so that RP speakers will need to modify this aspect of their speech.
 - The intervocalic /t/ of RP is preferred to the voiced alveolar flap [ɾ] of GA in words like *better*.
 - Most substitutions of the dental fricatives (for example, [t] for /θ/ and [d] for /ð/) are permissible. Pronouncing *three* as [tɹi:] and *that* as [dæt] would be seen as unproblematic. The same applies to substitutions of postvocalic /l/, such as the process of L vocalisation involved in pronouncing *bill* as [bɪʊ] instead of [bɪɫ].
 - Close approximations to core consonant sounds are permissible while certain L1-influenced substitutions are not, for example the

pronunciation of *very* as [βeri] by Spanish learners, which tends to be heard as *berry*.

2. Phonetic requirements:

- The plosives /p, t, k/ must be aspirated in initial position, to avoid confusion with /b, d, g/.
- The phenomenon of ‘pre-fortis shortening’ whereby the vowel in *cap* ([kæp]) is shorter than the vowel in *cab* ([kæb]), whilst being phonemically equivalent, should be retained.

3. Consonant clusters:

- Initial clusters should not be simplified, so that the Hong Kong English pronunciation of *produce* as [pədju:s] (Chan and Li 2000: 82) would not be permitted.
- Medial and final clusters can be simplified, but only according to L1 rules of elision. This means, for example, that [t] and [d] may be deleted from words with three-consonant clusters like *facts* and *pounds*, or from two-consonant clusters where the next word begins with a consonant (such as *strict rules*).

4. Vowel sounds:

- Vowel length (quantity) contrasts should be maintained.
- L2 vowel qualities are permissible if consistent, but /ɜ:/ should be preserved.

5. Nuclear (tonic) stress:

- Nuclear stress placement and the division of speech into ‘word groups’ (tone units, or intonation units) are important for intelligibility.

There is thus an emphasis on segmental rather than suprasegmental features in the LFC core, which Jenkins admits is ‘an almost complete reversal of current phonological orthodoxy’ (2000: 135). Features other than those above are classified as ‘non-core’, with varying degrees of usefulness. Aspects of connected speech such as weak forms ‘may impede intelligibility’ (Jenkins 2000: 159), and stress-timed rhythm is either ‘unnecessary’ or it ‘does not exist’ (Jenkins 2005: 201). As mentioned above, one radical implication of the LFC is that some NS features actually hinder communication and should therefore be avoided by native speakers themselves, as well as being eliminated from pronunciation teaching models.

2.3.3 The LFC and its relationship with ELF and World Englishes

Jenkins (2000) situates her LFC proposals within the overall movement towards what she then called English as an International Language (EIL), and ends her book by wondering whether ‘EIL is the future of English’ (2000: 235). In her later work, the term English as a Lingua Franca (ELF) appears to have superseded EIL. Clearly, if the LFC can be demonstrated to be a workable model it will play an influential role in the development of English. Regarding the applicability of the LFC, Jenkins stresses that it is mainly designed for learners of English as a Lingua Franca (ELF) rather than of English as a Foreign Language (EFL). ELF interactions ‘typically occur between NNSs of English’ (Jenkins 2005: 200), whereas EFL is seen to involve NNS-NS communication. This distinction between ELF and EFL sits somewhat awkwardly with her insistence on the increasing prevalence of NNS-NNS interaction and the increasing irrelevance of native speakers. The distinction also appears difficult to use in practice, and there are many learners in supposedly ‘EFL’ countries who also need the language for international use. In Hong Kong (whether classified as ‘ESL’ or ‘EFL’), some learners may wish or expect to communicate mainly with NNSs, others with NSs, and still others may wish to achieve flexibility in different contexts. Along with many other concepts (including that of ‘native speaker’), there are arguments for reviewing the EFL/ESL distinction in today’s heterogeneous world, in which English is ‘no longer being learned as a foreign language’ (Graddol 2006: 19).

Jenkins (2007) includes as one of the common ‘misperceptions’ of the LFC the idea that it is being prescribed for all learners of English. Other misperceptions are that it is a model for imitation, rather than a core of communicatively important features, that it constitutes a universal norm, and that it is intended to make the task of pronunciation teaching ‘easier’ rather than enabling a more efficient use of the available time. But it is worth noting that arguments could be made for the introduction of LFC-type innovations in almost any situation, including those usually classed as EFL contexts, especially if it is accepted that learners of English in today’s world are more likely to encounter non-native speakers than native speakers, or if it can be shown that some features of native speaker models are in fact unnecessary. This tendency towards decentralisation can be observed in some of the recent literature. Deterding and Kirkpatrick (2006: 406) suggest that at some time in the future ‘learners from ASEAN countries will no longer always have to refer to external norms for their teaching materials’ and Kirkpatrick (2007b) makes a detailed case for a ‘local bilingual model’ in Hong Kong.

While the link between the LFC (and thus, a putative ELF) and local varieties has not been made explicit as yet, there are calls for the relative acceptance already won by outer circle varieties to be extended: ‘it is high time that the legitimacy which has already been accorded to Outer Circle Englishes should be extended to the Expanding Circle’ (Seidlhofer and Jenkins 2003: 152). Applying the LFC criteria to local varieties, in order to characterise internationally intelligible sub-varieties, would seem to be a logical and a legitimacy-enhancing progression. This would help to begin the process of description and codification called for by Kirkpatrick (2007b), and would also help to establish a link between the World Englishes (WE) movement, with its pluricentric orientation towards the description of new varieties of English, and ELF. However, the compatibility of the two approaches has been questioned by, for example, Berns (2008), who notes that the promotion of a ‘common core’ of pronunciation features sits somewhat uneasily with the valorisation of ‘autonomy and creativity’ (Bolton 2000) in the World Englishes movement.

Despite their admittedly tentative nature, their limitations and their possible incompatibilities with other research perspectives, the ELF proposals hold considerable appeal. If used to inform language teaching, their advocates claim they will not only maintain and enhance intelligibility for international users of English, but will also save time in the classroom and lead to an empowering reassessment of the value of local varieties; Kirkpatrick (2006: 81) believes that a description of Lingua Franca English will help to ‘liberate the millions upon millions of people currently teaching and learning English from inappropriate linguistic and cultural models’.

2.4 The evaluation framework

The problems and possible solutions facing English language teaching have been outlined above, and this section will define some important areas for consideration in the evaluation of pronunciation models. The ELF proposals are based on intelligibility considerations, but intelligibility is only one of the criteria that need to be employed. Whether evaluating varieties or features, evaluation criteria can be grouped into four categories or quadrants by using two intersecting axes. One axis (the vertical axis of Figure 2.1, below) takes the form of the commonly-used distinction between linguistic and non-linguistic factors. In the present study, linguistic factors are seen as being closely related to language form and function, whether in terms of acquisition (such as L1/L2 contrasts) or use (such as intelligibility). Non-linguistic factors are those which impinge on language, while not being directly linguistic in nature; they may include individual factors such as identity, and social-psychological phenomena manifesting themselves in the form of language attitudes.

Another useful distinction can also be made between individual and social factors. The second, horizontal axis in Figure 2.1 thus takes the form of an internal/external distinction, not the ‘language-internal’ versus ‘language-external’ division frequently encountered in sociolinguistic research, but rather a division between factors that operate mainly at an individual level (again, such as identity) and those that tend to

be more social or collective in nature (such as language attitudes, from the perspective of the speech community). The diagram below (Figure 2.1) uses the terms ‘user-internal’ and ‘user-external’ to label this axis, reflecting the status of L2 speech community members as language users with diverse proficiencies and backgrounds, rather than simply as language learners (after Hansen 2006). Of course, such a division runs the risk of being somewhat arbitrary in some of its classifications; language attitudes clearly operate at both individual and social levels. Nevertheless, this classification helps to schematise the evaluation criteria used in the study, and at a later stage it forms the basis of an explanatory model that uses similar concepts to show how feature use develops within individuals.

Figure 2.1. The four-quadrant evaluation model used in the study.

An evaluation framework oriented towards feature norms is also presented by Brown (1991: 112), who describes a technique of prioritisation by which the features of non-native pronunciation may be assessed in terms of various criteria. This part of the present study can thus be seen as an attempt to extend Brown’s evaluation

framework and to take account of recent research into international intelligibility. It also uses the framework to conduct an evaluation of the phonological features of alternative pronunciation models, especially those that occur in the phonology of Hong Kong English. The rationale for this is that if certain features occur in the local English accent, and if they can be shown to be unproblematic in terms of intelligibility (and some of the other factors shown above), there are grounds for accepting them in language teaching and testing. The following sections will explain the four quadrants of the framework in more detail.

2.5 Quadrant 1: linguistic, user-internal factors

There can be little doubt as to the considerable influence of the L1 in second language learning. In terms of phonology, ‘evidence of L1 transfer is clearly seen’ (Ringbom 2007: 66), but the nature and extent of this transfer may vary considerably. In considering the factors that affect language acquisition, including the acquisition of L2 phonology, there are two main areas to examine: L1 transfer, and developmental factors. L1 transfer means that sounds or rules from the L1 affect pronunciation; for example, Mandarin learners of English show a tendency towards vowel epenthesis in words ending with plosives, and this can be seen as a consequence of the L1’s prohibition of obstruents in coda position (Radwanska-Williams and Yam 2001: 36). Developmental substitutions are defined by Major (1987: 107) as ‘processes that occur in general L1 acquisition but do not operate in the learner’s NL [native language]’; the devoicing of final stops by Portuguese learners is given as an example. However, Major (1987: 107) admits the possibility of ambiguous processes that occur in both the adult L1 and in child L1 acquisition. For example, TH stopping, an accent feature in which the voiced dental fricative /ð/ is replaced by [d], could be explained as being due to either the transfer of an articulatorily similar sound or to more universal, developmental factors (these sounds are typically acquired late in child language development). As a general guide, features that occur widely across different varieties of English are more likely to be a result of developmental factors, while those which are confined to specific L1 backgrounds are more probably related to transfer.

To explain the widespread tendency towards terminal devoicing in L2 English (for example the pronunciation of *have* as [hæf]), Eckman (1977) proposes the Markedness Differential Hypothesis (MDH). This states that L1/L2 differences are not sufficient to explain learners' likely areas of difficulty; differences in markedness must also be considered. Markedness here refers to specific language features and is defined cross-linguistically in terms of implicational relationships:

A is typologically marked relative to B (and B is typologically unmarked relative to A) if and only if every language that has A also has B but not every language that has B also has A (Eckman 1977: 320).

Thus while there are no languages which allow voicing contrasts in final position but disallow them in initial or medial position, all languages (including English) that permit final voicing contrasts also permit initial or medial voicing contrasts (Jenkins 2000: 103). The voicing of consonant codas is thus a relatively marked phenomenon, and this goes some way towards explaining many learners' tendencies to devoice them. However, as in the case of TH stopping, terminal devoicing has also been observed in L1 child acquisition (Edwards 1979, in Jenkins 2000: 105). The conclusion of Jenkins' (2000) metastudy is that transfer errors tend to occur earlier in acquisition, while developmental processes come to the fore later on. This is also the position taken by Major (2001) and Hansen (2006).

Hecht and Mulford (1982) conclude that while transfer determines the likely areas of difficulty, developmental factors determine the ways in which these difficulties are resolved (for example, which substitutions will be employed; in Jenkins 2000: 105). The overall conclusion is that the importance of transfer cannot be neglected, and substitutions which are predicted by both developmental and transfer processes 'tend to remain in interlanguage the longest, sometimes as permanent fixtures' (Jenkins 2000: 109). This has important consequences for any assessment of L1-induced difficulty and the establishment of pedagogical priorities, but in Jenkins' view it is intelligibility that should act as the final arbiter. Whether the source of the difficulty is transfer-related, developmental or some combination thereof, 'the outcome depends on intelligibility' (Jenkins 2000: 120).

Further insights into the transfer-development issue are provided by Optimality Theory, or OT (Prince and Smollensky 1993). Hancin-Bhatt and Bhatt (1997: 331) note that OT ‘provides a more explicit account of the interactions between transfer and developmental effects in L2 syllables’ (cited in Radwanska-Williams and Yam 2001: 35). Thus for Mandarin learners, there are L1-derived markedness constraints against both consonant clusters and obstruents in final position, but general ‘faithfulness constraints’ are also posited, so that both vowel insertion and consonant deletion would be non-optimal in the case of a word such as *fit*. The output that violates the fewest constraints overall is thus [fitə], with schwa paragoge, while deletion, giving rise to [fi], is also possible. In the case of Cantonese learners, however, there are also L1-derived constraints against released obstruent codas, and the optimal output is [fitʰ] with an unreleased final plosive (Radwanska-Williams and Yam 2001: 36; see also Chan and Li 2000: 78).

While OT offers some potential for predicting and explaining learner difficulty, its complexity (for example, when deciding on the ordering of constraints) militates against its direct application to classroom teaching. However, its general approach can inform pedagogy to some extent, as noted by Radwanska-Williams and Yam (2001: 38):

Therefore, we recommend raising the learners’ language awareness of phonological phenomena to equip them for possible restructuring over life-long learning. From our study, we can conclude that a specific feature of which learners need to be made aware is the difference in the ranking of the faithfulness constraints. All of the phenomena we have described (epenthesis, deletion, unreleased plosives) affect the intelligibility of English pronounced with a “Chinese accent”, because they violate the OT faithfulness constraints, which are highly ranked in English.

The practicality of the proposed solution, that of making learners aware of faithfulness constraints in some way, remains to be evaluated. However, it appears to be worth establishing which features occur in learner interlanguage, whether temporarily or as ‘permanent fixtures’, in Jenkins’ terms; there are few language-

specific guides to L2 English pronunciation patterns (but see Swan and Smith, 2001). Applying intelligibility (and other) criteria will also be important, as pronunciation teaching materials often fail to distinguish between features in terms of their relative importance. An evaluative study of this nature therefore needs to consider the phonological features of the local accent, whether these originate from transfer or development.

2.6 Quadrant 2: linguistic, user-external factors

This section will consider the nature of intelligibility in some detail, as it is an important consideration in the evaluation of pronunciation models. As has been mentioned, the study's empirical research focuses on acceptability, so placing intelligibility on a firm theoretical foundation and considering the various sources of evidence is important. Drawing on various studies that have investigated international communication, and focusing on those that have featured non-native speakers and listeners, some of the factors that may reduce or maintain intelligibility will be identified. The section also examines related linguistic factors such as functional load and frequency.

2.6.1 Intelligibility

Definitions of the term 'intelligibility' have come under the spotlight as existing models and their associated ideologies have been questioned. Jenkins (2000) notes a shift in research orientations away from NS-centred conceptions of intelligibility towards more nuanced positions that take account of various types of listener and of the listener's role in determining intelligibility.

The search for an operational definition of intelligibility can be frustrating. There is some appeal in the position taken by Kenworthy (1987: 13), that intelligibility is 'being understood by a listener at a given time in a given situation'. A more complex approach is taken by Munro *et al.* (2006: 112), who use the three constructs of intelligibility (the extent to which a speaker's utterance is actually understood),

comprehensibility (the listener's estimation of difficulty in understanding an utterance) and accentedness (the degree to which the pronunciation of an utterance sounds different from the expected pattern). Of these, the last two are the most problematic. Comprehensibility here resembles Jenkins' concept of acceptability, and relates more to listener attitudes (as does accentedness). In an earlier work, Derwing and Munro (1995: 2) defined comprehensibility as 'native speakers' *perception* of intelligibility' (italics in original), and this is doubly problematic; it turns intelligibility back towards its earlier, pre-globalised definition and also opens up the possibility that attitudes may interfere with intelligibility, that there may be both perceived and actual intelligibility.

The predominant classification in current research is that developed by Smith and Nelson (1985), who use a tripartite distinction between intelligibility, comprehensibility and interpretability. The categories draw upon the speech act theory of Austin (1962) and the notions of locutionary, illocutionary and perlocutionary force, as noted by Brown (1989). In this classification, intelligibility means the recognition of words or utterances in themselves, comprehensibility is concerned with recognition of their meaning at a higher level of processing, and interpretability with the intentions behind an utterance.

The search for a precise definition of intelligibility seems to lead to somewhat unwieldy constructs, and in keeping with the practical orientation of this study a concise and workable definition is sufficient. The approaches above largely concur on their definitions of intelligibility itself, and the present study uses the term to mean 'word and utterance recognition', as does Jenkins (2000: 76). While comprehensibility and interpretability are certainly also relevant, this study has a greater concern with intelligibility. A case can also be made for the relative importance of intelligibility, as compared with the other two aspects. Although Kachru (2005: 200) mentions the need to 'rescue' the concept of intelligibility from a unidimensional interpretation by also considering comprehensibility and interpretability, there is a sense in which it acts as an enabler of the other two. If an analogy is drawn with written communication, then comprehensibility and interpretability are partly analogous to cohesion and coherence. Intelligibility is

analogous to legibility; if words cannot be recognised they cannot form part of a message, unless they are supplied by the receiver using contextual clues. Of course, the analogy should not be taken too far. In spoken communication both speakers and listeners make use of nonverbal clues to meaning. In practice, comprehensibility and interpretability are unlikely to be hindered by occasional intelligibility problems, but it seems reasonable to suggest that they should be avoided, unless of course the speaker is unconcerned about being intelligible.

2.6.2 Sources of intelligibility problems

It is interesting to consult pre-ELF views of the sources of unintelligibility, such as those of Kenworthy (1987). These were largely based on native speaker intuitions or on native speakers' experience of listening to non-native pronunciation. According to Kenworthy, the features that reduce intelligibility include substitutions, deletions, insertions (such as epenthetic vowels) and links between words (so that the Italian English pronunciation of *It's a big one* as *It's a big-a one* is an 'un-English' way of linking words that can 'make life difficult for the English listener' (Kenworthy 1987: 18). Although there is some agreement with Jenkins' LFC, for example in terms of the overall importance of most phonemic distinctions, there are several points of departure. Firstly, the relative importance of sound substitutions is not considered; the LFC would deem the dental fricative substitution involved in pronouncing *thick* as *sick* to be unproblematic. Secondly, the LFC sees native-speaker suprasegmental phenomena such as word linking to be unhelpful in international communication (although the Italian English example also makes use of vowel insertion, which would appear to be problematic). While for Kenworthy the lack of smooth transitions makes the speech of Chinese learners sound 'staccato and jerky' (1987: 18), from an LFC perspective this may actually enhance intelligibility by maintaining the regularity of sounds at word boundaries.

These differences may also stem from different conceptions of the listener. Kenworthy's learning goal of 'comfortable intelligibility' neatly encapsulates the two dimensions of acceptability and intelligibility, but the former appears to assume a native-speaker listener: 'when we listen to a foreigner speaking our native language

we expect to have to work a little bit harder' (1987: 3). Some commentators would no doubt dismiss such views as belonging to an outmoded EFL paradigm, but it is worth remembering that acceptability is also a consideration for non-native listeners. While Kenworthy does not refer to empirical evidence in support of her claim about word linking and its effect on intelligibility, there is also no convincing evidence for Jenkins' position regarding these suprasegmental features. We cannot assume that the presence of word linking phenomena causes intelligibility problems merely because the absence of such features was unproblematic for the learners in Jenkins' study. On the other hand, Jenkins' research offers 'persuasive evidence', according to Levis (2006: 251), of the intelligibility problems caused by segmental errors in NNS-NNS communication.

Unfortunately, a survey of the research into intelligibility reveals a preoccupation with the relative intelligibility of varieties, rather than the contribution of particular features. An exception, in addition to Jenkins' LFC research itself, is Deterding and Kirkpatrick (2006). The researchers employed a similar methodology to that of Jenkins, namely recording groups of NNSs conversing (all were teachers from ASEAN nations). An analysis of the factors that reduced intelligibility identified five pronunciation features: a vowel substitution (that of [ɑ:] for /ɜ:/ in *pearl*); a consonant deletion (that of the /r/ in *three*, which was pronounced [ti:] rather than the non-problematic [tɹi:] of some other speakers); two consonant substitutions (pronouncing *holes* as [hounz] and *sauce* as [ʃɔ:s]); and a consonant insertion (*us* pronounced as [ʌts]). These features are all proscribed by the LFC, as Deterding and Kirkpatrick themselves note (2006: 406). This indicates both the importance of segmental features, especially consonantal ones, and the usefulness of the LFC in predicting likely intelligibility problems.

Finally, and although a feature-based analysis is indicated, it should also be remembered that there are other threats to intelligibility of a more general nature. One of these is simply a lack of language proficiency, which may lead either to phonemic substitutions or to a large number of phonetic deviations, the cumulative effect of which is that 'the attention required to process the speech signal becomes too great' (Levis 2006: 252). This suggests that speakers who satisfy phoneme-based

intelligibility criteria may still fail to be intelligible. Others may use several phenomena at the same time, none of which is proscribed individually by the LFC, but whose simultaneous occurrence causes intelligibility problems. A range of other threats come from the listener, who may be unfamiliar with the accent in question or prejudiced against it for some reason. It is necessary to acknowledge that intelligibility does not only reside in the speaker, but is also negotiated between speaker and listener (Jenkins 2000: 78-79).

2.6.3 Features contributing to intelligibility

As well as those features which reduce intelligibility, it should be possible to identify some of the features which increase it. However, research in this area is sparse and recommendations are often inferential in nature. In their study of NNS-NNS interaction, Deterding and Kirkpatrick (2006: 406) claim that some non-standard features were actually helpful: ‘it seems likely that some of the features, particularly the avoidance of reduced vowels in unstressed syllables and also the clear bisyllabic enunciation of triphthongs, actually enhance understanding’. Some of the possible innovations in L2 varieties of English that may contribute to international intelligibility will be considered in the following sections. A critical perspective on the evidence will be taken, in order to build up a detailed picture of the possible effects of the features concerned. While not all of the features will be included in the subsequent evaluation of Hong Kong English, it seems advisable to include various possibilities, in case there are connections between some of them.

2.6.3.1 The avoidance of vowel reduction and word stress

The avoidance of reduced vowels in unstressed syllables has been shown to be a feature of many new varieties of English (NVEs). In the case of Singaporean English, Gek and Deterding (2005: 60) conclude that while speakers use fewer reduced vowels than British English speakers, they do use some reduced vowels, particularly when there is an *a* or *u* in the spelling (for example, in words such as *afford*). The existence of variable patterns of use, perhaps influenced by spelling patterns, is thus indicated. The conclusion of Deterding and Kirkpatrick (2006: 399,

drawing on Kirkpatrick 2004) is that ‘it is likely that many actually find that the clear enunciation of all syllables enhances intelligibility’. However, despite its intuitive appeal, there seems to be little empirical evidence in support of this position. There is also an emphasis on the needs of the listener, which may not be appropriate in all circumstances.

It is important to be clear about what ‘the avoidance of reduced vowels in unstressed syllables’ actually means in terms of the phonological system. In Deterding and Kirkpatrick’s data, it can mean both the avoidance of schwa in the unstressed syllables of multisyllabic words such as *communicative*, *compare* and *lessons*, and the use of full vowels in function words such as *from*, *to* and *that*. The first aspect may mean that word stress is seen as unimportant for intelligibility, another point of similarity with Jenkins (2000: 41), who found that ‘[o]nly rarely did word stress deviations alone present difficulties’. Jenkins’ conclusion (2000: 41) is that in the instances where such deviations were problematic there were also associated phonemic errors (for example, pronouncing *hopeless* as [ɒpʰles] and *alone* as [ˈelɒn]).

A consideration of the nature of word stress in English provides some possible reasons for this. Roach (2000: 95) states that while the prominence of stressed syllables is a result of four factors (loudness, length, pitch and quality), the strongest effect is produced by pitch, with length also being important; loudness and quality have ‘much less effect’. In NS varieties there are relatively few vowels that can occur in unstressed syllables, the most frequent being /ə, ɪ, u, i / (Roach 2000: 95), and long vowels and diphthongs occur more frequently in stressed than in unstressed syllables. While it could be argued that replacing reduced vowels with full vowels might affect perceived word stress and thus intelligibility, it seems that as long as the prominence of stressed syllables is maintained through pitch and length, vowel quality will not matter so much. Kenworthy (1987: 18) provides an example of misplaced word stress contributing to misunderstanding, namely *written* being pronounced with second syllable stress so that the listener – NS in this case – heard *retain* (from Bansal 1969). Kenworthy’s conclusion is that if the learner ‘doesn’t stress one syllable more than another, or stresses the wrong syllable, it may be very

difficult for the listener to identify the word'. But this would appear to be independent of whether reduced vowels are used or not; if *written* is pronounced as [ˈrɪtən] with stress on the first syllable, the use of a full vowel in the second syllable should not affect intelligibility. But if the stress is perceived to be on the second syllable, i.e. [rɪˈtən], then it is likely that many listeners will hear *retain*. Again, vowel quality by itself does not appear to be crucial in this area.

On the other hand, if learners are not aware of the way word stress tends to be related to vowel length and quality, the chance of so-called 'spelling pronunciations' occurring may become greater. According to Wells (2005: 104), in cases where NSs make differences in pronunciation that are not reflected in spelling, NNSs tend to ignore them (as in the case of *front* being pronounced [frɒnt], for example). It is possible that the instance mentioned above (pronouncing *alone* as [ˈelɒn]) was spelling-influenced; another plausible pronunciation might be [ˈælɒn], by inferring the vowel from *apple*. A more advanced learner would probably realise that the final orthographic 'e' of *alone* is associated with a longer vowel or diphthong (the so-called 'magic e' of phonics teaching). But once again, as long as the word has correct stress placement ([ˈelɒn] rather than [eˈlɒn]), the absence of a reduced vowel does not appear to be critical.

The examples above illustrate a psycholinguistic principle with relevance for intelligibility studies, that of lexical competition. The perception of *written* as *retain* is an example of what Weber and Cutler (2004: 3) call the 'unwanted activation of spurious competitor words' (in Trudgill 2005: 221). Although there is an intuitive appeal in the idea that this might be less of a problem for non-native speakers (because they may have smaller vocabularies and hence fewer competing words in their mental lexicon), Weber and Cutler's study found that lexical competition was in fact more problematic for non-native speakers. Trudgill (2005: 220), drawing on Dalton and Seidlhofer (1994a: 26) makes the additional observation that NNS listeners often have less knowledge of word frequencies and probabilities, and hence rely far more on acoustic information alone.

In summary, while the general importance of phonemic contrasts is indicated, there do not seem to be any intelligibility problems associated with the use of full (rather than reduced) vowels in the unstressed syllables of polysyllabic words, assuming that word stress patterns are maintained. This ‘innovation’ can be observed in several NVEs, including Hong Kong English (for example, the use of an /ei/ diphthong in the final syllable of words such as *appropriate*).

2.6.3.2 The avoidance of vowel reduction in function words

Another aspect of vowel reduction is the way it affects function words in native speaker accents, giving rise to so-called ‘weak forms’ of function words such as articles and auxiliary verbs. Deterding and Kirkpatrick (2006) believe that the avoidance of weak forms enhances intelligibility, although there is no direct evidence of this in their study. Jenkins (2000: 147) concludes, though again not on the strength of empirical evidence, that ‘weak forms may actually hinder intelligibility in EIL’. Once again, there is an intuitive appeal in this, augmented by anecdotal evidence that native speakers ‘dramatically decrease their use of weak forms in situations where they are taking extra care to be understood’ (Jenkins 2000: 147).

There is little research data about the effect of weak forms on the non-native listener; although such words are harder to hear, they do not carry essential meaning. The use and processing of weak forms is likely to be related to proficiency level, with more advanced speakers tending to use more (although this may depend on language background, among other factors). From the speaker’s perspective, the possible value of weak forms is that they allow more information to be included in a shorter space of time, with less articulatory effort. The importance of weak forms from this perspective is considered by Dauer (2005: 548):

It would be very difficult for anyone to speak English at a natural speed and produce all the consonants, consonant clusters and long stressed vowels of English precisely without reducing syllables, either in length or in quality. Simplification is inevitable: A speaker can either drop consonants (a typical solution for NNSs) or significantly reduce unstressed syllables, especially in

function words (a common solution for NSs). In either case, the burden of speech production is lessened.

While intelligibility studies tend to focus on the listener, Trudgill's (2005) concept of the 'speaker-listener equilibrium' should also be considered. As Smith and Nelson put it, 'intelligibility is...interactional between speaker and listener' (1985: 337; in Fraser Gupta 2005: 139). There may be a general tendency for language users to optimise the efficiency of speaking while preserving the accuracy of listening; Trudgill (2005: 222) refers to Dressler's (1984: 31) contention that 'the goals of better perception and better articulation often conflict with each other'. An additional consideration here is that the optimal resolution of this conflict for NS-NS interaction may not be equally optimal for NS-NNS or NNS-NNS interaction. Flexible performance appears to require a command of the continuum from less careful (but more rapid) speech to more careful (but slower) speech. The term 'hypo- to hyper-articulated continuum' is used by Shockey (2003) and Blevins (2004) to describe this range.

Further research is clearly needed to ascertain the status of weak forms in international communication. There are some specific problems that may occur if weak forms are not used, however. These may include a reduced ability to produce contrastive stress, in that asking the question *Are you going to Korea?* with the strong form of *to* may be interpreted as a request for confirmation ('please confirm that you are going to, not coming from, Korea'; example from van den Doel 2007: 31). Similar problems could arise with common function words such as *for* (e.g. *four people* versus *for people*).

2.6.3.3 *The avoidance of vowel reduction and prosodic factors*

A possible concern regarding the importance of reduced vowels is that they may be related to the prominence of nuclear stress (a core feature in the LFC); the absence of reduced vowels may make it harder to identify the tonic syllable. Jenkins (2000: 146) is 'not at all convinced by the argument that it is necessary to weaken an unimportant item in order to highlight an important one, provided that the latter is adequately

stressed'. Her solution is to focus on the extra length of the tonic syllable (Jenkins 2000: 155). This is supported by the observation that pitch and length are the most important factors in determining prominence (in the case of word stress; Roach 2000: 95).

There has also been a great deal of discussion of the relative merits of stress timing and syllable timing. Deterding and Kirkpatrick (2006) note that vowel reduction is a factor in determining the relative degree of 'stress timing' or 'syllable timing', further pointing out that the Pairwise Variability Index (PVI) method of measuring rhythm developed by Low, Grabe and Nolan (2000) depends entirely on the relative duration of neighbouring vowels. The concepts of stress and syllable timing are themselves controversial (see for example Roach 1982; Marks 1999). There seems to be a consensus that these notions represent points on a continuum, rather than a dichotomy, but that languages will differ in which type of timing predominates (Roach 1982: 78). The degree of stress or syllable timing may also depend on the speaker, and whether the content is familiar; 'the rarer the word, the more likely it is to be syllable-based' is an observation made by Crystal (1996: 12) regarding the stress patterns of a Ghanaian speaker and another West African speaker. A connection between vowel reduction and rhythm is posited by Roach (1982: 78), who observes that 'languages classed as stress-timed may be more likely to exhibit vowel reduction in unstressed syllables'. Roach does not preclude the possibility that languages classed as syllable-timed, such as French and Japanese, may tend to have simpler syllable structure. Thus there may be linguistic constraints operating in English, some of them historically derived, which affect interrelated factors such as rhythm, vowel reduction and phonological structure.

While Kirkpatrick (2004) argues that 'the acceptance of syllable-based rhythm in the classroom in ASEAN countries and many other parts of the world can be liberating both for teachers and for learners', Crystal (1996: 11) urges that general impressions regarding the rhythmic qualities of NVEs must be interpreted cautiously. There may be intermediate styles, or patterns related to speech rate or fluency. Crystal describes one Indian speaker as having 'a rapid and fluent speech style, in which the unstressed vowels...are produced with somewhat increased tension, resulting in a greater

evenness of articulation than would be found in stress-based speech'. However, he concedes that 'the impression that there is some kind of syllable-based speech among second-language English learners is widespread, and apparently affects all areas where new varieties are emerging, in Africa, South Asia and South-East Asia'.

Regarding studies of intelligibility, there is a dearth of research involving non-native speakers' experience of native-speaker, stress-based rhythm (as noted by Crystal 1996: 12). But Jenkins (2000) and Deterding and Kirkpatrick (2006) both conclude that syllable-based rhythm was unproblematic in NNS-NNS interaction. Thus stress-based rhythm is not part of the LFC, and the acceptance of syllable-based rhythm is actively promoted by Deterding and Kirkpatrick (2006) and by Kirkpatrick (2007b: 398):

Research over more than two decades has shown that varieties of English spoken by people whose first language has a tendency towards syllable timing – and Cantonese is one such language – are more easily intelligible in the international arena than are speakers of stress-timed varieties – and these include all native-speaker varieties of English (Smith 1992; Hung 2000, 2002; Kirkpatrick and Saunders 2005). This is not surprising, as speakers of stress-timed languages reduce vowel sounds, commonly resulting in the schwas in 'native' English. In contrast, speakers of English from ASEAN countries share a tendency towards syllable timing and a concomitant lack of reduced vowel sounds, and this is one explanation for why they experience few misunderstandings when using English as a lingua franca in intercultural communication.

This is an important passage because here Kirkpatrick is implying that syllable timing actually increases intelligibility, and that the avoidance of vowel reduction is an explanation for why they 'experience few misunderstandings'. However, although some studies of mutual intelligibility (Smith and Rafiqzad 1979; Smith 1992) have lent support to the contention that native speaker varieties are not necessarily the most intelligible, with the exception of Hung (2002) the studies mentioned by Kirkpatrick do not identify stress or syllable timing as a contributory factor. The conclusion of Smith (1992: 88) is that 'being a native speaker does not seem to be as

important as being fluent in English and familiar with several different national varieties', which suggests that proficiency level and exposure to different accents, rather than features such as stress timing, may be more important predictors of intelligibility. The need for a features-based, as opposed to a varieties-based, approach to questions of intelligibility is highlighted.

Similarly, Hung (2002: 8; in Kirkpatrick 2006: 74) believes that 'the massive reduction and neutralisation' of vowels in NS varieties may affect intelligibility, but there does not appear to be any experimental evidence for this. The study of Kirkpatrick and Saunders (2005) demonstrates that the 'well-educated variety' of Singapore English was highly intelligible for both native and non-native listeners, but there is no evidence of any connection between syllable timing and intelligibility (and no mention of to what extent the extracts were syllable timed); this was not the aim of the study. Another point to consider is that fluent non-native speakers of English may also use vowel reduction; it is not exclusively an NS phenomenon. Vowel reduction is not inherent to particular varieties, as speakers may alter the amount of vowel reduction they employ depending on the context (i.e., according to their positioning on a hypo- to hyper-articulated continuum).

It may therefore prove difficult to establish a relationship between stress- and syllable-timing and intelligibility, because of the influence of other factors such as proficiency level, experience of speaking, voice quality, accommodation skills, knowledge of the audience, topic and so on. There are also associated listener variables. The digital manipulation of speech samples offers one solution to the problem of multiple causation. The study of Tajima, Port and Dalby (1997), while limited by its use of native speakers as listeners, uses this method and produces some evidence that temporal modifications, including 'stress-related durational contrast', made the Chinese learners' utterances more intelligible. The researchers' conclusion is that the 'intelligibility of foreign-language speakers may be enhanced if explicit training is provided on temporal properties of their speech' (Tajima *et al.* 1997: 1). However, this enhancement of intelligibility may only apply to native speakers, or to those whose first language tends to be stress-timed.

Once again, it seems premature to promote syllable-based rhythm if enhanced intelligibility is the main consideration. We can be certain that there are highly intelligible speakers whose speech tends to feature one type of rhythm, but it seems unlikely that the two types are mutually exclusive within speakers. It is also probable that both types of rhythm can be found in less intelligible speech samples. More research into developmental patterns would be useful, as would studies of the interrelationship between prosodic factors and intelligibility. In the meantime, a possible observation is that features related to the hypo- to hyper-articulated continuum (such as weak forms, and vowel reduction in general) do not need to be taught; rather, they develop naturally as speakers extend their range along the continuum, while taking account of intelligibility.

2.6.4 Other studies of intelligibility

Within the ELF and World Englishes paradigms, the general attitude towards intelligibility can be summarised in this quote from Deterding and Kirkpatrick (2006: 398): ‘It is simply not true that inner-circle pronunciation is always the most straightforward to understand or the most appropriate as a model for learners’. A frequently-cited study in support of this position is Smith and Rafiqzad (1979). The researchers recorded speakers from eleven countries and regions, one from the inner circle (the US), seven from outer circle / ESL environments (Bangladesh, India, Malaysia, Sri Lanka, Nepal, the Philippines and Hong Kong) and three from expanding circle / EFL countries (Japan, Korea and Thailand). The recordings were then played to listeners from eleven NNS backgrounds, who completed tasks designed to measure intelligibility.

Smith and Rafiqzad (1979: 375) found that there was a ‘basic consistency’ in the measures of intelligibility; the NS American was always among the least intelligible speakers, while the Japanese, Indian and Malaysian speakers were almost always in the top five. No attempt was made to identify the features that affected intelligibility, and since there was only one NS some caution should be exercised when interpreting the results. Nevertheless, the researchers’ conclusion (1979: 380) appears reasonable: ‘[s]ince native speaker phonology doesn’t appear to be more intelligible than non-

native phonology, there seems to be no reason to insist that the performance target in the English classroom be a native speaker’.

However, a study by Major *et al.* (2002) provides evidence that is to some extent contrary to the findings of Smith and Rafiqzad. The study investigated the relative comprehensibility of four accents (Standard American English, Chinese, Japanese and Spanish) for non-native listeners from three language backgrounds, and found that the listeners ‘appeared to be disadvantaged by the use of NNS accents’ (in Levis 2006: 257). This was the case even when the listeners shared the same L1 as the speaker. Munro, Derwing and Morton (2006) examined the intelligibility of non-native speech for non-native listeners; a conclusion shared with both Smith and Rafiqzad (1979) and Major *et al.* (2002) was that there is no consistent intelligibility benefit for speakers and listeners with the same L1, thus challenging the view that non-native speech communities with the same L1 are unaffected by the substitutions, deletions and other modifications in each other’s speech. Munro *et al.*’s conclusion suggests that it is phonological features, not varieties, that are the proper focus of intelligibility studies. The researchers conclude that ‘properties of the speech itself were a potent determinant of the listeners’ responses’ (2006: 125). There is thus scope for quantitative studies of intelligibility that focus on the linguistic code, as well as for more qualitative studies based on interactional processes (as noted by Pickering 2006).

The importance of linguistic features indicates that a certain level of proficiency is needed. This aspect of communication is somewhat neglected in the ELF literature (see Nunn 2005). Rajadurai (2007: 102) observes that ‘core linguistic features that constitute a minimum threshold level’, as well as ‘overall proficiency and an adequate repertoire’ are prerequisites for communication. In other words, while we may accept Smith and Rafiqzad’s conclusion regarding the intelligibility and acceptability of non-native models, this does not imply a relaxation of standards or, perhaps more dangerously, an uncritical acceptance of such models without reference to their specific features. The need for an LFC-type core of features is once again highlighted, if international intelligibility is to be maintained. An awareness of proficiency is also required, as long as this is not defined with reference to irrelevant

native speaker norms. The existence of relevant native speaker norms is not denied by this statement, and arguably it is these norms that create the skeleton, the ‘dark matter’, of the LFC.

A critique of intelligibility studies in general is provided by Rajadurai (2007). Citing Smith and Nelson’s (1985) observation that ‘intelligibility studies are marked by confusion’, she questions methodological practices in the areas of speech sample collection (frequently leading to ‘artificial and inauthentic data’; 2007: 90); experimental settings that fail to take account of the specificity of topic, participants and situation; and the failure to appreciate the interactive roles of speakers and listeners in constructing intelligibility, for example via processes of accommodation. Rajadurai also ‘interrogates the problematic relationship between methodological practices and ideological beliefs’ (Seidlhofer 2007: 99) by drawing attention to the following ‘misconceptions’ and ‘myths’ (2007: 91-95):

- Only non-native speech is accented. (This mistaken belief is also remarked upon by Trudgill (1999: 118), who states that RP is a ‘standardised accent of English and not Standard English itself’.)
- Non-native speech lacks intelligibility.
- The non-native speaker is responsible for communication problems. (There is some anecdotal evidence to the contrary; Graddol (2006: 87) claims that research is beginning to show ‘how bad some native speakers are at using English for international communication’.)
- The native variety should constitute the norm. (This is argued to be inappropriate, as it ‘robs multilingual teachers of any sense of confidence’, and unrealistic, because it fails to reflect the lingua franca status of English).
- The native speaker is always the best judge of what is intelligible.
- The native speaker is always the best representative of what is intelligible. (Not only is there considerable variation in native speaker accents, but there is no evidence of their automatic superiority in terms of intelligibility, as demonstrated by Smith and Rafiqzad 1979).

This list of misperceptions provides a useful guide to the directions that future research into intelligibility, and in fact into international communication in general, should take in a globalizing world.

2.6.5 Other linguistic factors: functional load and frequency

Although few recent studies have made use of the concept, it seems likely that linguistic factors such as functional load help to explain the results of empirical, communication-based intelligibility studies. According to Brown (1991: 80), ‘the simplest expression of the functional load of a phonemic contrast is the number of words which this contrast serves to distinguish’. This measure is complicated by the fact that many minimal pairs involve different parts of speech (Brown 1991), so that while *thought/taught* and *fate/faith* are potentially confusable, other pairs such as *those* and *doze* are less so. Another consideration, according to Brown (1991) is the frequency of occurrence of the members of minimal pairs. The pair *look/Luke* is frequently used in pronunciation teaching materials to illustrate the /ʊ, u:/ contrast, but the effective functional load of a contrast depends on how frequent both words are (Rischel 1962, in Brown 1991: 83). The probability of confusing *look* and *Luke* would appear to be very low because of both part-of-speech and differential frequency considerations. A detailed measurement of functional load is beyond the scope of this study, and where necessary it will refer to the list provided by Brown (1991: 82) as an indicator of the relative importance of different minimal pairs.

While word frequency affects the likely impact of functional load, the frequency with which phonemes occur is also a consideration. Thus while the voiced dental fricative /ð/ is usually thought to be the most frequently-occurring phoneme in English, its voiceless counterpart /θ/ occurs less often. This may affect the noticeability of the sound (and of any substitutions), although there are complex interacting factors. Some of these will be considered in section 2.10, which evaluates these and other phonological features from various perspectives. The following subsection will consider functional factors from a general perspective and assess the extent to which they are relevant in evaluating pronunciation models.

2.6.6 Functional factors, synchronic variation and diachronic change

In some ways, the debate about pronunciation models could be seen as a debate about language change: variants arise out of L1 influence or communicative conditions and are subject to various interacting factors, including linguistic factors (such as functional load) and non-linguistic factors (such as pedagogical intervention and language attitudes) that determine whether they survive or not. A specific reason for including evidence from language change in an evaluation of pronunciation models is that if certain features can be shown to be likely sites of change within the language as a whole, this could strengthen the arguments for classifying them as acceptable variants. If, on the other hand, an assessment of the various factors indicates that they are not sites of change, careful consideration should be made before they are accepted in pedagogical models. However, the counterargument is that previous patterns of change may not always serve as a guide to the future, and such an approach would also tend to disregard the possible contributions of L2 or lingua franca Englishes to 'the language as a whole'. But despite this, a consideration of language change provides useful theoretical evidence and orientation.

A theme that recurs in some of the literature regarding language change is the interdependence of synchronic variation and diachronic change. A useful comparison of approaches to explanation in phonology is given by Blevins (2004). Her classification distinguishes between approaches according to whether they have a synchronic or a diachronic focus, and whether they are teleological or non-teleological in nature. This refers to whether or not they characterise language changes as 'improving' language in some way, for example by making it easier to understand or easier to pronounce. Synchronic approaches tend to downplay the importance of historical factors, on the grounds that 'a segment does not know where it comes from' (Lass 1984: 178, cited in Blevins 2004: 3).

Diachronic approaches, on the other hand, accept the importance of historical factors in explaining sound change in that 'synchronic sound patterns are best understood in terms of their historical origins' (Blevins 2004: 15). Thus there are seen to be certain parallels between synchronic variation and diachronic change; Blevins (2004: 4)

regards as noteworthy the fact that ‘the majority of commonly attested sound changes are mirrored by synchronic alternations of precisely the same type’. It is often claimed that L vocalisation, for example, represents both a synchronic alternation (in many varieties of English) and a long-term historical tendency. The focus of Blevins’s Evolutionary Phonology is on ‘acoustic auditory signals’ that are inherently ambiguous or easily misperceived; sound change occurs at an individual level because of this ambiguity, although the question of whether or not these changes spread through the community is outside the scope of the model. An important corollary of this focus is that sound change is seen as non-teleological or non-optimising, and it occurs because of ‘the way we produce and hear speech. It does not happen in order to improve speech in any way’ (Blevins 2004: 16).

However, although such a model would appear to deny the importance of functional factors, they are not completely ignored by Blevins’s non-teleological, evolutionary approach. The disappearance of the contrast between /w/ and /ʍ/ in many accents of English, for example, is explained by Blevins as being partly due to the low functional load of the contrast, as well as to its weakness in acoustic terms. Its survival in some varieties, on the other hand, is maintained through ‘unnatural means’ such as enforced practice and spelling pronunciation (Blevins 2004: 30).

The importance of functional factors in language change is not accepted by all, although their potential importance is clear: if functional load can be shown to be a determinant of whether, for example, certain phonemic mergers occur in languages, then it is clearly an important factor that operates during language use and acts as a force that preserves contrasts and intelligibility. This was the basis of the hypothesis put forward by Martinet (1961), but the available evidence tends not to support it (e.g., Labov 1994). Thus the best summary of the relationship between functional factors, synchronic variation and diachronic change seems to be that while synchronic variation often indicates potential diachronic change, the role of functional factors as causative agents remains uncertain. They may play a role in either inhibiting or hastening change, but the evidence suggests that other factors are equally important.

In the case of L2 users, however, it seems likely that functional factors have greater importance. Discussions of language change involve L1 speech communities that, by and large, have already acquired a common phonological system. Not all L2 users will have acquired this system, most parts of which are important for communication (this is in fact what is suggested by the LFC core). The desire to be understood, and the need to disambiguate meanings, are functional factors that almost certainly play a role in L2 phonology acquisition. Both functional load and frequency effects will play a role, as it seems likely that language learning and processing in human beings includes the ability to pay increased attention to frequent and important features in the input. The present study will therefore retain a consideration of functional factors, although the explanatory chapter of the study will also consider the effects of competing factors.

2.6.7 Summary: the predictability of intelligibility

Before moving on to non-linguistic factors, this section will provide a synthesis of the above discussions of linguistic factors. In doing so it will also explain and attempt to justify the present study's intended focus on the non-linguistic factor of acceptability, by arguing that the intelligibility of segmental phonological features is largely predictable – or at least, explainable – if factors such as functional load, psycholinguistic processes and the evidence from synchronic variation and diachronic change are considered. The section summarises the arguments put forward in Sewell (2010).

To recap briefly, the intelligibility-preserving LFC is given in its briefest form in Jenkins (2007: 23):

- All consonant sounds, except for the dental fricatives /ð/ and /θ/
- Vowel length contrasts
- Initial and medial consonant clusters
- Nuclear (tonic) stress

The relatively low functional load of the dental fricative phonemes in English provides an initial explanation of why substitutions are mainly unproblematic. There are few minimal pairs involving these sounds (Brown 1991). The voiced dental fricative /ð/ occurs frequently in English, and its occurrence in the definite article *the* provides a distributional reason for the intelligibility characteristics of substitutions. Other things being equal, variants such as [də] for *the* are unlikely to lead to misunderstanding – function words being, by definition, those which have little lexical meaning and carry little functional load (here used in the broader sense of ‘carrying important information’. Furthermore, in the case of the definite article, the fact that it is normally unstressed reduces its prosodic prominence; this also applies to any substitutions that are made.

A consideration of the psycholinguistic or information processing level helps to understand what happens during communication as a result of functional load. From a functionalist standpoint a recurring concept is the ‘speaker-listener equilibrium’ (Trudgill 2005: 222). Schreier (2005: 217-218) uses the concept of iconicity to explain why some sound modifications may affect communication more than others:

The basic assumption is that an increase in form entails an increase in information; following this, a decrease in form necessarily leads to loss of information. In order to be efficient and non-redundant, it is crucial to determine what form (or type of form) can be omitted with the level of information being high enough so that communication does not break down. How much of a form can be lost (and where), with the loss of information being minimal?

Schreier’s account is designed to explain consonant cluster simplification, but the general principle holds, that of simultaneously achieving ‘economy of production’ and the ‘reduction of ambiguity’, in the words of Williams (1987). Wells (1982: 94-97) calls these opposing forces ‘the principle of least effort’ and ‘the necessity to preserve intelligibility’, and there is also Dressler’s (1984: 31) observation that ‘the goals of better perception and better articulation often conflict with each other’ (cited in Trudgill 2005: 222). Using these concepts it can be seen that dental fricative substitutions may benefit the speaker by virtue of being easier to pronounce;

substitutions such as [d] may be ‘more natural’ (Wells 1982: 97). They also do not appear to hinder the listener’s ability to process information, because of their low functional load or participation in information-carrying.

The argument can be broadened to include postvocalic /l/ substitutions such as L vocalisation. There are few minimal pairs involving these sounds, but the issue is made harder to assess by the fact that vocalisation may lead to vowel changes and possible homophony (Wells 1982: 313). Turning to synchronic evidence, a general tendency towards L vocalisation is suggested by its occurrence in many NS varieties, so that in New Zealand English ‘vocalized /l/ is now so prevalent that many people cannot make a dark [ɫ] preconsonantly’ (Bauer 1986: 231, cited in Shockey 2003: 35). Diachronic evidence also suggests that L vocalisation is a commonly attested phenomenon, not only in English but also in other languages (Vulgar Latin *alter*, French *autre*; Portuguese *mau*, Spanish *mal*; Shockey 2003: 112). This suggests that L vocalisation may be something of an inexorable development in English.

A consideration of functional load, psycholinguistic factors and both synchronic and diachronic evidence can help to explain the limited importance of these consonantal features and the viability of their common substitutions. Consonant clusters can also be used to illustrate the same point. According to the LFC core, final consonant clusters can be simplified, but only according to native speaker patterns of elision. Initial consonant clusters, on the other hand, cannot be simplified. Final consonant cluster simplification is of course a widespread phenomenon that ‘operates in all varieties of English and is classified as a universal process of spoken English’ (Labov 1972, in Schreier 2005: 32). The greater economy of production achieved by the speaker through simplification partly explains this, but is there also an intelligibility explanation? Again, a psycholinguistic perspective (such as that taken by Schreier, 2005: 220) predicts the greater information-carrying role of initial clusters:

Lexical processing is a crucial factor to explain why initial clusters are more stable than final ones. Information lost at the beginning of words impedes word recognition whereas

information lost at the end of words often occurs at little cost, word recognition being completed already.

Although the greater importance of initial clusters for intelligibility can thus be predicted, there is a lack of data regarding the effects of different types of final cluster reduction. The differences between lexical clusters (such as in the word *kind*) and inflectional clusters (such as in the word *planned*) are not well understood, although Schreier (2009a: 60) proposes the following ordering of morphosyntactic factors that make cluster reduction more likely (the symbol ‘>’ means ‘more likely than’):

monomorphemic (e.g. *guest*) >
redundant bimorphemic (e.g. *slept*) >
bimorphemic (e.g. *guessed*).

While this ordering is probably related to universal processes and constraints, including the need to preserve grammatical information, the possibility of non-native patterns of cluster reduction being inconsequential for intelligibility should not be ignored. There may be a sense in which the LFC is overly proscriptive in that it limits permissible simplification to L1 patterns; this essentially means the deletion of /t, d/ in the centre of triconsonantal clusters in syllable-final position or across word boundaries, for example in the words *facts* or *pounds*, or in the sequence *recent news*.

Of the two other features in the LFC core – nuclear stress and vowel quantity – the importance of nuclear stress can also be gauged with reference to a broader conception of functional load. The placement of nuclear stress is obviously important for information processing, and will not be considered in detail here. The reason for the differences between vowel quality and vowel quantity are not clear, especially as cross-variety evidence suggests that vowel quantity plays a less important role than quality in signalling contrasts (Schneider 2004: 1128). The importance of vowel quality thus needs careful consideration. Jenkins (2000) claims that variations in vowel quality did not cause intelligibility problems in her data; however, given the difficulties in measuring vowel quality without acoustic analysis, the actual rate of

occurrence of vowel quality substitutions in Jenkins' data is uncertain. It may be the case that her subjects actually made very few outright substitutions, and that while a range of vowel qualities existed, instances of misperception were rare. If this is the case, it further suggests that vowel quality is in fact important for intelligibility, along with most other types of phonemic contrast. If there was an absence of substitution or merger in the language systems of learners in such intelligibility studies, it implies that they had already acquired these contrasts, perhaps in response to the communicative demands of intelligibility.

This is not to suggest that alternative realisations of vowels are problematic, or that all quality contrasts are equally important (on the basis of functional load, the combined quality/quantity contrast between /ʊ/ and /u:/ would seem to be unimportant, as would the /ʊ, ʌ/ contrast, further evidenced by the absence of contrast in many L1 English accents). But in most cases, categorical perception requires the maintenance of sufficient distance between vowels, in order to avoid confusion. This is to some extent predictable from language universals and psychoacoustic factors. According to Blevins (2004: 11), 'vowel systems are preferred to the extent that the perceptual space between vowels is maximized'. The maintenance of contrast is also required even if the speakers share the same variety. Abbott (1991: 233) gives the example of two Ugandan speakers for whom the distinction between *destruction* and *distraction* was blurred by the absence of an /æ, ʌ/ contrast.

Generally, then, linguistic and psycholinguistic factors, combined with evidence from synchronic variation and diachronic change, help to explain the results of the available intelligibility studies. The number of allowable deviations from 'standard' phonological systems is actually quite small, and in fact the features that the LFC allows – essentially, dental fricative and postvocalic /l/ substitutions, final cluster reduction, some variations in vowel realisations, as well as certain suprasegmental features – can be seen as some of the 'weak links' of English (Williams 1987: 168). They are variable features in many L1 and L2 varieties and exemplify the 'synchronic alternations' (Blevins 2004: 4) that may indicate possible diachronic change. Although there is clearly a need for further research, the theoretical

foundations of intelligibility appear to be solid enough to justify a focus on the acceptability characteristics of features in this study.

2.7 Quadrant 3: non-linguistic, user-internal factors

Quadrant 3 of the evaluation model relates to non-linguistic, ‘internal’ factors such as personal goals and identity. It is generally accepted that pronunciation is a ‘sensitive’ area (Widdowson 2004), one which relates at a fundamental level to our sense of who we are in the world. Much of what we know operates at an unconscious level; children ‘unlearn’ their initial ability to hear phonemic distinctions from languages other than their own (Randall 2007: 42), and thereafter concentrate on minute distinctions in their own language that may have far-reaching consequences for their social and economic lives.

If this is true for native speakers, then much of it is also true for non-native speakers. The choice of pronunciation model may be an irrelevance for some, but for many it is likely to have important consequences. However, people’s reactions to pronunciation models and accents is often based on non-linguistic factors, and despite the difficulty of researching them, such factors are crucial determinants of how language variation will be perceived – and hence of how learners will respond to pronunciation models in the classroom. There is a wide range of non-linguistic factors that may affect users’ attitudes towards language features or accents. Jenkins (2007: 198) sees attitudes and identity as being inextricably linked, and cites Pavlenko and Blackledge (2004: 1) in this regard:

[l]anguage choice and attitudes are inseparable from political arrangements, relations of power, language ideologies, and that [sic] interlocutors’ views of their own and others’ identities.

The changing economic and demographic landscape is noted by Jenkins (2007: 198), who sees, in China and elsewhere in the expanding circle, groups of English speakers who are both ‘economically powerful and numerically large’. This in turn is seen as

having the potential to alter the ‘ELF identity “landscape” in ways that are at present only dimly discernible’. The main effect of globalization, according to Jenkins (2007: 201-202) is to offer ‘new identity options’. Identity is no longer seen as fixed, and the existence of ‘multiple identities’ is posited. This in turn raises the possibility of ‘speakers being aware of what an NS of English would do at a given point, but choosing to do something else, perhaps in order to signal a shared identity with a particular NNS interlocutor’.

This is a plausible scenario, but the possibility of *not* being aware of what an NS would do might also explain the entailing language variation. Interestingly, this ‘multiple identity’ concept implies that something resembling native speaker competence and proficiency may still be required and desired by some speakers, even though they may choose not to deploy it in all situations. The results of Jenkins’ (2007) investigation of identity, however, revealed few signs of these nascent identities. The 17 teachers interviewed by Jenkins (2007: 231) appeared to have ‘very mixed feelings about expressing their membership of an international (ELF) community or even an L1 identity in their L2 English’. As mentioned above, it is left for Jenkins to offer the hope that globalization will bring changing attitudes and identities in its wake. The effects of conflicting identity, divided loyalty, or perhaps what Bamgbose (1998) refers to as the ‘love-hate’ attitude felt by many non-native speakers towards native-speaker varieties, are also apparent.

In her 2007 work, Jenkins reveals a more nuanced viewpoint than some earlier discussions of accent and identity. These tended to assume that speakers want to express their identity through their L2 English: ‘speakers of a NVE will want to preserve their identity’ (Kirkpatrick 2000, in Kirkpatrick 2002: 215). Perhaps the safest conclusion is that learners need to be given choices; they may wish to signal their identities on some occasions but not on others, or they may wish to achieve maximal intelligibility in some situations but not in others. This implies that they need to be given access to a wide range of models and their associated features, as far as is practicable. A criticism of planning interventions could be that it is the learners, and not applied linguists, who need to make the important decisions about ‘identity’. At the same time, this line of argument could be criticised as being a cover for the

maintenance of the *status quo*. Research needs to take account of the features that characterise varieties, and of the ways in which L2 proficiency levels affect these. It also needs to consider the differences between speakers of NVEs, rather than assuming that they belong to homogeneous speech communities with undifferentiated identities.

2.8 Quadrant 4: non-linguistic, user-external factors

Quadrant 4 of the evaluation model includes non-linguistic factors that are ‘external’ in the sense of being more related to social interaction and including evaluations made by others. According to Dalton and Seidlhofer (1994a: 10), ‘acceptability’ depends largely on the value people attribute to each other’s accents, and on whether these are seen as appropriate to the occasion and to their respective roles. Acceptability is thus a sociolinguistic consideration. Its crucial importance is recognised by Dalton and Seidlhofer, who point out that intelligibility is ‘often overridden by cultural and economic factors’ (1994a: 11). The concept of acceptability can be placed under the general heading of language attitudes; for Jenkins (2007: 88), intelligibility and attitudes are ‘symbiotically linked’. Jenkins (2007: 88) notes that research from Wolff (1959) onwards has shown that there are ‘other factors than basic understandability...implicated in perceptions of accents’. Munro *et al.* (2006) found that accentedness was rated more harshly than comprehensibility, and Rubin (1992) showed that by being encouraged to assume, from visual clues, that a speaker had a particular non-native speaker accent, listeners tended to ‘hear’ that accent where none actually existed (cited in Jenkins 2007: 88).

It is clear from abundant research that at present, L2 learners of English tend to have negative attitudes toward their own accents. Forde (1995) found that Hong Kong students reacted least favourably to English spoken with a Hong Kong accent, when also provided with samples of American, Australian, British (RP) and British (Yorkshire) accents; this could have something to do with the fact that the Hong Kong speaker was classed as ‘middle proficiency’ (Forde 1995: 64). Significantly, the local speaker was rated lowest on an ‘ability in teaching’ criterion. This pattern

can also be seen to some extent in the Singapore study of Chia and Brown (2002), who compared British (RP), British ('Estuary English') and Singapore English. In this study the Estuary English accent was rated lowest (behind RP, and then Singapore English) by Singaporean listeners on both 'intelligibility' and 'appropriacy for teaching' criteria. One reason for the exonormative attitudes displayed by the teachers in Jenkins' (2007) study, then, is that learners make an equivalence between 'standardness' and 'correctness'. However, this perception depends on the type of local accent used in research, as well as being possibly related to knowledge about intelligibility. In Hong Kong the study of Bolton and Kwok (1990) found that students were more likely to accept a 'mild' Hong Kong accent than a 'broad' one when asked about suitability for broadcasting purposes, although many did not appear to recognise the 'mild' version as being from Hong Kong.

In terms of individual reactions to L2 accents, it may be the case that 'irritation' arises mainly as a result of noticeable features such as phonemic substitutions, and that once the main phonemic inventory has been acquired such reactions decrease. Scheuer (2005) found there was a significant correlation between the degree of listener irritation and the degree of deviation from the target, and that 'what really irritates the listener is precisely non-native, L1-tainted pronunciation' (Scheuer 2005: 121). In Scheuer's study the listeners and speakers shared the same L1 (Polish), and while having different L1s might conceivably affect the results, on the whole there seems to be little basis for Jenkins' claim that L2 speakers of English tend to be less judgmental of each other's pronunciation (2000: 160). However, some evidence in support of this is given by the work of Beinhoff (2005), which also indicates an important difference between native speaker and non-native speaker attitudes. Beinhoff concluded that while non-native speakers are generally tolerant of each others' accents, they are stricter towards their own L1 group (in Jenkins 2007: 89). This is an observation that has relevance for studies of Hong Kong English.

On the one hand, Scheuer's finding suggests there may be a relationship between irritation and unintelligibility; one would certainly expect the lack of 'core linguistic features' or an 'adequate repertoire' (Rajadurai 2007: 102) to annoy the listener in most cases because of the greater effort required to achieve understanding. On the

other hand, one of van den Doel's (2007) findings was that while NS listeners tended to prioritise errors that impeded intelligibility, they also noted errors as significant because they caused 'irritation or amusement'. In the study of L2 Swedish conducted by Markham (1997), there appeared to be an inverse relationship between negative reactions to an error and its potential to cause lexical confusion (in Scheuer 2005: 117).

There is great complexity in the attitudes underlying these responses. Jenkins (2007: 89) cites the study of Fayer and Krasinski (1992) as evidence for non-native speakers having less tolerance towards the accents of their compatriots than native speaker judges listening to the same samples. There is the possibility that non-native speakers are 'embarrassed by their compatriots' struggles in the nonnative language', according to Fayer and Krasinski (1992: 321). A survey of research into learner attitudes towards inner, outer and expanding circle varieties leads Jenkins (2007: 105) to conclude that there is 'some sort of contradiction, ambivalence, or possibly deep-seated bias' among the participants. Jenkins' summary of her own research (2007: 141) into teacher attitudes reiterates the same themes:

[H]ow difficult these teachers, both NNS and NS, find the concept of ELF in general and ELF accents (particularly the notions of core and non-core) specifically; how closely they identify with an NS norm (usually RP or GA in terms of accents); how reluctant they are to disassociate notions of correctness from 'nativeness' and to assess intelligibility and acceptability from anything but a NS standpoint; and how, intuitively, they regard 'standard' NS English as being more widely understood than other varieties regardless of the context of use.

The study was based on a questionnaire given to 326 native speaker and non-native speaker teachers from a variety of language backgrounds, which asked them to rank accents according to their relative correctness, acceptability, pleasantness and familiarity. For the non-native speaker respondents, native speaker accents (especially US and UK accents) were rated as 'overwhelmingly the best' on all four criteria (Jenkins 2007: 166). It is not certain how the respondents interpreted the term

‘acceptability’, but the consistency of the results across respondents is clear. However, the study assumed that respondents had some knowledge of the accents in question, as it did not provide any actual samples. The study thus represents an account of perceptions of accents, perhaps gained through experience in some cases, rather than the accents themselves.

One study which did provide listeners with samples of L2 speech is that of van den Doel (2006, 2007). This had the advantage of considering the possible features of the L2 accent in question (Dutch English), rather than treating accents as generic entities. However, the study only included native speaker listeners, who were asked to comment on NNS speech samples. The comments reveal how their judgments largely depended on their attitudes towards the groups they associated with these features. The use of epenthetic schwa in the word *film* (pronounced [filəm]) attracted the following comments (from van den Doel 2006: 183):

There are some English people who would say filim for film, but usually they are considered Yorkshire bumpkins. [British listener]

But common pronunciation among southern Irish plebs. [Southern Irish listener]

Sounds like “fillum” adding a syllable to the word. This is the mark of an uneducated speaker. [American Southern listener]

It should be pointed out that these were not the responses of linguists, but they do show how a single sound can cause judgments to be made about a speaker’s geographical origin, social class and education. The existence of stereotyping and stigmatisation cannot be ignored in sociolinguistic studies of accents.

2.8.1 Stigmatisation

The concept of stigmatisation refers to the fact that listeners may make unfavourable judgments about a speaker based on accent features, often (as in the above examples)

related to their prior assumptions about social or regional groups. This is entirely what we would expect, in view of Milroy's observation that 'prestige is attributed by human beings to particular social groups' (Milroy 2001: 532). The actual origin, class or education of the speaker is irrelevant; the use of a non-standard form causes them to be identified with a particular group, and the stereotypical characteristics of the group are potentially attached to the individual speaker.

Another example of stigmatisation as it relates to dental fricative substitution is given by Wells:

One can conclude that any New Yorker who uses no plosives or affricates for /θ/ or /ð/ in careful conversation has been to high school and is not a manual worker – quite likely he is a professional. If he uses plosives and affricates sporadically, he has probably been to high school, but is not a professional; if he uses them a lot, he has probably had no high school education (1982: 517; cited in Brown 1991: 77).

Of course, this example of 'social stratification', to use Labov's phrase, is referring to attitudes within native speaker communities. One reaction to stigmatisation is to dismiss it as one of the 'little games' (Seidlhofer 2005: 63) of such communities; the LFC implicitly takes this position by assuming that intelligibility is the most important criterion for communication between non-native speakers. A similarly dismissive attitude position is apparent in Kirkpatrick (2007a: 196): 'while prejudice against varieties is likely to occur, these prejudices are simply that – prejudices'.

However, it is likely that some applied linguists tend to underestimate the power of prejudice. Milroy (2001: 538) wonders whether linguists have always 'fully understood the power of the ideologies that drive public opinion on these topics', adding that these opinions are 'deeply and sincerely held and are widespread in society, however ill-informed linguists may consider them to be'. Milroy (2001: 538) also notes the necessity for linguists to proceed with caution: 'If we tell people things about language that they believe to be untrue, they will mistrust us and reject what we say'. This could easily explain, among other things, why the discourse of applied linguists on the topic of local varieties is often at variance with wider public discourse, a fact noted by Joseph (1996) in the case of Hong Kong English, and more

generally by Bolton (2008). An alternative, pragmatic position is that stigmatisation and judgment are facts of the real – as opposed to an ideal – world (Scheuer 2005: 126) and that learners must be equipped to participate in it, perhaps through the inclusion of sociolinguistic elements in syllabi, where appropriate.

On the other hand, it is important to remember that ‘acceptability’ is not an immutable concept. Bamgbose (1998: 4) suggests that acceptability can be engineered, as the codification of an innovation, or its use by an influential figure, may predispose learners to accept it. Jenkins (2007: 188) concludes on a note of optimism, believing that globalization may be the force which overcomes language inequalities:

As English, the language of globalisation, spreads around the world and is appropriated by an ever-increasing range and number of NNSs, it is perhaps inevitable that their sense of inferiority in the language will one day begin to diminish and that they, the majority speakers of the language, will eventually start to see themselves as at least equals alongside NSs in the global lingua franca context.

This echoes the ‘world in transition’ analysis of Graddol, who as part of his ‘postmodern’ characterisation of the current situation believes that there is ‘a new model...English is no longer being learned as a foreign language, in recognition of the hegemonic power of English native speakers’ (2006: 18). One exception to the ‘NS is best’ pattern in the study of Jenkins (2007) was that the Chinese students were ‘particularly positive about their own English accents’ (2000: 165), although once again there is no indication of what these accents were like in terms of feature use and proficiency level. These positive attitudes may well be connected with a resurgence of confidence within China, proceeding from its growing economic, political and diplomatic influence.

2.9 A features-based evaluation of the LFC proposals

The previous sections have outlined the key areas in evaluating pronunciation models, and the following sections will take a features-based perspective and explore in more detail some of the proposals contained in Jenkins' LFC. The aim here is to begin to identify the likely candidates for removal from pronunciation teaching syllabi, in order to focus the research and discussion in subsequent chapters. In terms of the evaluation model, in section 2.6.7 above it was argued that the intelligibility characteristics of phonological features can be predicted, or at least explained, by considering various sources of evidence. The empirical research in this study will therefore focus on quadrant 4, on the acceptability of various features from the local listener's point of view. This approach is also justified by the frequent observation that while linguistic factors generally determine the features, alternations or possible changes that arise, it is non-linguistic factors that determine whether or not these features will actually spread through the community (see, for example, Blevins 2004; Altendorf 2003; Labov 1994).

In the following sections, certain phonological features of English will be evaluated using some or all of the quadrants. Evidence from other sources, such as that provided by language variation, or inter-variety comparisons, and considerations of pedagogical factors, will also be included where appropriate. The evaluations of features will thus be fairly comprehensive but general in nature, and those that emerge from the evaluation as being possibilities for removal from teaching syllabi will be further assessed for their acceptability in Hong Kong English in the experimental part of the study. Particular attention will be given to dental fricative substitutions, L vocalisation and consonant cluster reduction, as these are all commonly attested features of Hong Kong English phonology (e.g. Bolton and Kwok 1990; Chan and Li 2000; Hung 2000; Peng and Setter 2000; Deterding *et al.* 2008).

2.9.1 The dental fricatives

It has already been noted that the LFC's consonant inventory closely resembles that of existing native speaker models such as RP and GA. If the LFC is intended to

‘reduce the size of the task’ of pronunciation teaching (Jenkins 2007: 27) then there are few apparent reductions in this area. Why, then, have the dental fricatives been omitted from the list of ‘core’ sounds?

Firstly, linguistic factors will be considered. The dental fricatives are marked features of English in that they occur infrequently in cross-linguistic terms and are acquired relatively late in child L1 acquisition. In the case of L2 English speakers, L1 influences generate a wide range of possible substitutions for /ð/ and /θ/: [d] and [f] (in Hong Kong), [v] and [f] in Cockney, and [ɖ] and [ʃ] (by Japanese learners, according to Jenkins 2000: 137; the substitution of [s] for /θ/ also occurs). The LFC does not stipulate replacements. Jenkins’ own position is that all of these are acceptable substitutions, with the exception of the Japanese ones (they are ‘less familiar to all EIL receivers’) and [z], which while being familiar, apparently ‘elicited responses of irritation when used with high frequency’ (Jenkins 2000: 138). Jenkins’ recourse to familiarity does not really help her overall argument, however, as given the wide range of substitutions it is arguable that the ‘standard’ forms are likely to be equally familiar and less confusing in international interactions. There is some anecdotal evidence in support of this; confusion between *three* and *tree* was the source of a misunderstanding for a Chinese student in Singapore, who interpreted as ‘you dig tree tree’ the utterance ‘you take three three’ when asking about bus travel (quoted by Young, 2003). The substitution of /θ/ with [t] is widespread in south and south-east Asia, but may cause intelligibility problems for those who habitually use other substitutions.

The low functional load of the dental fricatives partly explains the viability of substitutions. Further empirical evidence in support of substitutions can be found in Deterding and Kirkpatrick (2006). Their methodology resembles that of Jenkins (2000), with twenty English teachers from each of the ten ASEAN countries being recorded in groups of three or four for twenty minutes. It was found that speakers from at least four countries substituted [t] for /θ/, and this did not lead to any instances of communication breakdown. However, the authors’ conclusion is that while this substitution may be acceptable within the ASEAN region, it may cause

problems ‘for listeners from more distant parts of Asia’ (Deterding and Kirkpatrick 2006: 396) – as happened in the example given by Young (2003), above.

Another example of this is provided by the study of Date (2005), whose first language is Japanese. When listening to a sample of Singapore English, Date found that dental fricative substitutions did affect intelligibility in these sentences:

we had thought of going backpacking together

I think it [gambling] is er it is quite a form of thrill to a lot of people

Date’s account (2005: 180) is that he ‘initially’ heard *trill* for *thrill* and *taught* for *thought*. On the one hand, these examples support the argument that dental fricative substitutions are inconsequential for intelligibility. For an advanced listener these misinterpretations would probably result in no more than momentary misunderstanding, as the words arising from substitution cannot occur in the context. But on the other hand, the experience of less advanced listeners may be quite different. Lacking the ability to use contextual clues (Dalton and Seidlhofer 1994a, in Trudgill 2005: 220) and relying more on phonological information, their misunderstandings may be more than momentary. The cumulative effect could be a loss of comprehensibility, as well as increased processing effort on the part of the listener. Date’s experience provides an example of the ‘unwanted activation of spurious competitor words’ (Weber and Cutler 2004: 2) occurring in ELF interaction, as a result of dental fricative substitutions.

Turning to considerations of frequency, matters become more complicated. According to Dalton and Seidlhofer, the voiced dental fricative /ð/ is the most frequent consonant in English (1994a: 145), but other sources (e.g. Fry, 1947; Denes, 1963; in Brown 1991) tend to rank it as the 6th or 7th most frequently occurring (Fry and Denes, respectively). The difference in rankings is due to the method of measurement: Dalton and Seidlhofer use running text, while Fry and Denes used reference materials such as phonetic readers. The high frequency of /ð/ in running text is of course due to its occurrence in articles, demonstratives and other high-

frequency function words. Its frequency of occurrence might be used as evidence against substitution, as it may exert a disproportionate influence on listener perceptions. However, given that it occurs mainly in unstressed syllables, its prominence would appear to be relatively low. The voiceless dental fricative /θ/, on the other hand, is ranked by both Fry and Denes as one of the most infrequently occurring consonants in terms of its distribution in the lexicon (23rd and 24th out of 24, respectively). The difficulty of interpreting linguistic evidence is suggested by the fact that some high-frequency words (such as *think* and *three*) contain this sound. All in all, the linguistic evidence in favour of accepting dental fricative substitutions is mixed but generally favourable in terms of markedness, functional load and intelligibility, and frequency.

In terms of acceptability, the fact that substitution with [z] was found to be ‘irritating’ in Jenkins’ (2000) study shows that the LFC is not exclusively concerned with intelligibility; acceptability is also considered, albeit only briefly. It is not clear who was irritated by the speaker in question, but given the research methodology this was presumably a NNS reaction. While there are many studies that attest to NS reactions to NNS speech (see, for example van den Doel 2007), the LFC sees these as irrelevant. By mentioning the issue of stigmatisation, Jenkins (2000) is clearly aware of the acceptability issue, but sees it as neither relevant nor durable: ‘at the time of writing, these sounds are still stigmatized in the L1 communities by speakers of RP, GA and other more standard L1 varieties’ (Jenkins 2000: 138). The implication is that once the non-native speaker majority has adopted the LFC substitutions, the native speaker minority will have to accept the new standard. On the face of it this appears to be a democratic development, but suggests that considerable ‘engineering’, in Bamgbose’s term, may be needed.

There is evidence, however, that some non-native speakers are uncomfortable with the possible identity-marking nature of dental fricative or TH substitution. In her study of attitudes towards aspects of ELF communication, Grau (2005: 268) asked trainee English teachers about such substitutions in *Thank you*. 59% did not agree with the proposition that ‘phonetic problems...should not be emphasised in the English language classroom if they are not a problem for understanding’; 22%

agreed and the remainder were unsure. The reasons given for disagreement included 'It sounds terrible and it sounds like: you want to but you can't'; and 'He or she is understandable with this problem, but he/she would surely be recognised by native speakers and that is what I (personally) don't prefer'. It could be argued that being recognised by NSs is less and less of a problem in NNS-dominated ELF communication, but even so there is a possibility that because many TH substitutions are L1-specific, not all L2 English speakers would like to advertise their identity in this way.

The additional, external evidence provided by inter-variety comparisons are an indication of general tendencies towards variation and change, and substitutions of both sounds are widespread. Although this is often used to support their use by non-native speakers, the evidence does not wholly support Jenkins' position. Brown (1991: 77) cites Wells (1982) as believing that while TH fronting (the substitution of [f, v] for /θ, ð/) is often popularly associated with Cockney accents, this does not mean that the dental fricatives are absent from Cockney consonantal inventories. Instead there is a variable rule which affects their realisation in particular contexts. Brown's conclusion is that such 'systemic conflation' is virtually absent in native accents of English, and that such conflations found in learners' speech 'should be viewed seriously' (1991: 77). The reality of intraspeaker variation is another dimension of language use that tends to be overlooked in studies of intelligibility and NVEs. If native speakers use these forms variably as part of their sociolinguistic competence (Hymes 1974), pedagogic intervention may be seen as denying learners the right to acquire this competence.

Jenkins' response would presumably be that the LFC does not take its bearings from 'native accents' and that such concern is misplaced, but there are some contradictions involved in basing a model for international communication upon existing varieties; arguably, the LFC is based on RP and GA insofar as it lists (the relatively few) permitted departures from these varieties. As van den Doel (2007: 30) puts it, 'if any of Jenkins' recommendations lead to increased intelligibility among non-native speakers, this is because many of the features of the Lingua Franca core are derived from native-speaker models'. It appears to be contradictory to use the presence of L1

accentual features such as TH substitution as a justification for their inclusion in the supposedly ‘post-NS’ LFC, as is noted by van den Doel (2007: 31), who believes that ‘if native speaker norms were actually completely irrelevant, it should not matter whether some non-native variation is similar to native variation’. The tendency for native speaker and non-native speaker systems to have relatively few differences at the phonemic level may be a result of the centripetal influence of the lexicon. According to Shockey (2003: 69), ‘the phonemic system is a product of the lexicon rather than the converse’.

The idea of removing these difficult sounds from pronunciation teaching syllabi has been put forward before. Brown (1974: 53) suggests that ‘when time is short it is probably not worth spending time on teaching /θ/ and /ð/ if the students find them difficult’ (in Jenkins 2000: 137). She goes on to recommend [f] and [v] substitutions, which are ‘acoustically similar...and bear a low functional load’ (Brown 1974, cited in Brown 1991: 72). However, going a step further by teaching substitutions, rather than simply accepting them as they occur, seems to raise several problems. The substitutions that are natural may not be the most intelligible, and so time and effort would be spent on teaching an unnatural alternative. These alternatives might also be unacceptable, for global users of English. The [f] and [v] substitutions recommended by Brown (1974) are exactly the substitutions used in Cockney, and this might not be a preferred identity for L2 users. Some teachers might use dental fricative substitutions variably, and some students might not find the dental fricative sounds difficult; thus there would be competing models in the classroom.

The conclusion of Deterding *et al.* (2008: 153) is that the dental fricatives are ‘something of a shibboleth in new varieties of English’. Although linguistic factors help to explain why substitutions arise, and suggest why they often persist in the feature pool, the interaction of linguistic and non-linguistic factors makes explanation and decision-making difficult. Applying the ‘precautionary principle’ to syllabus design and teaching leads to the conclusion that if these sounds are given a lower priority, learners should at least be made aware of the possible disadvantages of not acquiring them, although this of course depends on their level. The alternative would seem to be the kind of ‘engineering’ in which language learners form the

vanguard of change, altering existing linguistic and non-linguistic landscapes by sheer force of numbers.

2.9.2 L vocalisation

Another consonantal substitution supported by the LFC is that affecting postvocalic /l/ (the dark /l/, i.e. [ɫ], of RP, for example). The term ‘postvocalic /l/’ is used in this study, even though it may appear after coda consonants in the form of a syllabic /l/ (in words such as *middle*, for example). A common replacement for this sound is the L vocalisation that can be heard in many NS accents (e.g. Estuary English; see Rosewarne 1994, Altendorf 2003). As with the dental fricatives there seems to be a markedness explanation. Jenkins (2000: 139) claims that ‘the majority of RP speakers already pronounce pre-consonantal dark [ɫ] as /ʊ/ in non-careful speech’, although no evidence is cited. She adds (2000: 139) that ‘it thus seems unreasonable to have “higher” expectations of L2 speakers’.

The actual distribution of vocalised forms such as [mɪʊk] for *milk* is uncertain. In a survey of Estuary English (EE), Chia and Brown (2002: 35) conclude that ‘many features of EE are being adopted by RP speakers in order to sound less posh, i.e. these features are becoming features of RP too’. From the perspective of intelligibility, there seem to be few arguments against such substitution, as there are few conceivable minimal pairs; in the case of word-final dark /l/, however, it is conceivable that words such as *fill* and *few* could become homophones or near-homophones, especially given the fact that vocalisation may lead to vowel changes (see Wells 1982: 313). Jenkins (2000: 139), citing a personal communication from Joanne Kenworthy, mentions one such argument, however: the /ʊ/ substitution is not at present intelligible to American users. Jenkins concludes that due to its growing presence in (British) L1 accents, and the difficulty of acquisition experienced by L2 speakers, this substitution ‘is likely to spread in both L1 and L2 English’ (2000: 139).

It is unclear why, therefore, in an overview of ‘Euro English’ accents (Jenkins, Modiano and Seidlhofer 2001: 17) there is a different conclusion about ‘dark l’: ‘it

seems unlikely that this sound will be included in “Euro-English” pronunciation norms and more likely that it will be substituted with clear “l”. This may be because L vocalisation conflicts with the LFC’s other aim of reducing sound-spelling idiosyncrasies; as Szpyra-Kozłowska (2005: 156) notes, ‘this option seems to be ill-advised for international users of English whose pronunciation is largely spelling-based’. Brown (1991: 94) considers the options for pronunciation teaching in some detail. His conclusion is that L vocalisation poses no real threat to intelligibility, and may be used as a pronunciation target (along with ‘absorbed /l/’ after back vowels, as in pronouncing *fault* as [fɔ:t]) for those students who find dark /l/ ‘articulatorily too awkward’.

The acceptability question is less easy to answer, as there have been few specific studies into this feature. In terms of global influence, the Estuary English accent is as yet mainly confined to southern England, and its prestige value (as a ‘T-shirt among accents’, according to Rosewarne 2000; in Chia and Brown 2002: 34) is doubtful. However, L vocalisation also occurs in many L2 accents, including those of Hong Kong and Singapore. The results of one investigation into the perceptions of RP, EE and Singapore English accents amongst Singaporean listeners (Chia and Brown 2002) was that EE does not hold any great appeal for Singaporeans, but it is not clear whether this was due to its particular features or its overall impression. The conclusion of Brown (1991: 94) is that L vocalisation attracts a stigma ‘only in restricted phonological contexts’. These contexts are after non-labial consonants (for example, in *uncle*, *special* and *parcel*) and especially after alveolar plosives (e.g. in *little*, *middle*) where it would seem to be a ‘childish’ pronunciation (Gimson 1980: 203, in Brown 1991: 92).

Taking a broader perspective, and as mentioned in an earlier section, the external evidence from language change suggests that L vocalisation is a common development in many European languages, whose spellings in many cases reflect the historical phonological change (for example, French has *loyal* and *loyauté* where English has retained the orthographic ‘l’ in both *loyal* and *loyalty*). The retention of the orthographic ‘l’ in English spelling perhaps represents an unnatural restriction on a process of natural change, but it is nevertheless one that must be taken into account.

As with the dental fricative substitutions, postvocalic L substitutions will be the norm for many learners. However, their presence (and perhaps their variable occurrence, depending on situation and context) should be accepted as a variable feature of international English, without there being any particular need to actively promote their use.

2.9.3 Consonant clusters

Final consonant clusters are relatively uncommon in the world's languages, and their simplification in many varieties also appears to have an initial explanation based on markedness. The LFC recommends allowing learners to simplify consonant clusters in medial and final positions, but only insofar as they follow native speaker patterns of elision (Jenkins 2000: 159). The underlying pedagogic principles here are that addition is preferable to deletion, because the underlying form is more easily recoverable (Jenkins 2000: 142); that sounds in initial clusters should never be deleted; and that where elision occurs in a final cluster, /t/ or /d/ are the preferred candidates. In other words, and as in some other areas of the LFC, what is being recommended for learners is largely what native speakers already do; Gimson and Cruttenden (1994: 237) provide examples including the words *exactly*, *mostly*, and *facts* to show that elision of /t/ or /d/ often occurs when either forms the centre of a tri-consonantal cluster. In reality, many learners will adopt different patterns, as they try 'to approximate the target language as closely as possible while at the same time reconciling the structural requirements of the target language with those of their first language' (Peng and Setter 2000: 105). While the LFC assists in identifying which of these patterns are likely to be problematic, this area does not appear to offer any significant reduction in either the size or the scope of the task. Learners with L1s that do not allow complex syllable codas (for example, Cantonese) will need to pay special attention to this area.

An earlier recommendation for the teaching of consonant clusters was made by Brown (1991: 108). He recommends prioritising two-consonant final clusters ending with /t, d, s, z/ as they occur most frequently (often as a result of suffixation, as in *cats* for plural marking and *walked* for tense marking). Also, these inflectional or

bimorphemic clusters are less likely to be simplified in most varieties (see Schreier 2009: 60). In considering the elision of final three-consonant clusters, Temperley (1983) argues for a simplified target, taking the position that ‘since native speakers regularly fail to distinguish such pairs [as *tents/tense*, *bands/bans*], it seems unreasonable to insist that non-native learners should distinguish them’ (in Brown 1991: 109).

As has been mentioned before, it is not entirely clear whether the LFC is being too prescriptive in this area. There may be some patterns of non-native speaker simplification that are inconsequential for intelligibility. It is possible that both native speakers and non-native speakers will follow similar patterns of simplification, due to the operation of universal factors and constraints, although Schreier (2009: 68) notes that prevocalic final CCR is more widespread in non-native varieties of English. Little is known about the acceptability of such cluster reductions, and the experimental design of this study will allow further consideration of this area.

2.9.4 Other features: consonant substitutions

The LFC’s recommendation that RP will need to rediscover its rhoticity is neatly counterbalanced by its admonition against American intervocalic flapping; both groups of speakers will have to sacrifice something, perhaps providing custom for the ‘adult EIL classes’ foreseen by Jenkins (2000: 228). As noted by Trudgill (2005: 88), it is unclear why the LFC recommends a strongly retroflected American /r/, phonetically [ɹ], when all /r/s are rhotic; it is accents and their phonotactic rules, not the ‘r’ sounds themselves, that are rhotic. A possible problem is that the sound in question is relatively rare in the world’s languages, so that the LFC ‘postulates a segment problematic for many international learners of English’ (Szpyra-Kozłowska 2005: 153). There are many possible /r/ substitutions, depending on the L1, including uvular fricatives and alveolar trills. Specifying a particular substitution would therefore seem to be unhelpful. On the other hand, promoting rhoticity would remove a potentially confusing sound-spelling disjunct. Brown (1991: 104), in his survey of pronunciation models, provides some support for Jenkins by concluding that rhoticity in general brings ‘certain pedagogical advantages, including reliability of English

orthography...[i]t should therefore be given serious consideration as a possible feature of pronunciation models’.

In the LFC core, the British pronunciation of words with intervocalic /t/, such as *writer* [raitə], is preferred to the American [raɪrə], with its voiced alveolar tap. This gains its justification from the possibility of confusion between minimal pairs such as *writer* and *rider*, as well as the fact that the stipulated alternatives have a ‘more reliable relationship with orthography’ (Jenkins 2000: 140). The general recommendation of the LFC regarding consonants, that close approximations of consonant sounds are acceptable, is uncontroversial. This provides another example of how the LFC gains its bearings largely from NS accents – despite the repeated claim that native speakers are no longer relevant. As Trudgill (2005: 88) observes, the LFC’s proposals are ‘extremely modest’ in this area.

2.9.5 Other features: phonetic characteristics

While not as immediately relevant to Hong Kong English, some other features of the LFC will be briefly considered in this section. A somewhat puzzling requirement of the LFC is another feature of NS accents, namely the aspirated nature of the fortis plosives /p, t, k/ in initial position: ‘Without the help of this puff of air, a listener will find it more difficult to identify the sound as voiceless’ (Jenkins 2000: 140). There are many minimal pairs distinguished by, for example, the /p, b/ pair, so the maintenance of ‘perceptual distance’ here would appear to be important. However, and in a way which once again draws attention to the largely NS-centred nature of the LFC, this distinction is not equally important for all speakers and listeners. According to Szpyra-Kozłowska (2005: 157), in languages that do not make use of aspiration, such as Polish, Hungarian and German, it is voicing rather than aspiration that makes the distinction; this raises the question, somewhat problematically for the LFC, of the extent to which a ‘common core’ exists for either production or reception. Some research has indicated that sound contrasts may be processed differently by L2 listeners. Hung (2000: 344), for example, found that Hong Kong subjects were better able to differentiate words such as *hood* and *hoot* (containing /u/ and /u:/, in RP) when read by a Hong Kong speaker, than when pronounced by a

native speaker (although Hung explains this as being due to differences in the release phase of the final consonants). In addition, many accents of the world (for example, Indian English and Yorkshire English) feature reduced aspiration of initial plosives (Collins and Mees 2003; Wells 1982). Requiring aspiration would seem to add to, rather than subtract from, the learning burden and would also introduce an additional and somewhat technical concept into the classroom.

A second, related requirement of the LFC is the retention of the phenomenon of ‘pre-fortis clipping’ observable in pairs such as *cap*, *cab*, where the vowel of the former is phonetically shorter than, but phonemically equivalent to, that of the latter. Jenkins (2000: 141) gives two reasons for this stipulation: firstly, learners do not automatically shorten pre-fortis vowels, so pairs such as *seat*, *seed* may be hard to distinguish; secondly, such shortening eases articulation because longer vowels and fortis consonants require more muscular effort than shorter vowels and lenis consonants. As Jenkins notes, ‘[t]his is presumably why proficient speakers automatically make this reduction’ (2000: 141). This quote suggests that proficiency is a factor in explaining certain types of variation, and further implies that some processes are acquired naturally by all speakers, whatever their language background, as part of their acquisition of the system of contrasts.

The likelihood of ‘natural’ acquisition by L2 English speakers is further indicated by the fact that this feature is ‘not generally included, let alone prioritized, in pronunciation courses’ (Jenkins 2000: 140). The LFC would thus seem to be adding to both the learning burden, by specifying a feature which is probably acquired naturally at a certain stage, and to the teaching burden, because of the technical nature of the phenomenon. Jenkins contends that this ‘basic pedagogic rule’ is simple and learnable in the classroom, but the problem is likely to be in applying the rule. Successful production involves already being able to make the necessary distinctions (in terms of both quantity and quality) between long and short vowels, something which Dauer (2005: 347) calls ‘the bane of pronunciation teachers’. In short, Jenkins’ treatment of aspiration and pre-fortis clipping appears to oversimplify these complex areas, and there are few clear indications for pronunciation teaching.

2.9.6 Vowel quality

As many L2 varieties of English (including Hong Kong English) tend to have ‘reduced’ or ‘simpler’ vowel systems (see Hung 2000), some consideration of the possible effects of vowel substitutions and mergers is necessary. The core stipulation of the LFC regarding vowel quality is that ‘L2 vowel qualities are permissible if consistent’ (Jenkins 2000: 159). Assuming that ‘consistent’ means ‘L1-influenced’, this is uncontroversial. All speakers have idiolectal vowel qualities; Preston (2005), as a native speaker of American English, lists four areas in which his own acoustically analysed vowel qualities differ significantly from the GA model. What is more controversial, however, is that Jenkins appears to take no account of speakers’ abilities to produce vowel contrasts, recalling Brown’s concern (above) about ‘perceptual distance’. L1-influenced vowel qualities are not problematic in themselves, but may become so if speakers merge pairs of vowels such as the TRAP/DRESS contrast (in RP, /æ/ and /e/). This is one of the contrasts identified by Brown (1991) as having a high cumulative frequency in English. It is problematic for learners from many language backgrounds, including Cantonese (Chan and Li 2000); van den Doel (2007: 31) mentions it as being problematic in ‘countless varieties of L2 English stretching from Korea to Turkey and beyond’. This aspect of the LFC has attracted surprisingly little comment, but Szpyra-Kozłowska (2005) notes the tendency of Polish learners to employ fewer vowel contrasts, so that several vowels are realised as one. Although the LFC research did not uncover instances of miscommunication resulting from vowel quality, it seems highly likely that the large number of homophones in some learners’ productions constitute ‘a serious threat to intelligibility’ (Szpyra-Kozłowska 2005: 164).

Reducing the importance of vowel quality in the syllabus seems to be one of the main ways in which the LFC achieves its aim of reducing the size of the task: ‘[w]ith the removal of nineteen of the twenty RP vowel and diphthong qualities, this was a likely outcome’ (2007: 27). The one remaining vowel, according to the LFC, would be /ɜ:/ in the NURSE lexical set; interestingly, Deterding and Kirkpatrick (2006) also found instances of miscommunication resulting from the substitution of [ɑ:] for this vowel. But while we can conclude that some vowel contrasts are more worthy of

attention than others (and here considerations of functional load are once again highlighted), there seems to be little in the way of empirical evidence or linguistic principles to support downgrading the teaching and learning of vowel quality. The apparent failure of the LFC to distinguish between individual sounds and the system of contrasts in which they participate, and the neglect of developmental aspects of L2 vowel systems, are potentially serious weaknesses. Furthermore, given the fact that vowel substitutions are likely to be highly noticeable, the acceptability of these substitutions for other users of the same variety needs to be given careful consideration.

Trudgill (2005: 219), in a discussion of what he calls the ‘speaker-listener equilibrium’, points out some of the possible problems associated with reduced vowel systems in international communication. While systems involving fewer contrasts may be easier to learn from the speaker’s perspective, they present problems for non-native listeners because such listeners ‘typically require more information, not less’ – because of the greater difficulties involved in dealing with distractions such as background noise, and because of Weber and Cutler’s (2004) concept of ‘spurious competitor words...[e]ven though the non-native listener knows fewer words of the language than the native listener, the total competitor population may be larger, because inaccurate phonetic processing allows spurious phonemic matches’ (in Trudgill 2005: 221). Trudgill’s conclusion regarding the necessary equilibrium between the needs of speaker and listener is that the teaching model should contain the ‘maximum number of readily attainable contrasts’ (2005: 226).

An aspect of vowel quality that does not receive much attention in the LFC is the non-reduction of full vowels (or, the avoidance of schwa) in the unstressed syllables of words with two or more syllables. In Hong Kong this can often be heard in the initial syllables of words such as *contain*, and in the final syllables of words such as *appropriate*. This aspect of vowel reduction was considered in section 2.6.3.1, and was thought to be unlikely to cause intelligibility problems. Its acceptability may depend partly on how noticeable the substitution is; its occurrence in unstressed syllables suggests that its noticeability would be low. This study’s empirical research will attempt to further evaluate the acceptability of this feature.

2.10 General pedagogical factors

The preceding sections have considered various aspects of pronunciation models and phonological features. However, an aspect that is missing from the evaluation model is the pedagogical perspective. This section will therefore assess how teaching and learning considerations may affect the nature and implementation of pronunciation models.

2.10.1 Models and goals

There is a important but often neglected distinction that needs to be made in any consideration of pronunciation teaching, that between *models* and *goals*. This study will use the term *model* to refer to any examples or explanations of the target language (for example, the teacher's voice, recordings or lists of phonetic symbols) that are used for pedagogical purposes to teach aspects of pronunciation. Models will not always be intended for imitation, although repetition and pattern practice may follow; their purpose is to provide input, to develop learners' phonological systems, particularly in terms of their ability to make sound or meaning contrasts, and to raise awareness of the uses of phonological features in communication. A *goal*, on the other hand, is a criterion for achievement, a target for the learners to attain. It may be highly variable, depending on such factors as the age, motivation, existing level and future aims of the learners.

The above definition of a 'model' coincides with the concept of a 'pedagogical norm', as elaborated by Valdman:

A pedagogical norm, like all norms, is an abstraction. Its distinguishing feature resides in the fact that it is an artificial construct reflecting the special conditions of classroom foreign language learning (Valdman 1989: 272, in Magnan and Walz 2002: 28).

Pedagogical norms, in Valdman's view, should take account of sociolinguistic variation as well as linguistic description. By considering the conditions of language

use and the ease of acquisition, they seek to provide appropriate and attainable targets for the arduous task of language learning.

The distinction between models and goals is also made by Dalton and Seidlhofer (1994b: 2.6-2.7), who see models such as RP and GA as ‘points of reference and models for guidance’ and as ‘pedagogic means to achieve the end of communication for specific learners’ (in Jenkins 2000: 18). This may be taken to imply that the particular model does not really matter, as long as it is consistent. But while Jenkins is aware of the possible validity of the model/goal distinction, her concern is still the relevance of the model chosen; in other words, accepting the idea of separate functions for models and goals means ensuring that both are pedagogically appropriate.

Not all commentators accept the usefulness of a model/goal distinction. Brown (1991) is concerned about the potential for confusion created by differences between models (in the forms of teaching materials and the local teacher’s accent) and students’ actual performance. His recommendation is that the goal for students should be ‘the educated local pronunciation of the teacher, which is also the model contained in pronunciation materials used in classwork’. Although some kind of differentiation between models and goals may be a pragmatic choice in many learning situations, Brown’s suggestion has the appeal of pedagogical coherence and of valuing, rather than devaluing, the accents of local teachers. The importance of identifying the nature of an ‘educated local pronunciation’, and of evaluating its features from the competing perspectives of intelligibility, acceptability and so on, is highlighted.

Thus, a variationist, features-based approach to the evaluation of pronunciation models is also indicated by a consideration of pedagogical factors. Bamgbose (1998) identifies three kinds of norms: code norms, or the standard varieties of a language; feature norms, which refer to the properties of a language at any level, from the phonetic to the orthographical; and behavioural norms, which are the set of conventions underlying patterns of interaction. The question of which model to adopt

can perhaps be most satisfactorily answered by considering feature norms, rather than choosing between code norms.

2.11 Ideology and pedagogy

Finally, to understand the ‘models debate’ it is necessary to look in more detail at the interaction between ideological and pedagogical positions. Thus far, this chapter has skirted around the fact that in the debate about pronunciation models, there are ideological undercurrents that must be understood if the competing arguments are to be evaluated. The word ‘undercurrents’ is not meant to imply that there is anything dangerous or unhealthy about the presence of ideologies; it is only natural that innovations in any field are partly driven by movements and trends within wider socio-political and intellectual worlds. It is advisable, however, that whenever possible, ideological positions need to be identified and brought nearer to the surface, especially when they claim not to be ideologically based. As Joseph and Taylor (1990: 2) put it: ‘It is our belief that any enterprise which claims to be non-ideological and value-neutral, but which in fact remains covertly ideological and value-laden, is the more dangerous for its deceptive subtlety’ (in Milroy 2001: 531). This warning could apply equally to the ‘conservative’ view, in which unanalysed native speaker models are seen as the default option for language teaching, or to the ‘progressive’ view, as advanced by the ELF paradigm.

As a profession, ELT seems to be particularly vulnerable to the kind of ‘pendulum swing’ that leads to rapid changes in practice. The widespread adoption of variants of task-based learning during the 1990s is seen by some as being at least partly due to the ‘spirit of the times’ (Ur 2006, in Waters 2007: 353), including a prevailing ‘distaste for control, whether personal or linguistic’ (Swan 2005: 388). According to Swan (2005: 376), there is little empirical evidence to support this innovation, which exemplifies what he sees as ‘a recurrent pattern of damaging ideological swings in language teaching theory and practice’. Waters (2007: 358) believes that the promotion of ‘non-standard’ approaches, including that represented by ELF, is related to the influence of ‘political correctness’ in ELT. He sees an

‘unacknowledged ideological basis’ which regards NNSs as ‘a “victim” class about whom generalizations are just as freely made’. The perceived need to ‘liberate’ learners from oppression is a recurring theme in some of the literature:

In a very real way, the choice of a nativized model over a native speaker model is the choice of democracy over imperialism (Kirkpatrick 2006: 76).

It is time, then, for applied linguists to provide a description of lingua franca English, for by doing so they can liberate the millions upon millions of people currently teaching and learning English from inappropriate models (Kirkpatrick 2006: 81).

It is Waters’ contention that a ‘reductionist’ stance exaggerates the extent to which NSs exert hegemonic domination, and ignores the different perspectives of NNSs. Such viewpoints raise the questions of who is liberating whom from what, and whether ‘they’ wish to be liberated; a somewhat crude analogy could be made with Western attempts to ‘liberate’ or ‘democratise’ certain states. A frequently-cited example of a non-native speaker who is apparently resisting liberation is Kuo (cited by Holliday 2005; in van den Doel 2007: 29):

It’s been clear that I’m a language learner from the periphery and – listen to this –
I prefer to speak for myself!

Kuo herself argues (2006: 220) that L2 learners should be allowed to decide ‘which English to learn’ and believes that a native-speaker model would appear to be more appropriate and appealing than a description of English which is ‘somewhat reduced or incomplete’.

The problem for many teachers, on the other hand, is that they may not have sufficient time to engage fully with the debate. The politically-charged rhetoric may therefore be misleading. Canagarajah (1999b: 207) expresses the practitioner’s desire to steer a course between competing positions:

These are, after all, times when academic discourses, spawned freely in opposition to each other, swing wildly between extremes like a pendulum. As a teacher, focused on the concerns of my students, I negotiate with these divergent rhetorics to consider how they may develop a richer awareness of language and social life, enabling me to act more rewardingly in the classroom.

While it is certainly necessary to be aware of the possible dangers of ideologically-propelled ‘pendulum swings’, a climate of excessive conservatism could also be criticised as stifling much-needed innovation or as protecting vested interests. It is therefore necessary to evaluate the alternatives from a range of perspectives, while maintaining an appropriately critical stance towards the ideological issues that have helped to drive the debate.

2.11.1 Contested terminology

In any instance of actual or perceived conflict, language is itself prone to becoming a contested area. An awareness of this is visible in Jenkins’ work, for example in her earlier insistence on using the term ‘bilingual English speaker’ instead of ‘non-native speaker’ – despite the fact that many L2 English users are far from being bilingual in the generally accepted sense of the term. While it is plausible that the inherent negativity of the term ‘non-native speaker’ is disadvantaging in many cases, the possible alternatives are also problematic. Jenkins herself reverts to the NS/NNS ‘dichotomy’ in her later work (e.g. Jenkins 2007).

A similar attitude towards terminology can be seen in Brutt-Griffler (2002: 179), who posits the existence of a vertical, unidirectional ‘power relation’ between native and non-native speakers, so that authority and the power to influence other varieties inheres in native speaker communities. The reasons for this include the age of the variety, or power in general, but they are always ‘political criteria’. As an alternative, Brutt-Griffler proposes a bidirectional model (2002: 180) in which the point of reference or grounding of the language is the world community rather than the mother tongue English communities. There is nothing to object to in this formulation, but what is striking is the apparent belief that the world will change as a

result of this reconceptualization: ‘This conception does away with hierarchy among speech communities, so that community relations are better depicted’ (Brutt-Griffler 2002: 180). This is perhaps an example of what Bourdieu (1991: 53) refers to as ‘scholarly relativism’, a view that neglects deeper structures of inequality and the existence of prejudiced attitudes, even among those who are most disadvantaged by them.

Whatever the relationship between terminology and the real world, and whatever the relationship between applied linguistics and the real world, the examination of the various factors in this chapter has indicated that certain aspects of terminology need to be defined very carefully for the time being, and brought up for review whenever possible. These include the native speaker / non-native speaker ‘dichotomy’, and the undifferentiated use of ‘variety’ to denote highly variable language systems. This study will attempt to employ these terms with care.

2.11.2 Understanding the debate: ELF and globalization

Another useful approach to the debate is gained by considering different points of view about globalization. Dewey (2007: 334), drawing on the work of Held *et al.* (1999) identifies three ways of conceptualising the phenomenon: hyperglobalist, sceptical and transformationalist. Hyperglobalism sees globalization as the ‘key defining force of the current epoch’, one which is constructing new economic, social and political world orders with an overall tendency towards homogeneity; this viewpoint tends to equate globalisation with Westernisation. The sceptical position, however, sees the apparent interdependence brought by globalization as operating only at a surface level. Transformationalists, as their name suggests, appear to be optimistic about the potential for positive change brought about by globalization; it can lead to information exchanges becoming more democratized and less hierarchical (Dewey 2007: 344).

Within the ELF debate, according to Dewey, there appear to be hyperglobalisers such as Phillipson (1992), who are preoccupied with hegemonic power relations in the form of, for example, ‘linguistic imperialism’; sceptics such as Mollin (2006)

who are unconvinced of the need for major changes and tend to support the use of native speaker models; and transformationalists, including Jenkins (2000, 2007) and Kirkpatrick (2006). In the end, it seems likely that attitudes towards potential innovations such as ELF will tend to correlate with attitudes towards non-linguistic phenomena, such as globalization. They will also depend on whether one sees language attitudes as being malleable, and thus as potential causes of societal change, or as being primarily the effect of deeper, underlying social conditions. It is useful to bear this in mind while considering the somewhat polarised nature of the models debate.

2.12 Summary

This chapter has summarised the background to the models debate in language teaching, and has related it to wider discussion within the fields of applied linguistics and sociolinguistics, including the ELF and World Englishes movements. A consideration of various factors led to the development of a four-quadrant evaluation model, which included both linguistic and non-linguistic factors. After establishing the need for a features-based, rather than a varieties-based, approach, it was argued that the intelligibility of features is largely predictable with reference to linguistic and psycholinguistic criteria, with supplementary evidence being obtained from language variation and change. The non-linguistic factor of acceptability was therefore identified as the research focus of the study. The evaluation model was then used to assess the viability of certain changes to existing pronunciation teaching syllabi proposed by the Lingua Franca Core (LFC). While this preliminary evaluation demonstrated the complexity of the relationships between various factors, it was able to identify some likely candidates for further evaluation in the experimental part of the study. These included features such as dental fricative substitutions, L vocalisation, final consonant cluster reduction and certain types of vowel substitution. The evaluation was further extended by considering some of the pedagogical factors that may influence the selection and introduction of alternative models, and some of the competing conceptual and ideological positions that underlie the models debate.

CHAPTER 3

THE HONG KONG BACKGROUND AND THE PRELIMINARY STUDY

3.1 Introduction

This chapter will extend and localise the themes introduced in Chapter 2. It begins by discussing some of the social and educational factors that affect language attitudes in Hong Kong, and reviewing some previous studies of attitudes towards the Hong Kong English accent. Moving to a focus on phonological descriptions of Hong Kong English, it identifies some of its characteristic features by reviewing previous studies. As well as this features-based orientation, the chapter also argues for the desirability of obtaining data on language use by ‘high proficiency’ speakers of Hong Kong English. Accordingly, the results of a preliminary study of the occurrence and distribution of certain consonantal features of Hong Kong English within a ‘mini-corpus’ are also presented in this chapter. By showing which features tend to persist, this part of the study serves as a prelude to the main study and acts as a further indication of the possible candidates for acceptance in local language teaching and testing. The chapter concludes by reviewing the study’s research space and stating its research questions.

3.2 Language in education

As one of Hong Kong’s two official languages, English has an important status. This was certainly the case before the 1997 return of sovereignty, when almost all interviews for government or large business corporations were conducted in English (Talbot, Atkinson and Atkinson 2003). Since then, and despite the increasing importance of Mandarin Chinese or *putonghua*, English ‘still plays an important role in the business sector, in the workplace, and especially in higher education’ (Lee

2005: 36). It is seen as the language that carries ‘outer’ values including ‘success, stylishness, and academic achievement’, while Cantonese is associated with ‘inner’ values connected with ‘tradition, home and solidarity’ (Pennington 1998: 13; in Glenwright 2005: 206). This distinctiveness leads Glenwright (2005: 206) to conclude that there is an incipiently triglossic language situation in Hong Kong, further noting that ‘the position of English in Hong Kong society and its implications for identity and education represent a particularly contentious and divisive issue’.

One problematic area is the medium of instruction (MOI) policy. Before 1997, the popularity of English medium of instruction (EMI) schools was ‘unrivalled by the Chinese medium of instruction (CMI) schools’, according to Talbot *et al.* (2003: 276). Because the universities used English as their MOI, a good pass in English was essential for entrance. By 1997, 94% of secondary schools were EMI (Lee 2005: 36), despite the fact that a roughly equal percentage of children (97%) spoke Cantonese at home (City University of Hong Kong 1999, in Talbot *et al.* 2003: 277). The MOI policy arguably neglected the standard and quality of education for the majority, but also reflected the importance attached to English as a source of linguistic capital (Talbot *et al.* 2003), especially by parents.

Reforms after 1997 increased the number of CMI schools, and the proportion of such schools reached around 25% by 2003 (Talbot *et al.* 2003: 281). At the time of writing the MOI pendulum appears to be swinging back towards English, with more schools being allowed to choose their medium of instruction under a ‘fine-tuning’ policy initiative. As in many places, English language teaching in Hong Kong is to a considerable extent a market-driven enterprise, and the freedom to choose the MOI will almost certainly result in more EMI teaching. According to the *South China Morning Post* (2009c), the majority of schools in Hong Kong will use English as the medium of instruction as soon as the relevant legislation allows them to. But despite the contentious issues, Hong Kong’s overall attitude towards English can be summed up as being one of pragmatic acceptance of its economic and technical importance; according to Talbot *et al.* (2003: 285) this is ‘less of an ideological matter and more a signification of its global importance’.

3.2.1 Other issues in language education: declining standards?

Another ongoing debate in Hong Kong surrounds the perception of ‘falling standards’ in English. While proclamations by business leaders and popular discourse usually support the idea that standards have declined, scholarly opinion suggests that demographic and educational changes, including increased access to higher education, are the main reason for the perceived change. One observation is that ‘more people than ever are speaking “good” English, and more people than ever are speaking “bad” English’ (Bacon-Shone and Bolton, 1998: 84). For Joseph (2004: 147) the ‘emergence of Hong Kong English’ and the ‘decline of English standards in Hong Kong’ are actually the same phenomenon, seen from two different points of view; while linguists seem preoccupied with ‘emergence’, popular discourse sees ‘decline’. Educational policymaking has tended to follow the widely-held belief that standards need safeguarding, leading for example to the instigation of the Language Proficiency Assessment for Teachers (LPAT). A satisfactory grade in this test, which has both English and Putonghua versions, is a requirement for teachers of these languages who wish to work in government-funded schools. The entry requirements for new teachers have also been tightened, as the language teaching workforce had hitherto been regarded as ‘notorious for its overall lack of training’ (Lee 2005: 37). The combined effect of such initiatives is intended to be an increased professionalisation of the teaching workforce, although critics maintain that teachers are being unfairly blamed for a largely imagined ‘decline in standards’ (see Glenwright 2005).

3.3 The role of pronunciation in language teaching in Hong Kong

Turning to the actual position and nature of pronunciation teaching in Hong Kong, there is little evidence to suggest it has an important role. Factors such as the exam-oriented learning culture, large class sizes and heavy workloads (Lee 2005) tend to militate against focusing on speaking skills, including pronunciation, although there are speaking components in public examinations. Another constraint is likely to be teachers’ own lack of training in and awareness of English phonetics and phonology;

in fact, there is also concern about the teaching of Cantonese pronunciation. The Standing Committee on Language Education and Research (SCOLAR 2003) recommended that '[m]ore attention should be given to the teaching of grammar, phonics and phonetics in English Language and Cantonese pronunciation'.

3.3.1 Pronunciation models

Whether Hong Kong is classified as ESL or EFL in nature, models and norms are generally derived from outside the region. In common with other norm-dependent situations, local language teachers are faced with a dilemma. Teaching materials and examinations are oriented towards NS models such as RP or GA, but their own accents are likely to be quite different. As Kirkpatrick (2007b: 381) points out, they are 'required to teach a model which they themselves do not speak', although there is little data on language use by this group of speakers. This separation almost certainly has negative effects on their confidence. There is also a lack of awareness regarding which accent features may cause intelligibility problems, or which should be prioritised for other reasons.

Current syllabus specifications do, in fact, show some acceptance of the concept of partly divergent, but mutually intelligible, varieties. In an article written by two of the consultants involved with the development of the LPAT 'benchmarks' (Coniam and Falvey 2002), there are assessment scales and descriptors for both the Classroom Language Assessment (the CLA, involving 'real-life' assessment in the classroom) and the Speaking Test (a more controlled context involving reading aloud or discussion with an interlocutor). These are reproduced in Table 3.1 below.

Table 3.1. ‘At the benchmark’ descriptors for pronunciation from two tests (from Coniam and Falvey 2002: 29).

	CLA	Speaking Test
<i>Level</i>	<i>Pronunciation, stress and intonation</i>	<i>Pronunciation, stress and intonation</i>
At the benchmark	Pronunciation of sounds is generally acceptable although there are some errors in the pronunciation of sounds and/or word stress and a number of L1 characteristics are evident but not too obtrusive. Sentence stress and intonation patterns may sometimes be inappropriate but communication is seldom impeded.	Although there may be some errors in the pronunciation of sounds and/or word stress and a number of L1 characteristics are evident, pronunciation is unlikely to present problems for L2 speakers. The candidate is fairly confident about the pronunciation of words. Sentence stress and intonation patterns may sometimes be inappropriate but reading of the text is seldom impeded and is acceptable for classroom communication.

While these minimum requirements can be said to reflect recent thinking to some extent, insofar as L1 characteristics (albeit unspecified) are accepted, the ‘well above the benchmark’ descriptor below (from Coniam and Falvey 2002: 23) suggests that they are still, as noted by Kirkpatrick (2007b: 380), ‘measured against idealised native-speaker norms’:

Pronunciation is completely error-free with no noticeable L1 characteristics...[a]ny mistakes that occur can be categorised as lapses rather than systematic errors.

Luk and Lin (2006: 10) note that such descriptors ‘imply that *error-free* pronunciation is also *accent free*’ (emphasis in original). While the influence of the so-called ‘standard language ideology’ is clearly visible, according to Coniam and Falvey (2002: 19) this ‘educated’ L2 speaker model actually has its origin in the arguments put forward by Jenkins (1998: 125): ‘non-native teachers will, themselves, still be required to develop the ability to approximate more closely than their students to a standard native norm’. This is an interesting proposition, although in her later

work Jenkins does not appear to have maintained such a conciliatory standpoint. To some extent, it makes intuitive pedagogical sense, given that the level of learner attainment is likely to be lower than that of the teacher. However, once again there needs to be a clearer indication of how closely teachers need to approximate a native norm, and of which features are most important from the perspectives of international intelligibility, local acceptability, and so on.

For the time being native speaker models serve as the default standard, while within the discourse of applied linguistics there is no clear indication of what form an acceptable and intelligible model might take. In Hong Kong, this leads Kirkpatrick (2007b: 387) to call for a ‘codified description of the local bilingual variety of English at the phonological, lexical-grammatical and discourse-pragmatic levels, as exemplified by highly proficient users of English who are mother-tongue speakers of Cantonese’. The case for such a description is persuasive, but there remains the difficulty of selection and the likelihood that linguistic description will be insufficient – factors such as intelligibility and acceptability, as included in the evaluation framework, will also need to be considered.

The need for research into non-native norms is echoed by Luk and Lin (2006: 18), who call for ‘the establishment of a regional variety of English with high mutual intelligibility’. For Luk and Lin (2006: 17), this includes the possibility of multiple competences in different varieties, to suit different contexts and purposes. While aware of Jenkins’ research, they believe that her criteria ‘might not fully reflect phonological features of native languages in the outer and expanding circles, and might not fully address pronunciation-based communication problems’.

Luk and Lin’s first point reflects the tension, already mentioned in Chapter 2, between the ‘autonomy and creativity’ of local varieties and the need to maintain international intelligibility. One possible problem with establishing regional norms is the great variation that exists in terms of language families and L2 English pronunciation features. Luk and Lin’s suggestion for an ‘Asian Pacific variety of World English pronunciation’ (2006: 18) is attractive, and resembles the regional view of studies such as Deterding and Kirkpatrick’s (2006) survey of ASEAN lingua

franca English, which they claim has systematic features. But the ‘Asia Pacific’ region is normally taken to include countries as diverse as China, Japan, Korea, Malaysia and the Philippines. The wide range of L1-derived sound substitutions makes a single variety appear less likely than a ‘common core’ resembling Jenkins’ original concept, in which variation is permitted within the constraints provided by the need for mutual intelligibility. It is, after all, a common ‘misconception’ that the goal of ELF is ‘to establish a single lingua franca norm to which all users should conform’ (Jenkins 2007: 19). There is a need for more research into local varieties, including the subvarieties that may exist as a result of factors such as proficiency, attitudes and exposure to the language. Such research will also need to take account of the language attitudes that exist in society as a whole, if (for example) the attainment of ‘pedagogical acceptability’ is desired.

3.4 Hong Kong English: status and attitudes

There are divergent views on whether and on what terms Hong Kong English can be said to exist. That NVEs are an important area of study is undeniable, but there is often a wide gulf between the discourse of linguists and that of the rest of the population, as noted above. Even in the research literature there seems to be more doubt about the status of HKE than of other varieties such as Singapore English, although the consensus of recent research is that HKE is an ‘emerging’ variety that is starting to develop its own norms (see, for example, Bolton 2000; Setter 2008). This section will examine the status of HKE within a broader discussion of language attitudes in Hong Kong.

3.4.1 Assessing variety status

In a European context, Mollin (2006) assesses the variety status of ‘Euro-English’ (see Jenkins, Modiano and Seidlhofer 2001), concluding that it does not meet the three criteria of function, form and attitude. In Mollin’s analysis, ‘function’ is concerned with the use of a variety in different domains such as the media or literary creativity. ‘Form’ relates to the development of ‘unique linguistic features, which

need to be communal and systematic' (Mollin 2006: 198), while 'attitude' considers the degree of acceptance of the new variety by its speakers. Perhaps unsurprisingly, Mollin (2006: 199) concludes that according to these criteria nothing that merits a 'variety' label exists in Europe, and that the term 'Euro-English' should be discarded. In Hong Kong this is the general view that has been taken, to differing extents, by Luke and Richards (1982), Li (2000) and Stibbard (2004). Stibbard's study of phonological variation leads him to argue against the existence of Hong Kong English phonology on 'phonological grounds', namely that "the instability of the accent, the repeated co-occurrences of phonemic overlap in the data, and the fact that for the most part the pronunciation is clearly due to transfer from Cantonese, all undermine the attempt to establish a "phonology of Hong Kong English"" (Stibbard 2004: 140). This view is contested by Hung (2000: 337), whose study concludes that Hong Kongers share a 'common underlying phonological system'. Hung avoids drawing any conclusions about the existence of a separate variety, however.

These divergent opinions indicate a need for further research, and for a clarification of what is meant by 'Hong Kong English', but such arguments are often set against a backdrop of community-wide scepticism; the idea of HKE as a *bona fide* variety is not widely accepted in Hong Kong (see, for example, Luk 1998). Thus, using Mollin's criteria, it appears that while the forms of Hong Kong English certainly exist, the functions of the variety and societal attitudes towards it tend to weaken claims of varietal status. However, ambivalent attitudes are characteristic of certain stages in the development of language varieties (see Schneider 2003). It is also important to remember that language attitudes are not immutable, and that one of the goals of language policy may be to change them. In Hong Kong, the MOI policy was discussed in the *South China Morning Post* (2008) under the headline 'Language policy change to "alter public mindset"'. There may be ways in which language policy, whether in the form of official pronouncements or decisions about classroom language and teaching materials, can be the cause, rather than the result, of language attitudes.

Dynamic models of NVE emergence, such as the one proposed by Schneider (2003), provide a useful perspective in that they attempt to assess a variety's stage of

evolution. Schneider proposes that new language varieties must pass through the stages of 1) foundation, 2) exonormative stabilization, 3) nativization, 4) endonormative stabilization and 5) differentiation. His assessment of Hong Kong English is that it is currently at stage 3, with some traces of stage 2 still observable (Schneider 2003: 258). Evidence for nativization includes the educational policy objective of ‘mass bilingualism’ (Bolton 2000, in Schneider 2003: 259), and the identity construction of Cantonese Hongkongers who no longer have a colonial ‘them vs. us’ mindset and instead display a ‘distinctive and healthy Hong Kong identity’ (Hyland 1997: 207, in Schneider 2007: 136).

Schneider accepts Bolton’s (2000) assertion of the need for discussion and discourse about Hong Kong English, contrasting this viewpoint with the ‘denial’ position taken by Luke and Richards (1982; in Schneider 2003: 260). In terms of phonology, evidence for the stage 3 ‘nativization’ of Hong Kong English is also provided by Hung’s (2000) account of a Hong Kong English accent which can be described phonologically and which, *contra* Stibbard (2004), is not reducible to the phonology of either Cantonese or English (Hung 2000: 354). There are also signs of positive attitudes, as according to Bolton (2000: 277) the local accent is ‘beginning to be regarded as a positively evaluated source of identification’. However, as mentioned above, the ‘variation problem’ significantly complicates the process of description, and may compromise the applicability of some descriptions of HKE. There seems to be a great need for a description of high-proficiency Hong Kong English, as called for by Kirkpatrick (2007b).

3.4.2 Attitudes and identity

The existence of exonormative language attitudes in Hong Kong is well known; the highly norm-referenced standards of local teachers are documented by Tsui and Bunton (2002). This is despite the emergence of the World Englishes paradigm, which has generally resulted in ‘a more liberal attitude towards local varieties of English’ (Luk and Lin 2006: 3). In Hong Kong there is little evidence of the influence of this paradigm, except perhaps in the descriptive studies of Hung (2000) and Deterding *et al.* (2008). According to Luk and Lin (2006: 11), the signs of

exonormative orientation include the instigation of the LPAT test, a deferential attitude towards NETs (native English-speaking teachers) and the existence of media and public discourses on ‘proper’ English pronunciation. Interestingly, this discourse also extends to L1 Cantonese pronunciation; there are Cantonese pronunciation classes for adults in Hong Kong and the disapproving term ‘Lazy Cantonese’ occurs frequently in media discourse (e.g. *South China Morning Post* 2007).

While Luk and Lin acknowledge the growing awareness of ‘linguistic human rights’ (see Lippi-Green 1997), they also take a more cautious stance by noting that the assertion of ‘rights’ is associated with ‘gains and losses of social and cultural capital’, in Bourdieu’s (1991) terms. This may be an aspect of the frequently-mentioned ‘pragmatic’ nature of Hong Kong people, who at the moment appear to prefer striving for various forms of ‘capital’, rather than asserting endonormative language rights. This explanation is also offered by Li (2002), who asks whether Hong Kong people are ‘passive victims of imperialism’ or ‘active agents of pragmatism’. Although Li prefers the latter explanation (his account resembles the thinking of Brutt-Griffler (2002) in terms of the appropriation of English by local agents), things may not be as simple as they appear:

A pragmatic self-pursuit of English seems to be a personal choice on the surface, but may indeed be a self-naturalized uncritical acceptance of linguistic control under the coercive force of state apparatuses (Li 2002: 14).

While such explanations have the disadvantage of lending themselves to ‘false consciousness’ arguments, the subtlety of domination by consent (Gramsci 1971) or symbolic domination (Bourdieu 1991) also needs to be considered. The contention of Luk and Lin (2006: 12) is that while the domination of English in Hong Kong is a result of colonialism and linguistic imperialism, the perpetuation of what they call ‘BANA-centric’ (i.e. British / Australian / North American) linguistic norms ‘seems to be an ideology of local construction’. This is characterised as a ‘clear case of hegemony’ due to the subtle and not-so-subtle means mentioned above, including the use of state apparatuses such as examinations, the NET scheme and the media to construct and maintain ‘accent-based linguistic hierarchization’ (Luk and Lin 2006:

13). An explanatory approach which combines the effects of external ‘linguistic imperialism’ with those of internal ‘linguistic capital’ is also developed by Talbot *et al.* (2003).

Another possible explanation mentioned by Luk and Lin (2006: 14) is what Gandhi (1998) refers to as ‘post-colonial remembering’. There is the possibility that many government officials may be ‘affectively attached to the former colonizer’. The ‘distinctive and healthy Hong Kong identity’ referred to earlier may include at some level a desire for symbols that denote a ‘Hong Kong Chinese’ identity. According to Luk and Lin (2006: 14), one of these symbols may include the ability to ‘speak English with a ‘standard’ prestigious accent from the West’. This may also be one of the ways in which people from Hong Kong seek to construct an identity that differentiates them from their mainland Chinese counterparts.

Luk and Lin make suggestions for reform in three areas: assessment, research and curriculum. Assessment (for example in the LPAT test) needs to question its assumption that only pronunciation free from L1 characteristics is error-free; research needs to focus on the ‘establishment of the common denominator of the World English phonologies’ with a view to establishing a regional, perhaps Asian Pacific, variety of pronunciation; and in curricula, ‘more or less codified regional varieties of English’ should be introduced with the aim of raising awareness of the ‘diversity of acceptable linguistic variants’, including among native speakers (Luk and Lin 2006: 18). The present study will take account of these suggestions, especially with regard to evaluating the acceptable ‘linguistic variants’ that exist in HKE phonology.

3.4.3 HKE accent studies

Studies of attitudes towards accents in Hong Kong seem to support exonormative characterisations of language attitudes. Forde (1995) found that Hong Kong students reacted least favourably to English spoken with a Hong Kong accent, when also provided with samples of American, Australian, British (RP) and British (Yorkshire) accents. The accents were screened by a panel of native speakers to ensure that the

samples were representative of the accents in question. The Hong Kong speaker was also rated lowest on an 'ability in teaching' criterion, which provides an early indication that Hong Kong students have exacting standards. However, the speaker was simply classed as 'middle proficiency' (Forde 1995: 64), and there was no attempt to describe the phonological features of the sample. Similarly, in the study of Candler (2001) students at a Hong Kong secondary school were asked to identify and rate a total of 12 native and non-native accents, including an HKE accent. The rate of recognition was highest for the Hong Kong accent, at 92.7%. The study found that most students wanted to speak with a native speaker accent and considered them 'better'. The HKE accent was favoured by some but generally not seen as an aspect of Hong Kong identity. Although Candler's study suggests that some students may positively evaluate the local accent, the findings are once again limited by the lack of precision about the type of accents used.

The tendency to treat the HKE accent in an undifferentiated way is also visible in Luk (1998), who surveyed secondary students' attitudes towards a local English accent and an RP accent. The representativeness of the local accent was verified by pre-screening, as in Forde (1995), and by the inclusion of a questionnaire item which asked respondents to agree or disagree with the statement 'I think that most Hong Kong people speak like him or her'. As in Candler's study, the rate of recognition was high (between 77.4% and 88.2%). Again, and perhaps unsurprisingly, the great majority (86%) of the respondents wanted their teachers to have RP-like accents. But if the students had been presented with samples of educated, high proficiency Hong Kong English that had been shown to meet international intelligibility criteria, there may well have been a different outcome.

One study that did address variation was that of Bolton and Kwok (1990), which included among the accent samples 'mild-accented' and 'broad-accented' Hong Kong speakers. These samples were presented along with RP and American English accents, also in the differentiated forms of 'mild' and 'broad' (or 'advanced RP' and 'near-RP', in the case of the British speaker). The RP accents were ranked highest on a questionnaire item that asked about suitability for a broadcasting position. However, once again it is not clear how 'mild' and 'broad' were defined, either in

terms of the accent features present or with reference to more global criteria. An interesting finding relating to a ‘choice of model’ question was that male respondents were more likely to choose a ‘Hong Kong bilingual’ option than were female respondents, confirming the pattern found by many accent studies (e.g. Labov 1966; Trudgill 1974).

Given that only a minority of students were able to recognise the ‘mild’ Hong Kong accent in Bolton and Kwok’s study, there would seem to be little awareness of this subvariety. But this is accompanied by a general lack of awareness of accent differences (Bolton and Kwok 1990: 170-171). This is also remarked upon by Luk (1998: 104), who believes students should have ‘more awareness of the existence of different accents in the world’. On the ‘choice of model’ issue, Luk (1998: 103) concludes that ‘the easy availability of an idealised exonormative model of English, coupled with the high status attached to it has made the development of an institutionalised endonormative model unnecessary’. However, there is a need for a reassessment of this question, following a research procedure that acknowledges the variation in HKE phonology. The following sections will consider the nature of this variation in more detail.

3.5 The phonology of Hong Kong English

This section first compares the descriptions of the phonology of Hong Kong English (HKE) presented by Hung (2000) and Deterding *et al.* (2008). This enables a fairly consistent and sufficiently detailed picture of HKE phonology to emerge. Some of the consonantal features will then be selected for inclusion in the preliminary study of HKE phonological features.

3.5.1 Two previous studies compared

For Hung (2000), the concept of systematicity is central, thus challenging the LPAT descriptor’s depiction of ‘systematic errors’ with the idea of ‘systematic features’ (Hung 2000: 354). However, the concept of systematicity needs careful

consideration. Hung's study involved recording 15 undergraduate students in Hong Kong, and his study concentrated on features that were 'not idiosyncratic but shared by a number of HKE speakers' (Hung 2000: 339). The focus of Hung's study was on the phonological properties of HKE rather than their frequency of occurrence, but this means that the group of speakers may not have been representative of wider patterns of language use. This is acknowledged by Hung (2000: 339), who recognises that there is a large amount of variation in HKE data. Nevertheless, for this group of speakers, there were a number of features in common. Among these were:

- Vowels: 'Hong Kong speakers in general operate with as few as 7 simple vowel contrasts...one important systematic feature is the lack of the tense/non-tense or long-short distinction, which more than anything else accounts for the smaller number of vowel contrasts in HKE' (Hung 2000: 343). A vowel inventory of 15 vowels (seven monophthongs and eight diphthongs) is postulated, compared with the 19 vowels of RP. The mergers of /ɪ, i:/, /e, æ/, /u:, ʊ/ and /ɔ:, ɒ/ in HKE account for the difference.
- Fricative consonants: 'for the great majority of speakers, there is no evidence of a voiced/voiceless contrast' (Hung 2000: 347). There are thus only four such consonants in the proposed HKE inventory (/f, s, θ, ʃ/); the voiced equivalents did not appear in the data.
- In onset position, [l] and [n] are often interchanged by HKE speakers (Hung 2000: 351). It appears that if the syllable contains a nasal, interchange is more likely (for example, *line* produced with initial [n], and *number* produced with initial [l]). The conclusion, however, is that [l] and [n] are in free variation in the onset of a syllable (Hung 2000: 352).
- Consonant clusters: the [kw] cluster shows deletion of the [w] component before rounded vowels, explaining the pronunciation of *quarter* as [kɔtə].
- /l/ in coda position (i.e. postvocalic /l/): the phonological rules operating on the production of /l/ in coda position can be summarised as follows:

/l/ > [w] / [-back vowel] ____ (e.g., *feel* is realised as [fiw])

/l/ > Ø / [+back vowel] ____ (e.g., *cool* is realised as [ku])

Hung's study certainly constitutes 'valuable work on the description of certain features', as noted by Kirkpatrick (2007b: 385). However, and while it is not intended to have pedagogical applications, most of the above features appear to fail the test of international intelligibility. If the findings of Jenkins (2000) and Kirkpatrick (2006) regarding international intelligibility are accepted for the time being, then all but one of these features would be potentially problematic. Vowel length distinctions are part of the LFC core, but they are not made in this version of HKE. Intelligibility-threatening consonant substitutions would inevitably arise from the lack of a voiced/voiceless fricative contrast, and also from the interchangeability of [l] and [n]. The above pattern of initial consonant cluster simplification also violates the LFC's core requirements. Of these features, only the substitutions of postvocalic /l/ and of the dental fricatives appear to be unproblematic from the perspective of international intelligibility (although the voiceless dental fricative /θ/ did appear in the data, along with substitutions). Of course, intelligibility was not the focus of the study, but in addition this subvariety of HKE may not be acceptable to all local users.

The approach of Deterding *et al.* (2008) uses interview data, rather than word lists, and thus provides a guide to the ways words are pronounced in connected speech. The 15 subjects were teacher trainees, who might be expected to be more 'norm-focused' than the first-year arts and science undergraduates in Hung's study. In terms of the vowel system, Deterding *et al.* confirm Hung's general finding about the reduced vowel system of HKE. For the monophthongs, the tendency to merge the /æ, e/ and /ɪ, i:/ pairs was noteworthy and confirms Hung's vowel chart data. The separation of /ʌ/ and /ɑ:/ (or /ʌ/ and /ɑ/ in Hung's postulated HKE phonemic inventory, where he sees a systematic absence of long/short vowel distinctions) is also indicated by both studies. Hung believes this may be due to the existence of a similar pair in Cantonese, namely [sʌm] (*heart*) and [sam] (*three*). One difference is that LOT and THOUGHT (/ɒ/ and /ɔ:/, in RP) were seen as merged by Hung, but Deterding *et al.* (2008: 162) conclude that 'it is not clear if they are fully merged'. For the diphthongs, Hung's observation that HKE has 8 diphthongs and thus 'differs from many NVEs, such as Singaporean or Indian English, which have a simpler inventory of true diphthongs' (Hung 2000: 346) is confirmed by the later study,

which found no evidence of the monophthongisation of the FACE and GOAT (/eɪ/ and /əʊ/) diphthongs.

Turning to the consonants, Deterding *et al.* include measurements of the frequency of dental fricative substitution. The results for words containing possible contexts for the voiceless dental fricative /θ/ or voiceless TH are reproduced in Table 3.2 below.

Table 3.2. The percentage of word tokens containing variants of /θ/ (adapted from Deterding *et al.* 2008: 154). Ø = zero consonant.

Phonological features	No. of tokens	% of tokens
• Voiceless TH in initial position:		
[θ]	27	64
[f]	14	33
[t]	1	2
• Voiceless TH in medial position:		
[θ]	6	60
[f]	0	0
[t]	2	20
Ø	2	20
• Voiceless TH in final position:		
[θ]	2	22
[f]	6	67
[t]	0	0
Ø	1	11

While the data show that substitution is more likely in final position, all four tokens of the word (*youth*) came from the same speaker, as noted by Deterding *et al.* (2008: 154). In general, the small number of tokens makes generalisation difficult and is a limitation of the study. This is probably a result of the use of unscripted speech data (some sounds are simply less frequent, including the /θ/ sound). There is less detail about the frequency of voiced dental fricative substitution, but the most common substitution is thought to be [d], as in Hung's data. The other voiced fricatives (found to be absent from HKE in Hung 2000) were not investigated in Deterding *et al.*'s study. Substitution or deletion of postvocalic /l/ was an observed feature in both studies, with both vocalisation and deletion appearing. The conflation of [n] and [l], while occurring in up to 37% of word tokens in Hung's study, was on the contrary

found to be ‘rare’ by Deterding *et al.*, with only two examples (in one of these, *Canada* sounded rather like *calendar*).

Consonant cluster patterns are considered in some detail by Deterding *et al.* In the case of word-final clusters before a pause or before a word beginning with a vowel (for example, *most of*, in which /t/ would not normally be subject to elision in native speaker varieties), the subjects elided the final consonants /t, d, k/ in just over half (53.6%) of the 69 tokens of potential word-final plosives in the data. In initial clusters, /kr/ combinations were often realised as [kl] (e.g. *crowded* [klaʊdɪd]), while /pr/ clusters were less likely to be changed (although informal data suggest to Deterding *et al.* that the word *problem* is often realised with a [pl] initial cluster; see also Chan and Li 2000: 82). Consonant deletion also occurred, mainly after bilabial consonants (e.g. in *applied* [ʌpaɪd]).

Deterding *et al.*’s study also looked at suprasegmental features including rhythm and sentence stress. While the expected syllable-timed rhythm occurred, the study’s authors believe that more research is needed to ascertain whether this is systematic. Some of the sentence stress and intonation patterns confirmed the findings of other NVE studies. There was an absence of de-accenting, so that repeated or predictable information was given emphasis; Low (1998) found this to be the case in Singapore English. It is uncertain whether this leads to problems in locating nuclear or tonic stress, which might have consequences for comprehensibility or interpretability. Pronouns and determiners were also stressed unexpectedly, as found by Deterding and Kirkpatrick (2006) in their study of ASEAN interaction.

A summary of the similarities and differences between the two studies is shown in Table 3.3 below.

Table 3.3. Similarities and differences in two descriptions of HKE phonology (adapted from Sewell 2009: 39).

HKE phonological features	Hung (2000)	Deterding <i>et al.</i> (2008)
a) Vowels		
• merger of vowel contrasts, especially /æ, e/ and /ɪ, i/	Yes (seven monophthongs)	Yes
• Fronted [ʌ]	Not reported	Yes
• FACE/GOAT as diphthongs	Yes	Yes
b) Consonants		
• Voiceless TH in initial position	[θ] for ‘at least half’	[θ] for 64% of tokens (33% [f])
• Voiceless TH in medial position		[θ] for 60% of tokens (20% [t])
• Voiceless TH in final position		[θ] for 22% of tokens (67% [f])
• Voiced TH	Does not exist in HKE	Not reported in detail
• Conflation of [n] and [ɺ]	Yes (up to 37% of tokens)	Rare
• L vocalisation or deletion	Yes	Yes
c) Consonant clusters		
• Initial	Not studied	/kr/ clusters often realised as [kl] e.g. <i>crowded</i> [klaʊdɪd]; /pr/ clusters less so (but <i>problem</i> may have [pl])
• Final	Not studied	Deletion can occur, mainly after bilabial consonants
d) Suprasegmental features		
• Rhythm	Not studied	Syllable-timed rhythm occurs but more data needed
• Sentence stress	Not studied	Absence of de-accenting; stressing of pronouns and determiners

3.5.2 The variation problem

While the above studies of HKE reveal many common patterns, there is also a great deal of variation, as noted by Hung (2000). In terms of inter-speaker variation, comparing the two studies reveals what may be part of a developmental path for many HKE speakers. In Hung’s study, the vowel system shows more signs of L1 influence and has fewer contrasts. The /ɒ/ and /ɔ:/ vowels are regarded as merged,

making COT and CAUGHT homophonous; this could also conceivably result from the influence of American English pronunciation. The conflation of [n] and [l] is more common in Hung's data. It could be hypothesised that the speakers in Deterding *et al.* had a higher level of proficiency, and as trainee teachers were probably more aware of the need for 'accurate' pronunciation. This may be said to show 'de-merging', but on the other hand the data collection methods could also account for these differences.

A possible dilemma that arises for linguistic descriptions of NVEs is thus: if speakers with a low level of proficiency are used, there are likely to be intelligibility problems from an international perspective, but if more advanced speakers are used there will be fewer and fewer differences from native speaker varieties, weakening the persuasiveness of the case for description and varietal status. The associated, crucial question is how much of this approximation is part of the normal process of learning a language, and how much is merely the uncritical or 'forced' adoption of native speaker norms. The concept of the 'interlanguage myth' is invoked by Kachru (2005) to criticise the tendency to see deviations from such norms as deficits (see also Jenkins 2006, and Brutt-Griffler 2002). Hung (2000: 354), who aims to conceptualise HKE 'on its own terms', nevertheless invokes the idea of an interlanguage continuum, with an 'idealised' HKE phonology at one end and a 'standard British or American phonology' (Hung 2000: 339) at the other.

An alternative perspective on variation is taken by Bolton and Kwok (1990), who make use of an 'acts of identity' approach in which many Hong Kong English speakers are seen as modelling their speech forms on those of educated bilinguals in Hong Kong. The uneven distribution of certain forms may be due to constraints such as the non-availability of NS norms 'for the vast majority of language learners' (Bolton and Kwok 1990: 164). As Bolton and Kwok note (1990: 161), in Le Page and Tabouret-Keller's (1985) formulation one of the constraints on performing acts of identity is the extent to which one is able to change one's behaviour. Thus a continuum is still likely to exist, although there is less suggestion of a 'target' than in an interlanguage phonology approach.

Although based on limited data, the indications of intelligibility studies are reasonably clear, however, and for certain types of research an important aim – not only in Hong Kong, but in any English-using country or region – is to try and determine the characteristics of a ‘local accent’ that do not compromise international intelligibility, while also being acceptable to local listeners. In Hong Kong, the local scholars Luk and Lin (2006: 17) agree on the need to distinguish between ‘local accents and careless speech’, and perceptions of ‘carelessness’ may affect both intelligibility and acceptability.

Of course, this presupposes that international intelligibility is desirable. Kachru’s concept of a ‘cline of intelligibility’ (Kachru 1992b: 65) relates to the intra-speaker variation or style shifting that may occur in some varieties. The speaker’s position on this cline is seen as depending on the context and the participants; intelligibility is thus context specific, to some extent. However, as noted above, while proficient speakers will be able to move in both directions along the cline, less proficient ones will only be able to shift ‘downwards’, unless they are skilled mimics. The ability to shift between ‘maximal intelligibility’ and ‘maximal solidarity’ arguably forms part of a contemporary definition of proficiency, as is argued by Canagarajah (2005). So while intelligibility may not always be a priority for individual speakers, it may still be a useful consideration when dealing with the ‘variation problem’. In Hong Kong, where there are limited opportunities for intra-ethnic communication in English, international intelligibility is arguably an important criterion. If linguistic descriptions are intended to have codifying or pedagogical applications, ignoring it will reduce their credibility.

For evaluative studies, there are two aspects of the variation problem. Inter-speaker variation means that the choice of subjects, in terms of factors such as age, educational level and proficiency, is important. While certain features may be common among some speakers, they may not be evenly distributed across the whole speech community. Another aspect of the variation problem results from intra-speaker, as opposed to inter-speaker, variation. While it is clear that some features occur in the speech of many Hong Kong English users, they may occur variably according to factors such as phonological context and the perceived formality of the

situation (and any associated style shifting). Their occurrence may also depend on the location of the speaker's utterances on a hypo- to hyper-articulated continuum reflecting the relative, situation-specific 'carefulness' of speech. One implication of intra-speaker variation is that if some features occur according to variable rules, users may have more than one variant at their disposal. This must be taken into account when making pedagogical recommendations, as the artificial selection of one variant over another would arguably reduce the flexibility of speakers along a number of dimensions.

Of course, many of the same arguments about variation could also be made regarding descriptions of any variety, native or non-native, standard or non-standard. However, in the case of L2 varieties there appears to be an additional source of variation, namely imperfect learning and its effects on proficiency. In Hong Kong, for example, some English users may have had limited exposure and experience, whether in formal or informal contexts, and may almost never use the language in their daily lives. On the other hand, some users have had extensive experience with the language and use it in a range of professional and social contexts. It seems to make little sense to base a description of the local variety entirely upon those who use the language only infrequently, or are still studying it. This has been the case to some extent with most studies of HKE phonology to date, which have generally used university students as the source of data (e.g. Hung 2000; Stibbard 2004; Deterding *et al.* 2008). The present study will therefore conduct a preliminary investigation of phonological feature use within a sample of proficient bilingual speakers in Hong Kong. This part of the study acknowledges and extends the approach of Bolton and Kwok (1990), who use the term 'higher-range' to refer to this subset of HKE speakers.

3.6 The preliminary study

In order to tackle the variation problem and gain an overview of the actual occurrence and distribution of HKE features within a sample of speakers, a preliminary study of HKE phonology was conducted. For both the preliminary study

and the main study, the samples were obtained from a ‘mini-corpus’ derived from local television programmes. Some of the advantages and disadvantages of this approach from a general methodological perspective are examined in more detail in Chapter 4, where the representativeness of the samples will be considered. This chapter merely provides an overview of the mini-corpus, in order to focus on the findings of the preliminary study. This part of the study is also reported in Sewell and Chan (2010).

3.6.1 The HKE mini-corpus

A database or mini-corpus of spoken HKE was created both for the purposes of the preliminary study and to provide samples for the main study. The mini-corpus consisted of 48 extracts from current affairs programmes on local television, representing 25 different speakers. All of the speakers were presumed, on the basis of accent, to have Cantonese as their first language. The total lengths of the samples for each speaker ranged between 16 seconds and 229 seconds, with an average duration of 74 seconds per speaker. The whole corpus contained just over 30 minutes of spoken Hong Kong English. Appendix 1 provides more details of the extracts and the speakers in the mini-corpus.

One of the main reasons for using ‘media English’ was to focus on speakers with higher proficiency levels, as mentioned above. Media English has the advantage of being preselected for proficiency to some extent, as presumably the speakers would not have agreed to take part in the programmes (or been allowed to) if their proficiency level had been inadequate for the task. Furthermore, a minimum level of intelligibility and comprehensibility is usually maintained via the recording and editing processes, as well as by the communicative contexts of the programmes. The disadvantages of using media English include the fact that the samples may not be representative of the way most people actually use English. The range of speakers was somewhat limited in terms of gender, occupation and age. All but four of them were male, probably reflecting actual gender imbalances in their occupations. Around half were involved in politics, although in Hong Kong this often means they are also representatives of occupational and professional groups, under the system of

functional constituencies. The other speakers included journalists, civil servants and spokespeople for professional, commercial and social organisations. The focus on experts and senior figures meant that most speakers were aged over 50, and certainly none appeared to be under 30. Again, further details of the speakers in the mini-corpus are provided in Appendix 1.

The main source of recordings was the current affairs programme *The Pulse*, with *The Pearl Report* also being used. Permission was sought and obtained from the broadcasters to use recordings for research purposes. Library DVD recordings of the programmes were obtained from the library of Lingnan University in Hong Kong and were viewed in order to identify suitable accent samples. Commercially-available audio-ripping software (Eufony™) was used to extract the recordings directly from the DVD soundtrack, in order to maintain sound quality. The samples were saved as CD-quality WAV files (16-bit stereo, 1611 kbps) to enable clear playback.

The contexts or genres of speaking in the programmes were mainly location interviews (about 50% of the samples) and studio discussions (about 30%), with studio interviews and speeches or other forms of public address making up the remainder. Most of the samples therefore represented unrehearsed speech, although the extent to which location interviews involve ‘soundbites’, which are usually rehearsed, is uncertain. The range of topics covered in the programmes was quite limited, with political and social issues forming the thread of most programmes.

3.6.2 Consonantal features of HKE

The preliminary study focused on HKE consonantal features, for two main reasons. Firstly, impressionistic listening during the collection of the accent samples suggested that while there were of course a range of distinctive vowel realisations, instances of outright conflation or merger were rare, while consonantal substitutions were present in a majority of samples. Secondly, the initial assessment of the intelligibility characteristics of features carried out in Chapter 2 indicates that while there are several HKE consonantal features that appear to be inconsequential for

intelligibility (for example, dental fricative substitution, L vocalisation and certain types of final cluster simplification), there are few comparable vowel features (with the possible exception of monophthongal forms of the FACE and GOAT diphthongs). Consonantal features have appeared as candidates for acceptance or codification in some other NVEs; for example, He and Li (2009) propose that dental fricative substitutions should be accepted in ‘China English’. A list of the HKE consonantal features considered in the preliminary study is presented in Table 3.4 below.

Table 3.4. HKE consonantal features considered in the preliminary study (adapted from Sewell and Chan 2010: 144).

Feature name	Explanation	Examples from the mini-corpus
TH stopping	Substitution of the voiced dental fricative /ð/ with [d] (Deterding <i>et al.</i> 2008, Hung 2000)	Most frequent word-initially, e.g. <i>that</i> [dæt]
TH fronting	Substitution of the voiceless dental fricative /θ/ with [f] (Deterding <i>et al.</i> 2008, Hung 2000)	Word-initially: <i>three</i> [fɹi:] Word-medially: <i>forthcoming</i> [fɔːfkʌmɪŋ] Word-finally: [bouf]
L vocalisation	The use of a vowel (vocalisation) in place of /l/ in postvocalic position, e.g. <i>will</i> as [wɪu] and <i>oral</i> as [ɔrou] (Bolton and Kwok 1990: 153, in Deterding <i>et al.</i> 2008: 161). Deletion may also occur, e.g. <i>cool</i> may be [ku] (Hung 2000: 350)	Preconsonantly: <i>chilled</i> [tʃɪud] Word-finally: <i>people</i> [pi:pou]
[n, l] conflation	[n] and [l] are in free variation in onset position (Hung 2000: 352). Word pairs such as <i>night</i> and <i>light</i> become homophones	<i>number</i> [lʌmba]
/r/ substitution	/r/ is produced as [w] in onset position (Chan and Li 2000: 80)	Word-initially: <i>rely</i> [wɪlaɪ]
/v/ substitution	/v/ is produced as [w] in onset position (Hung 2000: 348-349; Chan and Li 2000: 79)	Word-initially: <i>very</i> [wɛɪɪ] Word-medially: <i>advantage</i> [ɛwɑːntɪdʒ]
Initial consonant cluster modification (‘initial CCM’)	Initial consonant clusters are reduced, especially those involving /r, l/ after plosives (e.g. <i>produce</i> as [pədjuːs]; Chan and Li 2000: 82). Substitution may also occur, e.g. <i>crowded</i> as [klaudid] (Deterding <i>et al.</i> 2008: 159)	Word-initially: <i>providing</i> [pouvaɪdɪŋ]

While most of these features have already been discussed in terms of various criteria in Chapter 2, two require further explanation as they are features that appear to be specifically related to the L1. The phenomenon of [n, l] conflation is related to ongoing language change in Cantonese. Tong and James (1994: 6) note that the use of [l] in place of initial [n] is widespread among young speakers (cited in Deterding *et al.* 2008: 160), although free variation seems to be the pattern with English (Hung 2000: 352); *line* is sometimes heard as *nine*, causing occasional confusion when students refer to line numbers. While sometimes thought to be unique to HKE, this conflation also occurs in the English of speakers from central China (Deterding *et al.* 2008: 160).

The use of [w] as a substitution for /r/ and /v/ is fairly widespread in Hong Kong. Chan and Li (2000) note the former substitution and Hung (2000) the latter; Deterding *et al.* (2008) do not include these substitutions in their study of HKE pronunciation. In the case of /v/ substitution, Hung (2000: 349) notes the ‘[w]-like quality’ of the medial consonant of words like *revoke*, and also includes examples of substitution in initial position. Word-medial substitutions of /v/ can be either [w] or [f], with the latter usually occurring in unstressed syllables, such as in *even* [‘ifən]. However, [w] can also occur in unstressed syllables; Hung (2000: 350) gives the example of *advertise* [‘ədwətəɪs]. In this case, the use of [w] instead of [f] is probably conditioned by the preceding voiced consonant. Hung’s preferred hypothesis is that there is no /v/ phoneme in HKE, and that only /w/ and /f/ exist. This needs to be verified by further investigation of speakers at different levels, and in this study ‘/v/ substitution’ refers to substitution by [w], whatever the phonological context. It should be noted that the possibility of intermediate realisations, such as a voiced bilabial fricative [β], also exists, although the present study is not generally concerned with this level of phonetic detail. Instances of substitution by [f] (such as in *even*, pronounced as [‘ifən]) are classified as the devoicing of ‘voiced’ fricatives (see below).

Turning to /r/ substitution, this can also occur in either stressed syllables, such as in *red* [wəd], or unstressed syllables, such as in *rely* [wi‘laɪ]. It is likely to be related to the absence of a postalveolar approximant in Cantonese, resulting in the substitution

of the labiovelar glide [w]. It also occurs more generally in child L1 English acquisition, and it has been observed that many children operate a process of ‘gliding’ of liquids in which /r/ is produced as [w] (Johnson and Britain 2007: 303). In general, the fact that both /v/ and /r/ substitution are rare in other varieties of English suggests that they are more probably related to transfer; the relatively early acquisition of these sounds in L1 acquisition (compared to the dental fricatives, for example) also supports this view. As /r/ substitution is not mentioned as frequently in the literature, it may be the case that it is less widely distributed or more variable in HKE than /v/ substitution.

Table 3.4 therefore covers most of the consonantal features reported in such studies as Hung (2000), Deterding *et al.* (2008) and Chan and Li (2000), but there are of course other consonantal features in HKE. Munro and Derwing (2006) note the conflation of [s] and [ʃ] in initial position (see also Bolton and Kwok 1990: 153) and the replacement of /d/ with [z] in intervocalic position, but these were not noticeable in the mini-corpus. One other feature that did occur in the data, however, is that of the devoicing of voiced fricatives (or more accurately, the non-voicing of /v, z, ʒ/ and their apparent merger with /f, s, ʃ/. Hung (2000) goes to the extent of saying that the voiced fricative sounds do not exist in HKE, while Deterding *et al.* (2008) do not report them in detail. Bolton and Kwok (1990) include the devoicing of voiced consonants, including /b, d, g/, in their list of HKE features.

Cases of devoicing were noticed during the initial collection of accent samples, as in the *even* ([ˈiʃən]) example. However, because it can apply to many sounds and occur in many phonological contexts it creates a potentially diverse category, and may have little diagnostic value. The tendency towards final obstruent devoicing is widespread across many varieties and languages, and is even observable in native speakers; Shockey (2003: 30) concludes on the basis of impressionistic observations that ‘speakers of English avoid voicing in obstruents where possible’. It was therefore decided not to include this feature in the preliminary study, but it is acknowledged that the devoicing of voiced consonants is a common feature of HKE. The design of the main study will allow students to mark this and any other features thought to be relevant.

Similarly, final consonant cluster reduction (final CCR) also shows universal occurrence, and is another common HKE feature that has not been included. Final CCR was very prevalent in the data, but as final CCR occurs in all varieties it would be necessary to distinguish between types of final CCR that are specific to HKE and those that occur more generally. The occurrence of final CCR also depends on phonological context, and this seemed to add a level of analytical complexity that was unwarranted in a preliminary study. The feature list thus consists of segmental, consonantal features that occur in clearly-defined phonological contexts (although word-final L vocalisation, like final CCR, is also more likely when the following word begins with a consonant). Finally, a feature that was included was the modification of initial clusters (henceforward, ‘initial CCM’). Chan and Li (2000) describe the deletion of /r/ in the initial cluster of *produce*, and Deterding *et al.* (2008) include an example of substitution (*crowded* as [klaʊdɪd]). In the present study, initial CCM includes both deletion and substitution. It was regarded as being of interest partly because Jenkins (2000) found it to be one of the causes of intelligibility problems in her data.

3.6.3 Data analysis

To increase the reliability of the data and reduce the chance of filtering the speakers’ productions through the perceptual characteristics of my L1 English background, the process of data analysis involved the use of a research assistant (a bilingual speaker of Cantonese and English with some training in phonetic transcription). The author and the research assistant listened independently to the 48 extracts in order to determine whether or not each speaker used the features under consideration. Lists of word tokens were then compared, and instances of disagreement were subjected to repeated listening. Decisions were changed on some occasions, while avoiding any tendency towards convergence. After repeated listening the level of inter-rater agreement ranged between 97.9% of tokens, for instances of [n, l] conflation, and 75.6% of tokens for L vocalisation; this probably reflects differences in degrees of phonetic similarity and hence the difficulty of deciding which variant occurred. There was thus a generally high level of agreement between the two raters, and no attempts at acoustic analysis were made.

It should be noted that not all of the word tokens in the data were analysed, in some cases. If the use of a feature by a speaker could be established through the presence of at least one token, then not all of the tokens were analysed. Feature non-use, on the other hand, was established by analysing all the possible word tokens in order to make sure that no instances of feature use occurred. Although the principle of accountability (Labov 1973: 72) is normally applied to variationist data collection, so that all possible tokens are analysed, this part of the study was not concerned with establishing percentage rates of occurrence. Rather, the objective was to provide an overview of the frequency of occurrence and possible distributional patterns of the selected features; the nature and extent of intra-speaker variation was therefore not addressed in detail. Furthermore, implicational scaling (the analytical procedure used in this part of the study) normally involves binary, either/or categorisation rather than percentages. Accordingly, when combining the results of our analyses, if both raters agreed that a speaker used a feature in at least one context he or she was coded ‘Y’ overall for that feature. If no tokens were found by either rater, or if the raters could not agree on the presence of a feature within a word token, the relevant speaker was coded ‘N’ overall. A summary of the numbers of word tokens analysed, the frequency of occurrence across speakers, and the rates of inter-rater agreement is given in Table 3.5 below. If there were no possible contexts for the use of a feature, these cases were excluded when calculating the percentage of speakers using each feature. Thus in Table 3.5, the percentage of speakers using L vocalisation is calculated on the basis of there being 25 possible contexts (at least one for each speaker), while the percentage of speakers using TH fronting is calculated on the basis of there being 22 possible contexts (i.e., three speakers had no contexts for this feature in their utterances).

Table 3.5. Number of word tokens and average number of tokens per speaker, distribution of features across speakers, and rate of agreement for each feature category (table also in Sewell and Chan 2010: 147).

	TH stopping	TH fronting	L vocal.	[n,l] conf.	/r/ subst.	/v/ subst.	Initial CCM
Number of word tokens analysed	141	73	119	140	63	68	145
Average number of tokens per speaker	5.7	2.9	4.8	5.6	2.5	2.8	5.8
% (number) of speakers using the feature in at least one context	76.0 (19)	27.3 (6)	80.0 (20)	8.0 (2)	12.0 (3)	19.0 (4)	32.0 (8)
% agreement between raters	81.6	90.4	75.6	97.9	96.8	95.6	97.2

The initial analysis revealed that some features were far more prevalent than others. Phonemic substitutions such as those involving /r/ and /v/, and [n, l] conflation, were the least frequently occurring features, in terms of the percentage of speakers who displayed the features in at least one context (Figure 3.1 below shows these percentages). The two most frequently occurring features were L vocalisation and TH stopping.

Figure 3.1. The distribution of Hong Kong English consonantal features according to the percentage of speakers using them in at least one context (figure also in Sewell and Chan 2010: 147).

3.6.4 Implicational scaling

In the initial analysis of the accent samples there appeared to be implicational patterns; for example, those speakers who showed /v/ or /r/ substitution were also likely to show other features. An implicational scale was therefore a natural development of the analysis, rather than being a prior intention of the study. As there were two raters, it was first of all necessary to combine the results of our analyses. As mentioned above, if a speaker received a ‘Y’ code from both raters for at least one word token, he or she was given an overall ‘Y’ code for that speaker/feature combination. Following the procedure outlined by Rickford (2002), the resultant ‘Y’ and ‘N’ codes were then entered into a table with the 25 speakers as row labels and the seven features as column headings. Again, following Rickford (2002), the columns of the table were reordered horizontally according to the total number of ‘Y’ codes in each column, with the most frequently occurring features being placed on the left hand side of the table. The rows were reordered vertically according to the number of ‘Y’ cells in each row, so that the uppermost speakers were those with the

largest number of Hong Kong English features. An ‘n/a’ code was entered where there were no contexts for the feature within the speaker’s utterances. Table 3.6 below shows the resultant implicational scale.

Table 3.6. The implicational scale of Hong Kong English consonantal features, with speakers ordered vertically according to the number of “Y” cells within each row, and features ordered horizontally according to the number of “Y” cells within each column. “Deviant” or unexpected “Y” cells are circled. An “n/a” code represents the absence of word tokens within a speaker’s utterances. (Table also in Sewell and Chan 2010: 149).

		CONSONANTAL FEATURES							<i>Number of Y cells</i>
		L vocal.	TH stop.	Initial CCM	TH front.	/v/ subst.	/r/ subst.	[n, l] conf.	
SPEAKER NUMBER	19	Y	Y	Y	Y	Y	Y	N	6
	25	Y	Y	Y	Y	Y	N	Y	6
	22	Y	Y	Y	Y	Y	N	N	5
	21	Y	Y	Y	N	Y	N	N	4
	1	Y	Y	N	Y	N	Y	N	4
	3	Y	Y	N	N	N	Y	Y	4
	15	Y	Y	Y	N	N	N	N	3
	2	Y	Y	Y	N	n/a	N	N	3
	17	Y	Y	Y	N	n/a	N	N	3
	18	Y	Y	Y	n/a	N	N	N	3
	24	Y	Y	N	Y	N	N	N	3
	10	Y	Y	N	N	N	N	N	2
	11	Y	Y	N	N	N	N	N	2
	12	Y	Y	N	N	N	N	N	2
	13	Y	Y	N	N	N	N	N	2
	8	Y	Y	N	N	n/a	N	N	2
	23	Y	Y	N	N	n/a	N	N	2
	5	Y	N	N	Y	N	N	N	2
	14	Y	N	N	N	N	N	N	1
	16	Y	N	N	n/a	N	N	N	1
	9	N	Y	N	n/a	N	N	N	1
	20	N	Y	N	N	N	N	N	1
	4	N	N	N	N	N	N	N	0
	6	N	N	N	N	N	N	N	0
	7	N	N	N	N	N	N	N	0
<i>Number of Y cells</i>		20	19	8	6	4	3	2	

The underlying principle of implicational scales is that they depict ‘hierarchical co-occurrence patterns in the acquisition or use of linguistic variables by individuals or

groups, such that x implies y but not the reverse' (Rickford 2002: 143). 17 of the 25 speakers follow just such an implicational pattern, in which (reading along each row from left to right), an 'N' symbol is followed only by more 'N' symbols. In other words, the absence of a consonantal feature in a speaker's row implies the absence of those features to its right, while the presence of a feature implies the presence of those features to its left. In Table 3.6, for example, Speaker 10 follows an implicational pattern in that initial CCM does not occur, and the other features to the right are also absent. Speaker 19 also follows an implicational pattern as the presence of /r/ substitution implies that other features to the left will be used, and this is in fact the case: /v/ substitution, TH fronting, initial CCM, TH stopping and L vocalisation all occurred with this speaker. In terms of the codings, when read from left to right 'YY', 'YN' or 'NN' inter-cell transitions are seen as fitting the implicational pattern, while 'NY' transitions are seen as 'deviant' from the point of view of implicational scaling. An exception is provided by Speaker 3's row, in which both of the 'Y' cells following three 'N' cells were classed as deviant, despite forming a 'YY' sequence. In the above table, ten deviant 'Y' cells are visible (these are circled). An example is speaker 25, who does not show /r/ substitution and would not therefore be expected to show the conflation of [n] and [l]; however, this feature does appear.

Guttman (1944), the originator of implicational scaling, proposed a measurement of the 'index of reproducibility' (IR) to assess the scalability of such data, or in other words the extent to which the implicational pattern repeats itself. The formula is:

$$IR = 1 - \text{Number of "errors" (deviant cells)} / \text{Number of data cells}$$

In this case there are 175 data cells and ten deviant cells. If the seven 'n/a' or 'empty' cells are ignored the calculation becomes $1 - (10 / 168)$, giving an IR of 0.94. A stricter measurement would take account of the fact that the left-hand column cannot have an implicational relationship because it is not preceded by anything, and on this basis (150 data cells) the IR is slightly lower at 0.93. While the point at which the IR becomes statistically significant is uncertain, Rickford (2002: 157), citing Dunn-Rankin (1983) states that an IR of 0.93 'approximates the .05 level of significance'. Rickford also advises against having too many empty cells, noting that the proportion

of these in implicational scales has varied greatly, from 3.125 percent (Pienemann 1998) to 28.2 percent (Bickerton 1973). In Table 3.6, the proportion of empty cells is 4 percent.

It should also be pointed out that there are reasons to approach implicational scaling with caution. Fasold (1990: 199) notes that '[t]here is considerable freedom for manipulating the data in implicational scales'. Columns are usually ordered in such a way as to create as perfect a scale as possible, not in order to 'manipulate' the data but because this ordering is thought to represent the hierarchical relationship between features. The applications of implicational scaling are considered by LePage and Tabouret-Keller (1985), who criticise its neglect of social identities (cited in Fasold 1990: 197). In particular, the use of such scales may imply that there is a single focus, on the acrolectal or standard 'target'. It is important to remember that some of the variation in accent samples may not be purely linguistic, and that it may also indicate deliberate style shifting by the speakers concerned. Nevertheless, the implicational patterns may indicate how such style shifting may occur, in terms of the allowable combinations (as suggested by Bell 1984).

Implicational scaling has been used more widely in sociolinguistics (e.g. DeCamp 1971; Bickerton 1973), than in adjacent research paradigms, although Ho and Platt (1993) use it in their study of copula deletion in Singapore English. Williams (1987) notes that discussion of acquisitional processes has generally been avoided in research into NVEs, because it implies false notions of 'target' and may lead to a deficit perspective. According to Williams (1987: 164), '[a]lleged deficiencies in these varieties are potentially interpretable as deficiencies in their speakers'. However, as long as one guards against the deficit perspective and the assumption that all speakers move or wish to move along the scale, there may be a role for implicational scaling in the analysis of new varieties of English, as well as in descriptions of L1 varieties; Altendorf (2003: 119) presents 'implicational hierarchies' of features within the south-eastern accent continuum in the UK. The existence of implicational patterns in language universals is well known, and Eckman (2008: 97) notes that if a language has a voice contrast in syllable coda position it will also have such a contrast in syllable onset position, but not vice versa.

A problematic issue in this type of exercise is deciding on the categories, the selection and definition of which will significantly affect the outcome. In the present study, phonemic substitutions, such as those involving /r/ and /v/, could perhaps be grouped together as they appear to be transfer-related. Conversely, other features, such as TH fronting, may benefit from separation according to the contexts in which they occur. It is probably best to see the implicational scale as a general overview of feature patterning, and further studies could then subject features to analysis with programs such as VARBRUL. But despite the limitations of the approach, the tables and figures above give a useful overview of the frequency of occurrence and distributional patterns of some consonantal features in Hong Kong English. Perhaps the least that can be said is that ‘the scope of variability is significantly constrained’ (Rickford 2002: 143). An important observation with relevance for this study is the existence of considerable inter-speaker variation. While some speakers displayed almost all of the features within their utterances, others used none. It is therefore difficult to conclude, as did Hung (2000: 337), that the speakers share a ‘common underlying phonological system’. Rather, it confirms and elaborates the conclusion of Bolton and Kwok (1990: 167), who found a ‘clustering’ of items seen as being tokens of the Hong Kong accent, along with standard forms. Within this study, the above analysis appears to have achieved its objective of providing an overview of the distribution and co-occurrence patterns of some HKE consonantal features.

3.6.5 Explaining the implicational patterns

The question remains of what the implicational patterns may actually represent. The findings of the implicational scale will be integrated into the explanatory model presented in later chapters of this study, as it appears to encapsulate the possible interrelationships between language variation and factors such as intelligibility and L2 phonology acquisition. For the time being, a brief outline of these interrelationships will suffice. In terms of intelligibility, the three features on the right-hand side of the scale, namely substitutions of /v/ and /r/ and the conflation of [n] and [l], are more likely to affect intelligibility. Consonantal substitutions (except of dental fricatives and dark /l/) are disallowed in Jenkins’ *Lingua Franca Core*

(LFC) of features that help to maintain intelligibility (Jenkins 2000). On the left-hand side of the scale, L vocalisation and TH stopping are seen as being unproblematic for intelligibility. The positions of initial CCM and TH fronting would need to be reversed in order to create a perfect ordering of features according to their intelligibility characteristics, as the latter is not included in Jenkins' LFC. However, the general tendency for intelligibility-reducing features to occur less often is visible. A general indication of Table 3.5 is that most of the speakers (15 out of 25) would be highly intelligible internationally, according to Jenkins' (2000) criteria; these speakers are the ones below speaker 18, who do not use any intelligibility-reducing features. Of course, the actual effects of feature use on intelligibility need to be assessed in further studies before definite conclusions can be made.

The feature ordering may be related to the concept of a 'cline of intelligibility', as postulated by Kachru (1990, 1992a). While this concept highlights the role of intra-speaker variation, there is also an inter-speaker dimension in that speakers are likely to have been prioritising intelligibility over solidarity, or to use the terms of Kirkpatrick (2007a), focusing on communication, rather than identity. The inter-speaker variation that exists may therefore indicate the constraints that limit the ability to 'converge upwards' and perform acts of identity. If an interlanguage phonology perspective is taken and it is assumed that the differences in inter-speaker feature use represent different levels of phonological development, then the implicational scale may indicate an 'acquisitional hierarchy' (Mesthrie and Bhatt 2008: 94). This means that the features on the right-hand side of the scale tend to be acquired earlier, while those on the left persist for longer. Furthermore, the features on the right-hand side – /v/ and /r/ substitution, and [n, l] conflation – can arguably be attributed to transfer from the L1 (Cantonese), while the other features are more likely to be developmental in nature, possibly as a result of markedness. The tendency for transfer-related features to diminish as acquisition proceeds has already been identified as a principle of L2 phonology acquisition models, such as Major's Ontogeny Phylogeny Model (Major 2001) and the longitudinal study conducted by Hansen (2006).

In Hansen's model there is a four-stage developmental sequence constrained by L1 transfer effects, developmental effects and markedness (Hansen 2006: 153). At the first stage, users make equivalence classifications and consonants that are similar in type and position are transferred. Thus Cantonese does not possess close equivalents of /v/ and /r/, and [w] is used as a substitution. At the second stage of development, consonants are typically modified towards the emerging L2 repertoire, while transfer is still a constraint. At stage 3, transfer effects decrease and more marked consonants begin to emerge, with developmental and markedness effects continuing to affect some sounds. Stage 4 of Hansen's sequence is characterised by 'the approximation of a native speaker-like phonology, which may still include some errors' (Hansen 2006: 155). The 'errors', or in the terminology of this study the 'features' that persist for longest in high-proficiency HKE appear to be L vocalisation and dental fricative substitutions such as TH stopping and TH fronting. These features involve the avoidance of marked phenomena and are widely attested features of many L1 and L2 varieties. It is arguable that the speakers below speaker 18 in Table 3.6 have reached stage 4. They do not use any intelligibility-reducing features, and the features they do use also occur in many native speaker accents.

As mentioned in section 3.6.4, the above explanations should not be seen as resting on the assumption that speakers are moving along a cline of proficiency. As Bolton and Kwok (1990: 149) observe, there are many highly-educated speakers who still retain 'many localised features of speech, particularly at the accent level'. The mini-corpus reflects the fact, noted by Bolton and Kwok (1990: 150) that even among 'higher-range' speakers who appear in the local media, such as civil servants, businesspeople and educators, relatively few speak with 'native speaker' accents. Even those who do not use any HKE consonantal features are likely to have other HKE features, for example in the areas of vowel realisations and temporal properties. Nevertheless, the implicational scale provides more detail about how patterns of feature use can be used to differentiate accents. As already mentioned, and although intra-speaker variation has not been researched in detail in Hong Kong, there may be ways in which the speakers are 'designing' their speech for their audience (in the terminology of Bell 1984). The implicational scale may also reflect how this variation or style shifting actually takes place, given that intra- and inter-speaker

variation are seen to be mutually interdependent (Bell 1984). Thus the implicational scale allows phonological variation to be considered from both an ‘interlanguage phonology’ perspective and an ‘acts of identity’ perspective.

The preliminary study thus provides useful data about the actual rates of occurrence of certain features within a sample of relatively high-proficiency speakers. As such, it enables evaluation and discussion to focus on those features that appear to occur most frequently, or persist for longest, in HKE. The above explanations of speaker variation are necessarily tentative in nature, but will still be used to inform the discussion of listener reactions to the speech samples in the main study, as part of an overall explanatory model of phonological variation and development within L2 contexts.

3.7 The research space

There is thus a clear need for further research into the local accent, at several levels: its phonological features (taking account of variation and proficiency level, and thus avoiding the generic approach to HKE visible in most previous studies), and its characteristics in terms of intelligibility, acceptability and so on. There is a particular need for studies of acceptability, as this is important for pedagogical purposes and there seems to be an absence of research data obtained from a detailed, features-based perspective. The preliminary study, in conjunction with other studies of HKE phonology, has already indicated that feature use varies according to the position of speakers on a cline; this may be characterised as a cline of proficiency (representing phonological development and acquisitional hierarchies) or of identity (therefore possibly including both intra-speaker and inter-speaker variation, as speakers may have both ‘temporary’ and ‘permanent’ identities). The main study will investigate the acceptability of HKE accent samples for local listeners, following this features-based, variationist orientation.

3.7.1 Research questions

One of the aims of the study will be to ascertain whether or not local students are likely to accept a local model for pronunciation teaching, while another will be to discover how different features affect judgments of acceptability. Selected features will then be further evaluated using the multidimensional model, where appropriate, in order to construct tentative guidelines for pronunciation models in Hong Kong.

The research questions of this study can thus be stated as:

1. Are local students likely to accept a local model for pronunciation teaching purposes?
2. What are the effects of different phonological features on students' judgments of acceptability?
3. What are the implications for pronunciation teaching? For example, are there any phonological features of 'standard' models that can be omitted from local teaching syllabi or testing materials?

Question 1 will require the use of a range of accent samples, which will be subjected to acceptability judgments. To address question 2, statistical analysis will be used to assess the effects of phonological features on acceptability ratings. Question 3 will make further use of the approach to phonological feature evaluation introduced in Chapter 2. It will thus take account of external, indirect evidence relating to factors such as intelligibility, as well as the internal, direct evidence relating to the acceptability of HKE features. It will also be informed by the data on the distribution of these features that have been generated by the preliminary study. Chapter 4 will examine the methodological approaches and techniques used to achieve these aims.

CHAPTER 4

METHODOLOGY

4.1 Introduction

The research focus of the main study is thus on the acceptability aspect of the Hong Kong accent, or in other words on Quadrant 4 of the four-quadrant evaluation model. The rationale for this is that the intelligibility characteristics of features can be adequately assessed using the criteria outlined in Chapter 2, with reference to empirical research where possible. Of the non-linguistic factors in the evaluation model, while individual goals and identities should be taken into consideration, an assessment of acceptability is crucial in order to address the research questions. Given the pedagogical orientation of this study, it is vital that any proposed models are acceptable to their intended audience. There is also a lack of features-based data on the acceptability of different types of Hong Kong English accent, although the study of Bolton and Kwok (1990) suggests that higher acceptability is linked with ‘milder’ accents that contain relatively fewer local features.

The notion of acceptability is inextricably linked with language attitudes, including the unavoidable fact that evaluative judgments involve perceptions of the speaker as well as of his or her speech. Nevertheless, this study will rely on a carefully-constructed measurement of ‘acceptability’ and a rigorous methodological approach that takes account, as far as possible, of competing factors. Given the features-based orientation of the study, it will attempt to link audience ratings with the phonological features in the samples. The general approach involved collecting a range of Hong Kong accent samples and presenting them to student listeners for evaluation, thus avoiding native speaker judgments. The experimental design of such ‘accent studies’ needs to be carefully considered, and in this chapter the methodological approaches

employed in previous studies will be characterised before the study's own approach is outlined.

4.2 Previous studies

Accent studies have used various methods to attempt to satisfy the demands of statistical or explanatory rigour. The Hong Kong-based studies of Forde (1995), Luk (1998) and Candler (2001) have already been discussed in Chapter 3. While their goals were similar to this study's in that they investigated learners' attitudes towards different accents (including the Hong Kong English accent), these studies did not attempt to control the Hong Kong accent sample for its phonological features. In addition, they used only one Hong Kong English sample, thus ignoring the considerable inter-speaker variation that exists (see Chapter 3). A first objective for research, then, must be to acknowledge this variation by selecting accent samples according to meaningful criteria, and to avoid the 'generic' approach to accents. Accent studies in general have used samples of connected speech, rather than word lists, to investigate attitudinal reactions. There are two main ways to achieve validity and reliability: the use of standardised reading passages, and the matched-guise technique. These two approaches will be considered below.

4.2.1 Standardised reading passages

Within the 'hundreds' of language attitude studies that have been carried out all over the world, according to Cargile and Giles (1997: 195), the use of standardised reading passages is a popular way to reduce the variation between samples. The intention is that linguistic variables are controlled, so that variation is confined to the speakers and accents employed. An example of such a study is Bayard *et al.* (2001), which compared native speaker and non-native speaker evaluations of L1 English accents (New Zealand, Australian, North American and British). The evaluation questionnaires employed both 6-point Likert scales and multiple choice questions, and used four underlying constructs to select questionnaire items, namely 'status' (e.g. education and income), 'power' (e.g. authoritativeness and assertiveness), 'solidarity' (e.g. friendliness and cheerfulness) and 'competence' (e.g. intelligence

and reliability). The study used recordings of male and female native speakers with each of the accents. Although the phonological features of these accents were noted, the researchers concede that paralinguistic variation did exist in the samples (despite attempts being made to minimize the effect of variables such as reading speed and the number of hesitations). However, after performing a digital manipulation of the samples which removed one of these variables (intonation), the researchers conclude that ‘while intonation is clearly important, phonological accent appears to have the greater effect’ (Otago 2002). Once again, the accents were not considered from a detailed, features-based perspective.

While the use of standardised reading passages offers several advantages, there is the danger that hyper-articulated speech and spelling pronunciations will be used, reducing the authenticity of the samples; an approach that uses unscripted, spontaneous speech samples is more likely to reflect actual performance features. Reading aloud is a fairly common classroom activity in Hong Kong, but the ability to speak spontaneously is also important, both inside and outside the classroom. A possible methodological problem with using standardised passages in accent studies is that if there are only small differences between samples, as a result of trying to assess the effects of particular features, it is uncertain whether the listeners will be able to maintain attention and make meaningful judgments after, say, ten or twelve samples. Some listeners will also realise that the experiment is designed to measure their reactions to particular features, and by focusing on these it is more likely that they will react to the feature, and its stereotypical associations, rather than the speech sample as a whole. Furthermore, if it is accepted that some phonological features tend to occur in implicational patterns, as suggested by the implicational scale shown in Chapter 3, the accent samples would still need to contain more than one feature at a time. For example, attempting to measure the effects of TH fronting by including it and it alone in a reading passage would be unrealistic because it tends to co-occur with other features.

4.2.2 The matched-guise technique

The use of reading passages reduces the amount of linguistic variation in samples, but if different speakers are used it cannot control for speaker variables such as tone of voice. In order to achieve this, the matched-guise technique (Lambert 1967) is frequently employed in accent studies. An actor or other professional speaker reads samples according to detailed instructions from the researchers, thus controlling accent features and reducing extraneous variables to some extent. This approach is taken by van den Doel (2006, 2007). In van den Doel's study, the two actors were instructed to maintain the same accent (Dutch English) while manipulating different phonological features, in order to determine the 'hierarchy of error' for both Dutch and non-Dutch listeners. This study went to great lengths to eliminate unwanted extralinguistic and linguistic variables, so that the age and gender of the speakers were matched, and the performances were checked by phoneticians. However, in considering the results van den Doel (2006: 244) concludes that at least some of the variation 'may be derived from differences in performance between the two actors', and that this was a limitation of the experiment. Further analysis of the samples revealed that there were slight variations in prosody and at least one case of unintentional variation in segmental realisation between the two versions (van den Doel 2006: 305).

The matched-guise technique also has its limitations, and the researchers involved in the Otago study (Otago 2002) rejected it on the grounds that '[i]t seemed to us impossible to find speakers of each gender who could give convincing renderings of all four accents without falling into the trap of projecting a stereotype rather than the genuine article'. A further example of the unnaturalness of the matched-guise method is provided by the comment of the speaker cited in Cargile and Giles (1997: 201), who reported that it felt 'increasingly "unnatural" to speak English with a moderate and heavy Japanese accent as fast as he spoke it with a standard-American accent'. In this case, the unnaturalness was probably a result of linguistic factors. If the reader or researcher employed the avoidance of vowel reduction as one of the features of a 'heavy' accent, this would tend to slow down speech rate by increasing the time and physical effort needed for articulation. Once again, the possible

implicational patterning of phonological features makes the procedure somewhat unnatural and reduces the external validity of research.

4.2.3 Authentic data

There is no easy way to reconcile the imperatives of ‘hard’ data with the need for authenticity. Following her investigation of the intelligibility, comprehensibility and interpretability of South Korean speakers of English for South African listeners, Coetzee-Van Rooy (2009: 33) concludes that researchers should ‘continue the struggle to work with authentic data’, although this is time-consuming and yields ‘unwieldy’ data (Van der Walt 2000, in Rajadurai 2007: 96). In the present study it was decided to prioritise authenticity by using recordings of speakers on television programmes, subjecting the samples to feature-based selection and analysis while acknowledging the possible influence of competing factors through careful experimental and statistical design. However, while some of these factors may be quantifiable and amenable to statistical investigation, others will not be. The statistical analysis will attempt to identify the phonological features that had the greatest influence on the acceptability ratings, but the nature of ‘causation’ in this study does not lay claim to the kind of validity that would be required in the natural sciences (Altendorf 2003: 139). In particular, the explanatory part of the study does not attempt to identify single causative factors, but rather takes an ‘ecological’ perspective by considering the combined effects of various factors.

4.3 Sample selection

It has already emerged that accent samples representing ‘educated’ or ‘high proficiency’ users of English are desirable for a study of this nature; Kirkpatrick (2007b: 387) believes that Hong Kong needs a description of the local bilingual variety of English ‘as exemplified by highly proficient users of English who are mother tongue speakers of Cantonese’. The importance of proficiency is emphasised by Bolton (2008: 11), who notes that ‘the issue of proficiency in the Asian context has now gained a good deal of prominence’. A consideration of the possible

processes of development, as well as the nature of variation, is also central to the present study and has been integrated through the use of the implicational scale and its associated data.

A range of accents was required, so that listeners could make their own decisions about acceptability, perhaps in terms of an ‘acceptability threshold’. This concept is related to the idea of ‘tolerance testing’, originally developed by Faerch and Kasper (1983) and employed by Van der Walt (2000: 141): ‘the effectiveness of *communication in context*...should be tested on the most probable audience for this communication, so as to ascertain how well it is “tolerated” by this audience’. The general idea of a threshold related to proficiency is also advanced by Rajadurai (2007: 102), who believes that ‘core linguistic features that constitute a minimum threshold level’, as well as ‘overall proficiency and an adequate repertoire’ are prerequisites for communication. Rajadurai is referring to general language proficiency, but it is possible that similar considerations apply to phonological features. The idea of ‘core’ features in fact echoes Jenkins’ LFC, which emphasises the utility of most phonological contrasts. This provides further justification for focusing on high-proficiency speakers and excluding those who clearly do not possess an ‘adequate repertoire’, while not prejudging samples to such an extent that the resultant selection is unrepresentative. The mini-corpus of television programmes used in this study has already been described in Chapter 3; the following sections will examine some of the general advantages and disadvantages of using broadcast material.

4.3.1 Advantages of broadcast material

The use of accent samples from television programmes on Hong Kong’s English-language channels appeared to be the most effective way of achieving these desired outcomes. Broadcast material offered several advantages: authentic, mainly unscripted speech (although this depends on the genre); a wide range of accents, speakers and topics; the absence of an ‘observer effect’ insofar as there is no experimenter present; and ease of recording. In a study of the international comprehensibility of varieties of South African English (Van der Walt 2000), the

researcher reaches a similar conclusion: 'recordings of television and radio broadcasts meant for general consumption were found to be the most authentic type of communication that could be accessed, recorded and repeated (for subsequent testing) with relative ease' (Van der Walt 2000: 142).

It could be argued that by agreeing to appear on television, the speakers were themselves reasonably confident about their English proficiency, and were thus preselected for proficiency level (although there were wide variations in the use of phonological features). On the other hand, the samples did not seem to represent uncommon or unattainable levels of proficiency. The use of genuine communicative contexts also suggests that the speakers were focused on the content of their speech, further reducing the 'observer effect' and the threat of speakers self-consciously modifying their speech, either in response to being recorded or because of the influence of spelling pronunciation. The comprehensibility of the original utterances was also more or less guaranteed by the demands of the situation; presumably, if the interviewer or producer had doubts about the comprehensibility of an utterance he or she would have asked for clarification or repetition, or simply edited the unsatisfactory material from the final programme. In her comprehensibility study, Van der Walt (2000) provides three justifications for the use of television programme excerpts in accent studies, the first two of which are especially relevant to this study:

1. the message must be part of a communicative event which provides its context (in this case both the context of the communication created by interviewers and interviewees on radio and TV and the context of the viewer as audience and interested listener);
2. the message must be directed at a receiver whom the speaker assumes listens for a reason (radio and television programmes deal with topics regarded as interesting or useful for that audience (Van der Walt 1999: 11, cited in Van der Walt 2000: 142).

4.3.2 Disadvantages of broadcast material

There are also several disadvantages in using television programmes as a source of recordings, most of which relate to the threats to internal validity arising from the use of different speakers and authentic speech samples. Van der Walt (2000: 141) describes her approach as ‘an uncomfortable attempt to “live in both worlds”, to satisfy a need for “hard statistics” and for a more context-bound, socially responsive testing procedure’. With any accent study, listener ratings may potentially be affected by a number of variables that go far beyond the phonological features of the accents. Some of these variables are more or less quantifiable, such as grammatical and lexical complexity, speech rate and fluency markers such as pausing and rephrasing, and discourse context. Others are difficult or impossible to measure, such as tone of voice. Variations in speaker age, gender and occupation also have to be taken into account.

Another possible criticism of television programmes as a source of accent samples is that they represent a somewhat artificial context of use. Despite the promise of genuine communicative contexts, the main target audience of a programme such as *The Pulse* will presumably be native speakers of English (but not necessarily: most readers of the English-language *South China Morning Post* are in fact ethnically Chinese; *South China Morning Post* 2009a). Although media English reduces one kind of observer effect through the absence of the researcher, it introduces another in the form of an ‘audience effect’. The speakers may have been ‘designing’ their speech, using the term of Bell (1984), for the audience. Furthermore, in the case of studio discussions and studio interviews there is normally a native speaker interlocutor present in the role of programme host. The speakers might therefore be expected to approximate native speaker norms more closely than if no native speakers were present.

However, the threat posed by accommodation or audience design is not a serious disadvantage for this study, for two main reasons: firstly, accommodation can be seen as an additional source of variation, and is therefore still amenable to acceptability judgments; secondly, as noted in Chapter 3, speakers can only

accommodate as far as their phonological and articulatory systems will allow. While the existence of such variation is relatively unproblematic at the methodological level, at the explanatory level an awareness of the possibility of accommodation or style shifting must be maintained. The interdependence of linguistic and non-linguistic factors is acknowledged in this study, following the viewpoint of LePage (1980), who sees ‘no such thing as a grammar independent of social life’ (in Fasold 1990: 198).

4.4 Selecting the survey samples

The implicational scale presented in Chapter 3 provided an analytical tool for the next stage of the study, namely the selection of twelve accent samples. It was used to try and ensure that a range of accents was included, a range that contained all of the consonantal features investigated in the preliminary study, for example. Furthermore, certain principles of selection (such as the absence of grammatical errors) were also applied, and these made obtaining the required range more problematic; the implicational scale suggests that certain phonological features tend to co-occur, but it was also noticeable that some of these feature combinations tended to co-occur with grammatical problems. This suggests that there may be a ‘proficiency cline’ that includes grammatical, as well as phonological, features.

As the general orientation of the study was to allow student listeners to evaluate the samples and note their phonological features without them being prejudged by native speakers, a detailed analysis of the samples was not performed at this stage. They were subjected to an initial analysis using implicational scaling, however (see section 4.5.1 below). The samples were taken from both the existing mini-corpus and from subsequent television programme recordings, using the same criteria:

- **Completeness:** a complete ‘phonemic clause’ was the minimal acceptable unit. The concept was developed by Boomer (1978) and is defined by Crystal (2003: 348) as ‘a grammatical structure produced within a single intonation contour’. These clauses are often (but not always) related to syntactic clause boundaries (Cruttenden 1986: 75, in Laver 1994: 492).

- Length: an excerpt length of around ten seconds was thought to be ideal in terms of providing sufficient time for evaluation while allowing the desired number of samples to be played during the projected timeslot. Longer samples were thought likely to reduce listener concentration, given that in the actual survey twelve accent samples were provided. There was also the danger that longer samples might overload the listeners with phonological features and make it difficult for them to focus on the important features, possibly increasing the unwieldiness of the data.
- Fluency: samples with disfluency phenomena such as lengthy filled or unfilled pauses, false starts and so on were avoided as far as possible.
- Accuracy: samples with prominent grammatical errors were also avoided in the selection process. There were some difficulties in distinguishing between grammatical and phonological error, and this will be discussed further in due course.

Table 4.1 below shows brief biographical details of the twelve speakers, along with the sources of the recordings and the contexts of speech. Table 4.2 shows transcripts of the twelve accent samples. The tables provide an overview of the samples in terms of origin, length, and type of language.

Table 4.1. Speaker biodata and source of recording.

Speaker No.	Gender M/F	Origin (presumed)	Approx. age	Occupation	Source and context of recording	Type of speech
1*	M	Hong Kong	50s	Journalist	<i>The Pulse</i> December 2007 (studio discussion)	Unscripted
2	M	Hong Kong	50s	Politician	<i>The Pulse</i> February 2008 (recording of public address)	Scripted
3	M	Hong Kong	50s	Journalist	<i>The Pulse</i> May 2007 (studio discussion)	Unscripted
4	M	Hong Kong	50s	Politician	<i>The Pulse</i> May 2007 (studio discussion)	Unscripted
5	M	Hong Kong	50s	Government or industry spokesperson	From HKICE (recording of public address)	Scripted
6*	M	Hong Kong	50s	Journalist	<i>The Pulse</i> May 2007 (studio discussion)	Unscripted
7	M	Hong Kong	60s	Civil servant (retired)	<i>Pearl Report</i> March 2006 (interview)	Unscripted
8	F	Hong Kong	50s	NGO chairperson	<i>Pearl Report</i> March 2006 (interview)	Unscripted
9	F	Hong Kong	30s	Government or industry spokesperson	<i>The Pulse</i> June 2007 (interview)	Scripted (probably)
10	M	Hong Kong	40s	Politician	<i>The Pulse</i> April 2007 (studio interview)	Unscripted
11	M	England	30s	Journalist	<i>Pearl Report</i> March 2006 (studio interview)	Unscripted
12	M	Hong Kong	50s	Journalist	<i>The Pulse</i> December 2007 (studio discussion)	Unscripted

* ‘Speaker 1’ and ‘Speaker 6’ are the same person.

Table 4.2. Transcripts of the twelve accent samples.

Speaker No.	Transcript
1	They don't see an advantage in doing anything risky, and they don't have to because they think that they have all the cards now
2	You can see the words commitment, sustainability and pragmatism. In the past year the economy has continued to perform well and we have built up a considerable surplus
3	China itself is quite heterogeneous these days, and there are many local identities, so to speak. So if we take a more relaxed attitude of national identity, I don't think we should be too bothered by it
4	Well I think that the very concept of one country, two systems suggests the people of Hong Kong should try to at least maintain some of their own attributes
5	The applications of information technology in the clothing industry are diverse and varied, and it is impossible to cover all the options in two days
6	There's no reason why the new leadership in Beijing would be more forthcoming, you know, in terms of granting Hong Kong a high level of political participation
7	The question we need to ask is: does the public want KCRC run like a government department? MTR run like a government department?
8	There are many children who are not as privileged as we would like to think they should be
9	The accredited fish farm scheme aims at assisting the local fish farmers to enhance their operation and production standards
10	Actually I have been with the party for a long time, ten years to be exact, but I have been serving mostly as central committee member and standing committee member
11	The quality migrant attraction scheme seeks to attract talented people and also talented people to bring their families with them
12	And when they found, virtually there's no progress on democracy, I think people are confused

As mentioned above, it proved difficult to find samples that met the important criterion of not having grammatical errors, while still including a range of phonological features. In fact, this made it necessary to extend the search beyond the HKE mini-corpus, and the Speaker 5 sample was obtained from among the Hong Kong samples in the International Corpus of English (2010). Even so, Speaker 5 deletes the final consonant of *applications*, and the question thus arises of whether this was a grammatical or a phonological error. The fact that the utterance later contains the plural *are* suggests the latter, unless the speaker's notes were incorrect (this was a public address, rather than an interview, so the speech was probably scripted). However, while this sample was included in the initial assessments of acceptability, it was removed from subsequent statistical analyses because it was felt that its inclusion increased the risk of conflating the effects of grammatical and phonological error.

It was also hoped that the samples would provide equal numbers of male and female voices, but the television programmes viewed had fewer female participants. Ultimately, the twelve samples included ten male and two female voices. While studio recordings were thought to be ideal because of the low level of background noise, it was impossible to find a sufficient number of such extracts that also conformed to the other selection criteria. The contexts of the twelve samples thus consisted of two rehearsed speeches, probably given at press conferences (Speakers 2 and 5) and three non-studio interviews (Speakers 7, 8 and 9). Of the latter three, Speaker 9 is probably using scripted speech, while the other two seem to be extracts from unscripted interviews. The remaining samples all came from studio discussions. The possible differences between these contexts in terms of speaking style should also be taken into consideration when interpreting the results. Again, while the use of unscripted speech would seem to be preferable, it proved impossible to find a sufficient number of suitable samples. One reason might be that only the most confident and proficient speakers employ unscripted speech in media situations, and restricting samples to this subset would considerably reduce the range of speakers and phonological features.

Two of the twelve recordings, numbers 1 and 6, were of the same person. They were included to investigate whether the slightly different phonological features in each recording affected the ratings, and to provide an additional indication of rater reliability. Eleven of the twelve different speakers were assumed to be native speakers of Cantonese, judging from their English accents. One of the speakers (Speaker 11) was a British native speaker of English. The acceptability data pertaining to this speaker was compared with that of the Hong Kong speakers in order to provide a comparison, but was excluded from subsequent features-based analysis in order to confine the study to the phonological features of Hong Kong English. Overall, the samples appeared to offer a range of ‘typical’ Hong Kong accents, with one possible exception. Speaker 10 was born in Hong Kong and also has L1 Cantonese, but spent his teens and twenties in Canada. As a result he has a rather different accent, some features of which (such as rhoticity) are distinctively North American, and other features of which (for example, TH stopping) may be related to Hong Kong English. This speaker provides an example of someone who

has moved towards ‘native-speaker competence’ as a result of receiving tertiary education abroad (Bolton and Kwok 1990: 149).

4.4.1 Linguistic preselection

As mentioned above, it was thought to be desirable for the samples to include a range of accents. Initially this was achieved through global, impressionistic assessments of accent strength, but an implicational scale was then used to assess the consonantal features present in the twelve samples. Table 4.3 below shows this scale for the twelve speakers.

Table 4.3. Implicational scale for the twelve accent samples.

		TH stop.	L vocal.	/v/ subst.	Initial CCM	/r/ subst.	TH front.	[n, l] conf.	Number of Y cells
SPEAKER NUMBER	6	Y	N	n/a	Y	Y	Y	N	4
	1	Y	n/a	Y	n/a	Y	N	n/a	3
	12	Y	n/a	Y	Y	n/a	N	n/a	3
	3	Y	Y	n/a	N	n/a	N	N	2
	5	N	Y	Y	N	N	N	n/a	2
	9	Y	Y	n/a	N	N	n/a	N	2
	11	Y	Y	n/a	N	N	n/a	n/a	2
	4	N	Y	N	N	N	N	N	1
	8	N	Y	N	N	n/a	N	N	1
	10	Y	n/a	n/a	N	n/a	n/a	n/a	1
	2	N	N	n/a	N	n/a	n/a	n/a	0
	7	N	n/a	n/a	N	N	n/a	N	0
Number of Y cells		7	6	3	2	2	1	0	

It should be noted at this point that the subsequent survey design allowed students to mark any features in the samples that they considered to be important; they were not limited to consonantal features. The scale in Table 4.3 is thus intended to serve as an indication of the range of accents, rather than as a definitive list of their features. The scale has been prepared on the same basis as that for Table 3.6 in section 3.6.4 of Chapter 3, and the basic implicational pattern is still visible in the diagonal division of the table between feature presence (more prevalent in the top left corner) and feature absence (more prevalent in the bottom right corner).

However, the reordering of the columns according to the number of speakers using the features has resulted in a different ordering of features. TH stopping and L vocalisation were still the most widely distributed features, but TH fronting has shifted to the right (there was only one instance in the samples). This is most likely to be due to the limited duration of the samples, which were not able to provide a sufficient number of possible contexts for some of the features. One third of the cells in Table 4.3 are empty ('n/a') because of this. Another consequence of the shorter samples is that there are relatively more 'deviant' cells in Table 4.3 than in Table 3.6. In Speaker 6's row, for example, there are three 'deviant' cells as a result of the non-appearance of L vocalisation; a longer sample would have created more chances for this feature to appear. The reversal of the TH stopping and L vocalisation columns, (compared with Table 3.6) has led to three 'deviant' cells in the LV column, and this suggests that Table 3.6 more accurately reflects the implicational hierarchy of these features. No measurement of the index of reproducibility was made in this case, as the high proportion of empty cells significantly reduces the meaningfulness of the measurement. Nevertheless, the general patterning of the scale above reinforces the tentative conclusions drawn in Chapter 3 about the distribution and implicational relationships of certain HKE consonantal features.

It can also be seen that while most of the features considered in Chapter 3 are included, there were no instances of the conflation of [n] and [l]. This was accepted as reflecting the low occurrence rate of this feature (conflation was found to be 'rare' by Deterding *et al.* 2008: 160). Similarly, there is only one occurrence of TH fronting (with Speaker 6), reflecting both the relatively infrequent occurrence of the /θ/ phoneme in English (see Chapter 2, section 2.9.1) and the low proportion of speakers displaying the fronted variant (see Chapter 3, section 3.6.3, Table 3.5). Unfortunately, from the outset it appears that the likelihood of drawing any meaningful conclusions about the effects of these features is low. A further disadvantage of using authentic samples is thus that the occurrence of word tokens and the instances of feature use within them are subject to language-specific and accent-specific constraints, respectively. This problem is exacerbated by the use of high-proficiency speakers, who tend to use fewer local features. Nevertheless, an acceptability study can accommodate these constraints by considering them to be

aspects of authenticity – features that occur infrequently, whether as a result of language-specific or accent-specific distributional patterns, will stand less chance of achieving statistical significance in terms of their effects on acceptability. The danger is of course that potentially significant features will be missed, and that features that appear to be insignificant within the samples may be highly significant in other samples, for instance in those taken from low-proficiency speakers.

Because of the constraints mentioned above, the accent samples range between having four consonantal features (Speaker 6) and none (Speakers 2 and 7). This may seem to be a limited number of features, but as the study's survey design allowed students to mark any features that they considered to be important, the data on acceptability effects were not limited to these features. It was highly likely that other features, such as those relating to vowels, and suprasegmental features, would be marked. Viewed in this light, the selection above seems to be wide-ranging and representative of a similar range of variation as was seen in the earlier implicational scale, which included 25 speakers from the mini-corpus. It might be argued that this preselection followed native-speaker assessments of 'proficiency', but as the scale is based on the occurrence of features it seems to be neutral and objective.

4.5 Other speaker variables

Inevitably, despite the attempts to standardize the samples as far as possible, there remained some important differences between them (in addition to the differences in feature use described above). The survey design attempted to address this by allowing listeners to make open-ended comments about the speakers, but nevertheless this is an unavoidable weakness of the 'authentic' approach. An alternative is to increase the number of samples so that variables such as 'subphonemic and phonetic irregularities' (Munro and Derwing 2006: 524) are unlikely to be distributed non-randomly across the stimuli. Munro and Derwing used 19 speech samples in their study, while the optimum number in the present study was thought to be twelve. However, while some features and extralinguistic factors are likely to occur non-randomly, thus offsetting their potential influence to some extent,

it is possible that paralinguistic features such as intonation may also occur in an implicational fashion. This would mean, for example, that samples with many segmental errors will also include more intonational errors, each compounding the effects of the other.

It was therefore decided to measure selected linguistic factors under the general headings of accuracy, complexity, fluency and other prosodic variables (including a partial, non-contextual measure of intonation in the form of pitch span). The first statistical operation conducted was to assess the relative effects of these factors on the acceptability scores. The factors thus formed the independent variables, along with listener assessments of general phonological accuracy, while the acceptability scores formed the dependent variable. In addition to these measured factors, other factors, such as the contexts of speaking, were assessed and noted without being quantified, so that they would be available for consideration at a later stage. Phonological accuracy was measured by counting the error codings on the survey forms, and hence this factor will be introduced with the other findings of the study in Chapter 5. The selected linguistic factors and their measurements are explained in more detail in the following sections.

4.5.1 Lexical and syntactic complexity

Although lexical choice and complexity is arguably not as crucial a consideration as it would be in a study of intelligibility, there is the possibility that speakers who used more complex vocabulary would appear more proficient or sophisticated, thus influencing the evaluations, and this should be considered when interpreting the results. To assess lexical complexity, two measurements were made. Firstly, the average number of syllables per word was computed, to provide an overall indication of lexical complexity. Secondly, words of three or more syllables were checked against a non-lemmatized, frequency-ranked corpus of 4841 words derived from the spoken English component of the British National Corpus (Leech, Rayson and Wilson 2001). An arbitrary score of between 1 and 5 was assigned to these words according to their position in the frequency rankings, so that the first 1,000 words in

the corpus received a score of 1, the second 1,000 words received a score of 2, and so on. Details of the lexical complexity scores are given in Table 4.4 below.

Table 4.4. Lexical complexity measures.

Speaker No.	No. of syllables	No. of words	Ave. no. of syllables per word	Words of three or more syllables	Position in frequency list	Complexity score	Total complexity score
1	31	24	1.29	advantage	1,930	2	2
2	48	28	1.71	commitment sustainability pragmatism economy continued considerable	2,185 n/a n/a 1,547 4,604 2,449	3 6 6 2 5 3	25
3	58	37	1.57	heterogeneous identities attitude national	n/a n/a 1,648 527	6 6 2 1	15
4	38	29	1.31	attributes	n/a	6	6
5	44	25	1.76	applications information technology industry impossible	2,256 449 1,737 991 2,519	3 1 2 1 3	10
6	42	27	1.56	leadership forthcoming political participation	4,390 n/a 1,049 n/a	5 6 2 6	19
7	38	23	1.65	government x2 department x2	328 875	1 1	2
8	22	18	1.22	privileged	n/a	6	6
9	33	19	1.74	accredited assisting operation production	n/a n/a 1,373 1,385	6 6 2 2	16
10	44	30	1.47	committee	2,775	3	3
11	35	20	1.75	attraction talented x2	n/a n/a	6 6	12
12	23	15	1.53	democracy	2,661	3	3
<i>Ave.</i>	38	24.6	1.5				9.92

A measurement of the syntactic complexity of the samples was also made, using as the unit of analysis the T-unit (see Hunt 1965). A T-unit will often have the same boundaries as a sentence, as it is defined as a main clause plus any subordinate clauses, with subordinate clauses including all adverbial, adjective and nominal clauses (Hunt 1965, in Wolfe-Quintero, Inagaki and Kim 1998: 71). In compound sentences, each main clause is a T-unit (see McKay 2006: 111). By counting the number of T-units and the number of clauses in each sample, a measurement of the number of clauses per T-unit can be derived as an indication of syntactic complexity. In calculating this measurement, only clauses with finite verbs were included (see Hunt 1965: 15), thus excluding participle, gerund and infinitive verb phrases. A measurement of T-unit length was also obtained by dividing the number of words by the number of T-units. Table 4.5 provides details of these measurements.

Table 4.5. Syntactic complexity measures.

Speaker No.	Transcript (T-unit boundaries shown by /; finite verbs in bold)	No. of T-units	No. of clauses	Clauses per T-unit	Words per T-unit
1	They don't see an advantage in doing anything risky, and / they don't have to because they think that they have all the cards now	2	4	2	12
2	You can see the words commitment, sustainability and pragmatism. / In the past year the economy has continued to perform well and / we have built up a considerable surplus	3	3	1	9.33
3	China itself is quite heterogeneous these days, and / there are many local identities, so to speak. / So if we take a more relaxed attitude of national identity, / I don't think we should be too bothered by it	4	5	1.25	9.25
4	Well I think that the very concept of one country, two systems suggests the people of Hong Kong should try to at least maintain some of their own attributes	1	3	3	29
5	The applications of information technology in the clothing industry are diverse and varied, and / it is impossible to cover all the options in two days	2	2	1	12.5
6	There's no reason why the new leadership in Beijing would be more forthcoming, you know, in terms of granting Hong Kong a high level of political participation	1	2	2	27
7	The question we need to ask is: does the public want KCRC run like a government department? / MTR run like a government department?	2	3	1.5	11.5
8	There are many children who are not as privileged as we would like to think they should be	1	4	4	18
9	The accredited fish farm scheme aims at assisting the local fish farmers to enhance their operation and production standards	1	1	1	19
10	Actually I have been with the party for a long time, ten years to be exact, but / I have been serving mostly as central committee member and standing committee member	2	2	1	15
11	The quality migrant attraction scheme seeks to attract talented people and also talented people to bring their families with them	1	1	1	20
12	And when they found , virtually there's no progress on democracy, / I think people are confused	2	3	1.5	7.5
Ave.		1.83	2.75	1.69	15.84

Tables 4.4 and 4.5 show that there are substantial differences in lexical and syntactic complexity. Speaker 2 has the highest lexical complexity score of 25, but below-average syntactic complexity scores. This is probably a result of the speech genre (a prepared speech), rather than any difference in language level. The opposite case (high syntactic complexity accompanied by low lexical complexity) is demonstrated by Speaker 8. Although the two measures may appear to be negatively correlated, subsequent analysis revealed an absence of significant correlation. Comparisons with the native-speaker scores (Speaker 11) support the characterisation of these HKE speakers as being generally ‘high-proficiency’ or higher-range.

4.5.2 Fluency and speech rate

The speech rate of the speakers was measured by computing words per minute and syllables per minute (see Table 4.6 below). The effect of speech rate on evaluative judgments was investigated by Munro and Derwing (1998). They found that while the optimal speech rate for non-native speakers may be somewhat slower than the optimal rate for native speakers, there was no benefit derived from the strategy of deliberately speaking at a slower rate. In fact, the researchers conclude that non-native speakers who speak ‘especially slowly’ might benefit from increasing their speaking rates (Munro and Derwing 1998: 179). While this may be partly due to linguistic phenomena such as the effect of rate reduction on accentual features, it is worth considering that there may be associated evaluative judgments arising from speech rate. Munro and Derwing (1998: 179) note that ‘NNSs typically speak more slowly than NSs’ and further cite evidence that speaking rate ‘may be a reliable index of overall oral proficiency’ (Guion *et al.* 1997, in Munro and Derwing 1998: 163). Proficiency in general may be a desirable attribute, according to Llurda (2000), who found that evaluative variables related to intellectual capacities were highly correlated with language proficiency (as measured by TOEFL scores). So while one could conclude that the faster speakers had an unfair advantage, this may be a reflection of their underlying language proficiency (or at least, the students’ perceptions of their language proficiency).

A problem that arises when submitting speech rate measures to statistical analyses along with other factors is that there is likely to be an optimum speech rate, on either side of which ratings tend to decrease. Thus the use of correlational statistics may have limitations in this area. According to Munro and Derwing (1998), a speech rate of 207 words per minute was preferred by NSs listening to NS speech production (Foulke and Sticht 1966), and NS listener comprehension tended to decline when rates surpassed 250 words per minute (Foulke 1968). Among the twelve samples, speakers 1 and 10 both surpass the 207 wpm level, but neither exceed 250 wpm. The speakers with the below-average rates tended to be using rehearsed or scripted speech (speakers 2, 5 and 9). Speaker 7 employs a deliberately slow, rhetorical speech style in order to make his point (and ask his rhetorical questions). It therefore seems unlikely that speech rate was related to proficiency level, within these samples; the speech rate of the native speaker is below the average.

Table 4.6 lists the speech rates in syllables per minute and words per minute, and also lists disfluency phenomena in the samples. These were defined as filled or unfilled pauses with a duration of more than 0.25 seconds. In general, the data provide further evidence to support the characterisation of the speakers as being ‘high proficiency’, and none of the disfluency phenomena appear to be of a type unlikely to occur in the speech of native speakers.

Table 4.6. Fluency factors.

Speaker No.	No. of words	Length of utterance (seconds)	No. of syllables	Syllables per minute	Words per minute	Disfluency phenomena
1	24	6.5	31	286	222	1 unfilled pause after <i>because</i> (0.349 seconds)
2	28	12.1	48	238	139	None
3	37	12.0	58	290	185	1 filled pause after <i>many</i>
4	29	8.7	38	262	200	2 unfilled pauses after <i>people</i> , <i>should</i> (0.265 seconds, 0.405 seconds)
5	25	10.6	44	249	142	None
6	27	8.7	42	290	186	None
7	23	10.2	38	224	135	None
8	18	5.9	22	224	183	None (pause of 0.5 seconds after <i>privileged</i> but at phonemic clause boundary)
9	19	9.0	33	220	127	None
10	30	8.2	44	322	220	1 filled pause after <i>serving</i>
11	20	8.1	35	259	148	1 filled pause after <i>attract</i> ; repetition of /t/ in <i>talented</i>
12	15	5.6	23	246	161	None
Ave.	24.6	8.8	38	259	171	

4.5.3 Prosodic factors: pitch span

Other prosodic variables that might affect perceptions of a speaker include the use of voice pitch. While pitch range refers to the habitual range of pitch exploited by a speaker, pitch span is defined as the difference between maximum and minimum pitch within an utterance (Laver 1994: 155). Pitch range is classed as an ‘extralinguistic’ factor by Laver, along with other variables such as loudness and voice quality; these are seen as being ‘rich in evidential information about the identity of the speaker’ (Laver 1994: 23). Few studies of L2 speech have investigated

the effects of these factors, but there is some circumstantial evidence to show the importance of pitch range. The influence of Cantonese (a tone language) may mean that speakers are more accustomed to pitch changes within syllables, rather than over phrases or clauses (Tay 1973, cited in Chan and Li 2000: 82). It is common to hear criticisms of the lack of intonation in some speakers' productions, and Chan and Li (2000: 83) claim that 'native speakers of English, and even Cantonese speakers of English like the present authors, would sometimes perceive such an English pronunciation pattern as "flat and boring"'. On this basis one would therefore expect pitch span to be positively correlated with listener ratings of acceptability, although this may depend on the proficiency level of the speakers. As with speech rate, there may be a point at which further increases in pitch span become non-optimal in terms of listener reactions.

Measurements of each speaker's minimum and maximum frequencies during their utterances were made using the WASP program (part of the Speech Filing System developed by the Department of Phonetics and Linguistics at University College, London and available for download at: <http://www.phon.ucl.ac.uk/resource/sfs/>). The results are shown in Table 4.7 below. As there were only two females it was not possible to carry out any meaningful statistical analysis of the effects of this variable in their case, so only the ten male speakers are included. Figure 4.1 shows the maximum and minimum frequencies in graphical form.

Table 4.7. Pitch span measurements for the ten male speakers.

Speaker No.	Maximum frequency	Minimum frequency	Pitch span
1	190	55	135
2	190	100	90
3	210	60	150
4	205	70	135
5	180	60	120
6	185	55	130
7	240	60	180
10	145	60	85
11	160	60	100
12	180	55	125
<i>Ave.</i>	<i>189</i>	<i>64</i>	<i>125</i>

Figure 4.1. Maximum and minimum frequencies in Hz for the ten male speakers.

It is noted by Ladd (1996: 260) that pitch level and span are often not distinguished in research studies because they tend to co-vary (the higher the level, the wider the span). This is apparently the case with the above data, where the Pearson correlation coefficient between maximum frequency and pitch span is 0.88 ($p < 0.01$, two-tailed). It can be seen that minimum frequency was approximately the same for all the speakers, with the exception of Speaker 2 (this was a public address and the higher minimum frequency may have been a result of various factors, such as the presence of background noise, the acoustics of the venue and the desire not to appear too relaxed). The higher maximum frequency and pitch span of Speaker 7 is probably a result of the rhetorical questions in his utterance. It may be noteworthy that Speaker 11 (the native speaker) and Speaker 10 (the near-native speaker) both had lower than average pitch spans, as a result of lower maximum frequencies; the ultimate reason for this is unknown.

The measurements of complexity, speech rate and pitch span will be analysed along with phonological accuracy for their effects on the acceptability scores. The results of this analysis are presented in Chapter 5, section 5.3.2.

4.6 The survey respondents

The survey respondents were first year undergraduate students at Lingnan University in Hong Kong, who had nearly completed a fourteen-week course in Practical Phonetics. The majority of the students were English majors. The main reasons for this choice included convenience, language proficiency considerations, and the existence of a certain level of phonological awareness. The students were already grouped in three classes of roughly equal size, enabling the accents to be played over the classroom PA system. As English major students, they were thought to be less likely to experience comprehension problems when encountering the authentic samples used in the study. The students had almost completed a course in Practical Phonetics, which provided an overview of English phonetics and phonology. The students were thus familiar with the terminology used on the survey form, which required them to note and categorize phonological features. Finally, as university students and English majors, they arguably constitute a ‘language élite’ whose views regarding the acceptability of phonological features may be influential, while still being fairly representative of their generation.

The possible drawbacks of using this survey population include the fact that as English majors, they might be expected to have a more norm-oriented outlook than other majors with a lower proficiency level. The preponderance of female students may further increase this tendency. Although the Lingnan students’ proficiency profile is not as high as in some other universities in Hong Kong, it could be argued that their comparative advantage in English might lead them to favour standard over non-standard forms. However, given the strongly exonormative language attitudes found in Hong Kong, this would perhaps tend to occur to some extent with almost any sample. It is perhaps noteworthy that 26 of the 52 students were following a

four-year degree programme in Contemporary English and Education, and may be even more norm-oriented as probable future teachers.

While it might be possible that the presence of a native speaker as course lecturer and survey administrator also encouraged them to value external norms, the survey itself was designed to focus on the relative acceptability of the samples. Even if the overall ratings were lower than they might have been with a different survey population, the relative ratings and their correlation with phonological features would still be of interest. In general, there appeared to be considerable advantages in using a group with some phonological awareness, over more ‘naive’ listeners, in a study focusing on the acceptability of phonological features. Given the aims and content of the course they had followed, their knowledge of metalanguage should have been sufficient for an understanding of such terms as ‘consonant clusters’, even though their ability to perceive variation in these areas may have varied quite widely.

The students were already formed into three groups of around twenty students. Not all of the respondents reported their L1 as being Cantonese, and these were excluded from the data analysis. The total number of eligible respondents was thus 52 (two groups of 17 and one of 18). The respondents were asked to note their L1, gender and degree programme, and to provide a self-rating of their English pronunciation skills. Details of the three groups of student listeners are provided in Table 4.8 below.

Table 4.8. Details of the three groups of respondents.

	Group 1	Group 2	Group 3	Overall
Number of students	17	17	18	52
Gender balance	15 F, 2 M	16 F, 1 M	16 F, 2 M	47 F, 5 M
Distribution of majors	English & Education 6% English Studies 88% Other 6%	English & Education 94% English Studies 6%	English & Education 39% English Studies 55% Other 6%	English & Education 50% English Studies 46% Other 4%
Self-rating of English pronunciation skills	Good 59% Fair 35% Poor 6%	Good 29% Fair 65% Poor 6%	Good 11% Fair 89% Poor 0%	Good 33% Fair 63% Poor 4%

It seems that the English and Education students were relatively less confident of their pronunciation skills, despite their probably future careers as English teachers; although proficiency levels are difficult to assess, it is possible that they actually had a slightly lower level of proficiency.

4.7 Survey design

The final survey design was the outcome of a series of preliminary and pilot surveys. The common aim of these surveys was to enable the respondents to rate the samples according to several criteria, and to allow them to reflect on and indicate the phonological features, if any, that influenced their ratings. A space for comments of an open-ended nature was also included on the survey form. As well as attempting to measure the possible relationship between acceptability ratings and phonological features, this approach allowed for the collection of both quantitative and qualitative data.

Early survey designs involved the use of clines (for example, with ‘Many errors’ at one end and ‘No errors’ at the other). This had the advantage of providing respondents with a visual comparison of the ratings awarded to different samples, but analysing the data was time-consuming in that physical measurements of marks on the cline had to be taken to obtain quantitative scores. The final design used a six-point Likert scale, to discourage neutral scores while allowing for greater discrimination than would be possible with a four-point scale. This was also the design chosen by Jenkins (2007) in her survey of attitudes.

4.7.1 The concept of acceptability revisited

In designing the survey form, it was necessary to deconstruct the term ‘acceptability’ and decide which of its components were to be included. Linguistic definitions of acceptability normally see it from the perspective of the native speaker and often relate it to grammaticality judgments. According to Crystal (2003: 4), acceptability is defined as ‘the extent to which linguistic data would be judged by native speakers to

be possible in their language’, although judgments may differ because of variations in regional and social backgrounds. In phonology, an opposing position is taken by Jenkins (2000: 224): ‘we are no longer concerned with the acceptability of ‘non-native’ speakers to ‘native’ hearers, rather than that of ‘non-native speakers’ to each other’. The desirability of allowing non-native listeners to evaluate their own accents is thus indicated, as is the need to allow them to do so without the material being prejudged in significant ways, for example by inserting perceived ‘errors’. It may after all be possible that non-native speaker perceptions differ significantly from those of native speakers in this area. The second part of the survey form was open-ended in that listeners could mark whichever parts of the transcript they considered to be significant; the first part involved a combined construct of acceptability.

4.7.2 Part 1 of the survey form

In the survey form, acceptability was measured through several questionnaire items relating to different components of the concept. First of all, the concept was framed by considering the nature of the educational context and the aims of the study. The respondents were students who had recent experience of English teaching in school, and a number of them may become English teachers in future. A relatively direct measurement of acceptability was therefore achieved by including a question asking them if they thought the speakers could be used as models for pronunciation teaching purposes in Hong Kong. The ultimate aim of this study was to make recommendations for pronunciation teaching syllabi, so some direct measure of ‘perceived acceptability for pedagogical purposes’ seemed to be important. The orientation of this question resembles one used by Bolton and Kwok (1990), who asked about the acceptability of accent samples for use in broadcasting.

Although it is preferable to separate constructs such as intelligibility and acceptability, it is also likely that an accent with a large number of phonological errors might become tiring or irritating to listen to, so respondents were also asked to rate the overall ‘error density’ of the sample and assess how easy it was to understand the speaker. Another statement required the respondents to consider whether they liked the way the speaker sounded, something which risked conflating

phonological features with paralinguistic or extralinguistic features, but which seemed important to an overall measurement of acceptability. Finally, in view of the importance of perceived prestige, respondents were asked to rate the speakers' level of education and job status. Another statement ('This speaker sounds like a Hong Kong person') was added, in order to assess the ways in which a 'Hong Kong accent', however perceived, related to other measurements.

The six questions, as they appeared in Part 1 of the survey form, are shown in Figure 4.2 below.

Figure 4.2. Part 1 of the survey form.

	Agree strongly						Disagree strongly
a The speaker sounds like a Hong Kong person.	☐	☐	☐	☐	☐	☐	
b This speaker has a lot of pronunciation errors.	☐	☐	☐	☐	☐	☐	
c This speaker is easy to understand.	☐	☐	☐	☐	☐	☐	
d I like the way this speaker sounds.	☐	☐	☐	☐	☐	☐	
e This speaker's accent is acceptable as a model for pronunciation teaching purposes in HK.	☐	☐	☐	☐	☐	☐	
f This speaker has a high level of education and / or a high status job.	☐	☐	☐	☐	☐	☐	

In the questionnaire used by Jenkins (2007) to investigate attitudes towards English accents, there were four components: correctness, acceptability, pleasantness and familiarity. It is uncertain how these concepts were actually interpreted by the respondents, but the survey design in the present study also uses these components as parts of an overarching, somewhat wider-ranging, construct of acceptability. Correctness is addressed by item B; acceptability (in the sense of 'pedagogical acceptability') by item E; pleasantness by item D; and familiarity by item A. The separation of correctness and acceptability was thought to be necessary by Jenkins

(2007: 153), in order to assess ‘whether respondents had any concept of ELF accents as more acceptable than NS accents in international communication contexts’. However, and as noted before, this depends entirely on how the accents are defined by the researchers and conceived of by the respondents. In the present study, the use of accent samples and precisely worded questionnaire items helped to ensure the construct validity of acceptability, and as in Jenkins’ study the findings will permit an assessment of whether correctness and acceptability are seen as separate or related dimensions.

Not all of these items were actually used in the subsequent analyses; the ‘direct’ acceptability measure (item E) and the average of items B to F (‘overall’ acceptability) were the main measures employed. Item A was arranged in a correlation matrix with the other items, but was not included in the later analyses of acceptability. It will be noted that from the rater’s perspective, item B has a different polarity (‘right equals good’) to the items below it (‘left equals good’). This was intended to act as a test of the integrity of the questionnaire.

4.7.3 Part 2 of the survey form

A two-part survey design thus evolved, in which the first part elicited overall ratings while the second part focused on the phonological features that had been noticed by the students. In early pilot studies a transcript was provided from the outset and respondents were asked to mark the important features. One disadvantage of this was that some students marked a great number of features, making it difficult to decide which ones were actually important. Another was that some marks or comments were difficult to interpret in that it was not always clear whether the feature noted had been a positive or a negative influence on the rating. In a trial run conducted with different students just before the actual survey, respondents were asked to focus on the features that had negatively influenced their ratings, and also to code the importance of the feature using a number. This approach was inspired by van den Doel’s (2006) internet survey form, but this part of the survey avoided the term ‘error’, using instead the more neutral ‘feature’. This design appeared to work satisfactorily, except for the fact that few students actually used the numbers to code

the features they marked. It seemed likely that the concentration required to listen to twelve accent samples and complete six questions for each was already considerable, and the final design employed a simpler marking system that tried to limit the number of marked features to three, and did not require any coding of relative importance. Part 2 of the survey form is reproduced below.

Figure 4.3. Part 2 of the survey form.

Part 2: listen again and then try to decide which words, sounds or other features were most important in helping you make the decisions you made in Part 1. You can refer to any of these areas:

<i>Vowel sounds (V)</i>	<i>Consonant sounds (C)</i>	<i>Consonant clusters (CC)</i>
<i>Word stress (WS)</i>	<i>Connected speech: sentence stress, rhythm, linking etc. (CS)</i>	
<i>Intonation (I)</i>		

For ‘negative’ features, please mark the transcript by underlining the relevant parts and using the above codes. For example, if you think there is a consonant problem in the word ‘supermarket’ you can mark it like this:

C ↘
 supermarket

NB: Please do not mark more than THREE features per speaker. Decide which features were most important in forming your impression.

If you do not think there are any errors, you do not need to mark anything. You may note ‘positive’ features and/or further explain your Part 1 answers in the space provided.

During Part 2 a transcript of the samples was visible to the students. The ‘space provided’ below each transcript allowed students to note positive features or make other open-ended comments, which were collected as part of the ‘student comments’ data. The students were verbally encouraged to do this and were shown an example of a completed form before the survey commenced. A full copy of the survey form is provided in Appendix 2.

4.8 Administering the survey

When administering the survey, the twelve accent samples were ordered differently for each group, to reduce the possibility of ordering effects. The survey was introduced in the same way each time: first, a brief presentation was given, reviewing the content of the course and reminding students of the categories used in the survey, and then an example of Part 2 of the questionnaire was shown, including an example of a student's coding from one of the pilot studies. I also explained questionnaire item E by asking students to imagine that they were choosing speakers for a course or book to be used for English teaching purposes in Hong Kong. The survey procedure took around 50 minutes in total. The accent samples were played a total of three times, once for Part 1 and twice for Part 2.

CHAPTER 5

DATA ANALYSIS AND FINDINGS

5.1 Introduction

This chapter presents the findings obtained from analyses of the 52 survey forms. These forms yielded three types of data. Part 1 provided quantitative acceptability ratings, and Part 2 both indicated the phonological features that influenced the Part 1 ratings and provided further qualitative data about the students' reactions to the accent samples. After reviewing the data for its overall consistency in terms of inter-rater reliability, and for the effects of listener factors such as gender, the Part 1 data is first presented in the form of acceptability rankings for the twelve speakers. The relative effects of the measured linguistic factors (accuracy, complexity, fluency and prosodic factors) are assessed using a multiple regression analysis. The student error codings in Part 2 provided the input for a second regression analysis, which measured the effects of different phonological features on the acceptability ratings. The chapter concludes with a speaker-by-speaker analysis that attempts to combine the various sources of data, including the acceptability ratings, the measured linguistic factors, and the qualitative comments from Part 2 of the survey form.

5.2 Part 1 data

The six questionnaire items are listed again below. The individual ratings for the items were assigned a value of between 6 (Agree strongly) and 1 (Disagree strongly). Item B was intentionally stated with a different polarity to the other items as a test of rater attention to the task, and the numerical scores for this item were inverted in order to compute the overall acceptability.

- A. The speaker sounds like a Hong Kong person.
- B. This speaker has a lot of pronunciation errors.
- C. This speaker is easy to understand.
- D. I like the way this speaker sounds.
- E. This speaker's accent is acceptable as a model for pronunciation teaching purposes in HK.
- F. This speaker has a high level of education and / or a high status job.

Of the six statements, item A (referring to the perceived degree of accentedness) was not considered to be central to the measurement of acceptability, as mentioned in Chapter 4. Accent strength is not seen as a component of acceptability in this study, partly because there is no *a priori* reason for an accent perceived as 'strong' to be unacceptable, but also because accent features were assessed via Part 2 of the survey form. However, the extent to which accent strength correlated with acceptability as a whole was deemed to be of interest, and item A was thus included in the inter-item correlation measurements. For most of the analyses, therefore, a measurement of 'overall acceptability' comprising the average scores for items B to F was used. The internal consistency of this measurement and the treatment of item A as peripheral both received some support from an initial statistical analysis of the Part 1 questionnaire data.

5.2.1 Internal consistency

The first approach to the questionnaire data was to assess its reliability in terms of internal consistency, or the extent to which the questionnaire items measure the same general attribute (Perry 2005: 134). One of the most common measures of internal consistency is provided by Cronbach's alpha coefficient. Submitting all six questionnaire items to an SPSS analysis produced the following output table (Table 5.1).

Table 5.1. SPSS output table for Cronbach's alpha coefficient (six items).

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
A	19.1619	23.7988	-.4435	.8486
B	19.8149	13.5158	.5234	.4948
C	19.3614	13.8832	.5739	.4868
D	20.3189	11.5105	.6532	.4154
E	20.2484	11.5225	.6702	.4091
F	19.3397	14.1252	.5503	.4972

Reliability Coefficients

N of Cases = 624.0

N of Items = 6

Alpha = .6118

All of the twelve speaker samples were submitted to this analysis, giving a total of 624 cases (52 raters and twelve samples). The overall alpha coefficient is 0.6118, somewhat lower than the ideal value of 0.7 (Pallant 2001: 85). However, the 'Alpha if item deleted' column shows that eliminating item A would significantly increase the alpha coefficient, presumably because it is negatively correlated with the other items. There is thus a clear statistical and conceptual basis for eliminating item A and creating an 'overall acceptability' measure from the five remaining items. The alpha coefficient of these items was 0.8486 (see Table 5.2, below). This comfortably exceeds the 0.7 threshold, and indicates that the questionnaire items are measuring what is essentially the same underlying construct.

Table 5.2. SPSS output table for Cronbach's alpha coefficient (five items).

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)				
Item-total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
B	15.3277	16.3687	.6044	.8314
C	14.8742	17.7622	.5339	.8474
D	15.8317	13.8232	.7698	.7853
E	15.7612	13.8370	.7880	.7794
F	14.8526	17.2479	.6107	.8305
Reliability Coefficients				
N of Cases =	624.0	N of Items = 5		
Alpha =	.8486			

5.2.2 Inter-rater reliability

The inter-rater reliability was measured by means of intraclass correlation measures. The intraclass correlation coefficient (ICC) is a measurement of the proportion of variance that is attributable to the objects of measurement (McGraw and Wong 1996: 30). Thus in an experimental situation with multiple raters, an ICC of 1 signifies perfect agreement amongst the raters. In such a case, all of the variance would be attributable to the objects of measurement (i.e. the test takers, or speakers), rather than the test raters or respondents. Using the overall acceptability scores (the average of items B to F) and a two-way random effects model under a consistency definition, the intraclass correlation coefficient was 0.8486. This is the expected result, as in classical test theory the ICC is an extrapolation of Cronbach's alpha (Feldt 1990, in Kaplan 2004: 79). Similarly, the minimum acceptable value of the intraclass correlation coefficient is also normally taken to be 0.7 (Bott and Radó 2007: 62). Thus there was a sufficiently high level of agreement between the 52 raters.

The questionnaire data relating to overall acceptability were also assessed for the effects of gender, major (English Studies, English Education or Other) and self-rated pronunciation skills (poor, fair or good). The adjusted intraclass correlation coefficients and their 95% confidence intervals are shown in Table 5.3 below.

Table 5.3. Intraclass correlation coefficients according to gender, major and self-rating of pronunciation skills.

Category	Group	Intraclass correlation coefficient*	95% confidence interval
Gender:	Female	0.837	0.814-0.858
	Male	0.759	0.642-0.845
Major:	English Studies	0.848	0.819-0.874
	English Education	0.818	0.781-0.850
	Other	0.799	0.626-0.904
Self-rating of pronunciation skills:	Poor	0.723	0.483-0.867
	Fair	0.821	0.791-0.848
	Good	0.850	0.813-0.881

*Two-way random effects model, average measures, consistency definition.

Although all of the ICCs are above the 0.7 threshold, there are some interesting differences between the groups. It is noteworthy that the pronunciation skills level of the students, as measured by their self-ratings, appeared to affect the consistency of the ratings. The lowest ICC (and thus the greatest variation in ratings) occurred within the ‘poor’ group, and the highest correlation occurred within the ‘good’ group. This suggests that attributes such as listening skills and language awareness were perhaps greater within the ‘good’ group, with the ‘poor’ group being less able to perceive differences or simply being unaware of the significance of phonological features. However, the fact that even the ‘poor’ group displayed an ICC of greater than 0.7 further indicates that the experimental design was capable of providing reliable data. Two possible objections to the use of non-native speaker, student raters in accent studies might be that they are not able to reliably hear differences, or that they react to these differences in a non-systematic manner. There is little or no support for either of these contentions in the reliability data, although the accuracy and systematicity of their error marking (as opposed to their scale rating) has not yet been considered.

5.2.3 Inter-item correlations

Table 5.4 below shows the Pearson correlation coefficients between the scores for the items on the survey form. Most of the correlation coefficients were found to be significant at the $p < 0.01$ level, using two-tailed tests of significance ($N = 624$).

Table 5.4. Inter-item Pearson correlation matrix.

		Correlations						
		A	B	C	D	E	F	AVEBF
A	Pearson Correl.	1	-.368(**)	-.079(*)	-.456(**)	-.464(**)	-.331(**)	-.444(**)
	Sig. (2-tailed)	.	.000	.048	.000	.000	.000	.000
	N	624	624	624	624	624	624	624
B	Pearson Correl.	-.368(**)	1	.369(**)	.531(**)	.591(**)	.459(**)	.751(**)
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000
	N	624	624	624	624	624	624	624
C	Pearson Correl.	-.079(*)	.369(**)	1	.502(**)	.481(**)	.410(**)	.683(**)
	Sig. (2-tailed)	.048	.000	.	.000	.000	.000	.000
	N	624	624	624	624	624	624	624
D	Pearson Correl.	-.456(**)	.531(**)	.502(**)	1	.788(**)	.553(**)	.874(**)
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000
	N	624	624	624	624	624	624	624
E	Pearson Correl.	-.464(**)	.591(**)	.481(**)	.788(**)	1	.553(**)	.883(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000
	N	624	624	624	624	624	624	624
F	Pearson Correl.	-.331(**)	.459(**)	.410(**)	.553(**)	.553(**)	1	.739(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000
	N	624	624	624	624	624	624	624
AVEBF	Pearson Correl.	-.444(**)	.751(**)	.683(**)	.874(**)	.883(**)	.739(**)	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.
	N	624	624	624	624	624	624	624

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

The overall acceptability scores (the average of B to F) and the ‘direct acceptability’ scores (item E only) were highly correlated ($r = 0.883$, $p < 0.01$). Although the latter measure has the advantage of simplicity, for some data analyses (such as regression) there are advantages in using a composite measure as this generates scalar, rather than ordinal, data. The reliability data has already shown that there is an acceptable level of consistency within this overall measure.

As can be seen from the table, the inter-item correlations with values of $r > 0.5$ were:

- items D and E (liking and acceptability as a teaching model, $r = 0.788$);
- items B and E (few errors and acceptability as a teaching model, $r = 0.591$);
- items D and F (liking and perceived status, $r = 0.553$);
- items E and F (acceptability and perceived status, $r = 0.553$);

- items B and D (few errors and liking, $r = 0.531$); and
- items C and D (ease of understanding and liking, $r = 0.502$).

The ‘direct acceptability’ measure (item E) correlated most strongly with ‘liking’, with the items ranked as follows:

1. D (liking) $r = 0.788$
2. B (few errors) $r = 0.591$
3. F (perceived status) $r = 0.553$
4. C (ease of understanding) $r = 0.481$
5. A (sounds like a Hong Kong person) $r = -0.464$ (negative correlation)

These rankings give some indication of the nature of the acceptability constructs used in this study. In general, ‘acceptability as a teaching model’ appears to be most strongly correlated with ‘liking’ and ‘few errors’. It is not immediately clear why ‘liking’ should show a higher correlation with pedagogical acceptability than the apparently more functional items of ‘few errors’ and ‘ease of understanding’. An initial explanation might be that ‘liking’ is something of a cover term that actually overlaps with pedagogical acceptability, so that the students may have interpreted it as meaning ‘I would like to have this as a teaching model’. The above differences also indicate varying levels of agreement over some items, so that there was less consensus regarding ‘ease of understanding’ than there was with ‘liking’, for example. An initial finding is that there is no evidence for the potential separation of ‘correctness’ (item B) and ‘acceptability’ (item E). The study of attitudes towards ELF accents carried out by Jenkins (2007) also found these aspects to be closely related in questionnaire responses. However, it should be borne in mind that in the present study item E refers to pedagogical acceptability, and is thus likely to be rated more strictly than in other senses (for example, that of acceptability for international communication).

Looking at the inter-item correlations as a whole, it can be seen that item A (‘sounds like a Hong Kong person’) correlated negatively with all of the other items, suggesting that perceived accent strength had a negative effect on acceptability.

However, it is difficult to assess how the students interpreted ‘sounds like a Hong Kong person’, and there is a need for a more detailed, features-based investigation in order to identify some of the phonological components of perceived accent strength. This was performed by means of a regression analysis, which is explained in section 5.3. Section 5.2.4 below presents the acceptability data in more detail.

5.2.4 The acceptability ratings

Figure 5.1 below shows in graphical form the overall acceptability scores for the three groups; the high intraclass correlation can be gauged visually from the closeness of the three lines.

Figure 5.1. Overall acceptability scores for the twelve speakers and three groups of raters.

In Table 5.5 below, the twelve speakers are ranked according to both their overall acceptability scores and the direct acceptability scores.

Table 5.5. The twelve speakers ranked by overall acceptability scores (ranking 1) and direct acceptability scores (ranking 2).

<i>Ranking 1</i>	<i>Speaker No.</i>	<i>Overall acceptability score (average of B-F)</i>	<i>Ranking 2</i>	<i>Speaker No.</i>	<i>Direct acceptability score (E only)</i>
1st	10	4.92	1st	10	4.91
2nd	11	4.48	2nd	11	4.50
3rd	8	4.45	3rd	8	4.16
4th	3	4.15	4th	3	3.97
5th	4	4.09	5th	4	3.78
6th	2	3.85	6th	2	3.28
7th	9	3.52	7th	9	3.11
8th	5	3.51	8th	7	2.81
9th	7	3.43	9th	5	2.73
10th	6	3.29	10th	6	2.67
11th	1	3.20	11th	1	2.53
12th	12	3.18	12th	12	2.50

The overall rankings using the two measurements are almost the same, with only one transposition (Speakers 5 and 7 are in a different order). The wide range of average scores for ‘direct acceptability’ (between 2.50 and 4.91) suggests that the accent samples were quite diverse, and that the acceptability levels of these samples vary widely. It is noticeable that while the acceptability scores for the highest-ranked speakers are almost the same, regardless of which measurement is used, the direct acceptability scores decline more rapidly than the overall acceptability scores as one moves down the rankings. This relationship is shown in Figure 5.2 (note that here, the numbers on the horizontal axis apply to the speakers’ positions in the rankings, rather than to the speaker numbers).

Figure 5.2. Acceptability scores using the two measurements of ‘overall’ acceptability and ‘direct’ acceptability.

This suggests that pedagogical acceptability (as measured by item E) was given lower scores than some of the other questionnaire items, and that this tendency was more pronounced among those speakers given relatively lower overall ratings. Unfortunately there are no absolute measures of acceptability, but the presence of a native speaker (Speaker 11) serves as a comparison. However, while these data will be relevant when considering research question 1 (regarding pedagogical acceptability), the main aim of the subsequent statistical analyses is to investigate the effects of phonological features on the acceptability scores. In line with the features-based orientation of the study, this will enable the discussion to move away from an exclusive concern with speakers, and towards a consideration of some of the properties of their speech.

5.3 The effects of speaker variables

The above analyses have helped to assess the construct validity of ‘acceptability’, and the general nature of the acceptability ratings has begun to emerge. In the next stage, speaker variables in the categories of accuracy, complexity, speech rate and prosodic factors (pitch span) were assessed in terms of their interrelationships and

their effects on the acceptability scores. The overall purpose of this was to establish the relative importance of these variables in determining the acceptability scores.

5.3.1 Phonological accuracy

While the measurements for complexity, speech rate and prosodic factors were presented in Chapter 4, the measurements of phonological accuracy were obtained from an analysis of the survey forms and are introduced in this section. As Bolton and Nelson (2000: 259) observe, the use of such terms as ‘error’, ‘interlanguage’ and ‘feature’ is contentious. In describing the work of the International Corpus of English (ICE) project, Bolton and Nelson state that the approach is ‘features-based’ rather than employing the concepts of ‘error analysis’ or ‘interlanguage’. This study has generally preferred the term ‘feature’, and this will continue to be employed where possible. The measurement of ‘accuracy’ in this study was derived from student codings, in order to avoid prejudging the nature of ‘accuracy’. However, in Part 2 of the survey form the listeners were instructed to mark the features that negatively affected their acceptability ratings. Hence it seems reasonable to classify these as ‘errors’ for the purposes of statistical analysis, although subsequent analyses will attempt to distinguish between types of error in terms of their effects on the acceptability ratings. The term ‘phonological accuracy’ as used in this study thus refers to the number of ‘errors’, so defined, as marked by student listeners.

The total number of errors marked on the survey forms was tallied for each speaker and used to obtain the measures of ‘number of errors’, ‘errors per word’ and ‘errors per syllable’ in Table 5.6 below. It must be pointed out that these scores are raw accuracy measures based on unanalysed listener codings, and they may contain non-existent errors. Also, the figures for ‘errors per word’ and ‘errors per syllable’ in Table 5.6 are based upon the total number of errors marked, including those marked more than once. They do not therefore represent the number of discrete errors per word or syllable. However, for this part of the study a relative measure of accuracy appears to be adequate. In later analyses, the student codings will be examined in more detail in order to ensure that they represent actual instances of phonological deviation within the various categories of error.

Table 5.6. Measures of phonological accuracy.

Speaker No.	No. of words	No. of syllables	Total no. of errors marked	Errors per word	Errors per syllable
1	24	31	104	4.33	3.35
2	28	48	74	2.64	1.54
3	37	58	67	1.81	1.16
4	29	38	67	2.31	1.76
5	25	44	101	4.04	2.30
6	27	42	82	3.04	1.95
7	23	38	95	4.13	2.50
8	18	22	52	2.89	2.36
9	19	33	80	4.21	2.42
10	30	44	29	0.97	0.66
11	20	35	45	2.25	1.29
12	15	23	112	7.47	4.87
<i>Ave.</i>	<i>24.58</i>	<i>38</i>	<i>75.67</i>	<i>3.34</i>	<i>2.18</i>

5.3.2 Correlations between the variables

The analyses of the speech samples performed in Chapter 4, plus the above computation of the number of ‘errors’ marked on Part 2 of the survey forms, yielded a total of thirteen measurements pertaining to the categories of accuracy, fluency, complexity and prosodic factors. For the category of prosodic factors, only the measurements from the ten male speakers were included in order to provide comparable data. These measurements are summarised in Table 5.7 below, which also gives the abbreviations used in statistical analyses.

Table 5.7. The thirteen speaker variables selected for analysis.

Category	Abbreviation	Explanation / measurement
Phonological accuracy (derived from student error codings)	ERRPW	Errors per word
	ERRPSYLL	Errors per syllable
	NOERR	Number of errors
Fluency	WPM	Speech rate in words per minute
	SYLLPM	Speech rate in syllables per minute
Complexity: lexical syntactic	SYLLPW	Average number of syllables per word
	LEXCOMP	Lexical complexity
	WTU	Average number of words per T-unit
	CTU	Average number of clauses per T-unit
Prosodic factors	MINFREQ	Speaker’s minimum voice pitch during utterance
	MAXFREQ	Speaker’s maximum voice pitch during utterance
	SPAN	Pitch span (i.e. MAXFREQ – MINFREQ)
Other	LENGTH	Length of utterance in seconds

First of all, Pearson correlation coefficients were calculated for these variables in order to investigate their interrelationships. The detailed correlation matrix is shown in Appendix 3. The high correlations between factors that are obviously related, such as words per minute and syllables per minute, are visible in the matrix. These intercorrelations will be examined and dealt with during the next stage of the analysis.

5.3.3 Regression analysis

The next stage of the data analysis was to examine the relative effects of the speaker variables on the acceptability ratings. To do this, a multivariate statistical approach (multiple regression) was employed. As this method was used in both of the two main statistical analyses in this study, some of its general characteristics and assumptions will first be explained in this section.

Within the ‘quantitative paradigm’ in sociolinguistics the general value of multivariate statistical analysis has been acknowledged by, for example, Bayley (2002). Regression analysis has not been used particularly often, however, an exception being the use of logistic regression to investigate the effects of various independent variables on a binary dependent variable such as feature use or non-use. To my knowledge the present study is the first to use regression analysis to investigate the relationships between several dependent variables (categories of phonological error) and a numerical dependent variable (acceptability scores).

The basic aim of such analyses is to explore how well a set of variables is able to predict a particular outcome (Pallant 2001: 134). Consequently, the nature of both the independent and dependent variables, as well as the methods used to measure them, should be carefully considered. An important point about regression analysis is that there is no implication of causal relationships in regression equations. As Tabachnick and Fidell (1996: 130) note, an apparently strong relationship between variables ‘could stem from many sources, including the influence of other, currently unmeasured, variables’.

The practical considerations for regression analysis are in four main areas: sample size, multicollinearity, the presence of outliers and certain aspects of distribution and variation such as normality and linearity (Pallant 2001: 136). As a rule of thumb for calculating the minimum sample size for testing individual predictors, Tabachnick and Fidell (1996: 132) provide a formula of $N \geq 104 + m$, where m is the number of independent variables. Assuming that all the thirteen variables in Table 5.7 were to be employed in the analysis, a sample size of at least 117 cases would be required, and this is comfortably exceeded by the 520 cases (52 raters, 10 speakers) in this part of the study.

Multicollinearity refers to a situation where the independent variables are highly intercorrelated ($r \geq 0.9$, according to Pallant 2001: 136). To deal with the intercorrelations observed in the correlation matrix (Appendix 3), speaker variables showing high intercorrelations were combined before being used in the regression analysis. Pallant (2001), citing Tabachnick and Fidell (1996: 86), suggests that variables with a bivariate correlation of 0.7 or above should not be included in regression analyses. There are thus two choices: to eliminate one member of an intercorrelated pair, or to form a composite variable from the two variables.

An analysis of the intercorrelations and a process of combination yielded a list of six variables. The intercorrelated variables ERRPW and ERRPSYLL were averaged to produce a new variable, ERROR. The total number of errors within an utterance (NOERR) was excluded from the ERROR category, partly because it represents a non-relative measurement and also because it correlated strongly with both ERRPW and ERRPSYLL. To measure lexical complexity, the variable LEXCOMP was retained. The variable SYLLPW was excluded because it has a strong negative correlation with WPM; multisyllabic words tend to reduce speech rate as measured by WPM. Syntactic complexity (SYNCOMP) was measured by the average of CTU and WTU. In similar fashion, a SPEED variable was created by averaging SYLLPM and WPM, and an INTON (for 'intonation' variable was formed by averaging MAXFREQ and SPAN. This left two 'conceptually independent' variables of MINFREQ and LENGTH; it was decided to retain MINFREQ but to exclude LENGTH, on the grounds that there were moderate (though non-significant)

correlations with variables such as ERRPSYLL and LEXCOMP. The six resultant variables were thus:

1. ERROR (the average of ERRPW and ERRPSYLL)
2. LEXCOMP
3. SYNCOMP (the average of CTU and WTU)
4. SPEED (the average of SYLLPM and WPM)
5. INTON (the average of MAXFREQ and SPAN)
6. MINFREQ

The six variables represent the four categories of accuracy (actually inaccuracy, or ERROR), complexity (LEXCOMP and SYNCOMP), speed (SPEED) and prosodic factors (INTON), together with the remaining non-correlated variable, MINFREQ. The correlation matrix for these six variables is shown in Table 5.8 below.

Table 5.8. Correlation matrix for the six speaker variables.

Correlations

		ERROR	LEXCOMP	SYNCOMP	SPEED	INTON	MINFREQ
ERROR	Pearson Correl.	1	.069	.124(*)	-.508(**)	.480(**)	-.234(**)
	Sig. (2-tailed)	.	.158	.012	.000	.000	.000
	N	416	416	416	416	416	416
LEXCOMP	Pearson Correl.	.069	1	.069	-.167(**)	.057	.100(*)
	Sig. (2-tailed)	.158	.	.158	.001	.242	.042
	N	416	416	416	416	416	416
SYNCOMP	Pearson Correl.	.124(*)	.069	1	-.270(**)	.106(*)	.137(**)
	Sig. (2-tailed)	.012	.158	.	.000	.030	.005
	N	416	416	416	416	416	416
SPEED	Pearson Correl.	-.508(**)	-.167(**)	-.270(**)	1	-.495(**)	-.444(**)
	Sig. (2-tailed)	.000	.001	.000	.	.000	.000
	N	416	416	416	416	416	416
INTON	Pearson Correl.	.480(**)	.057	.106(*)	-.495(**)	1	-.233(**)
	Sig. (2-tailed)	.000	.242	.030	.000	.	.000
	N	416	416	416	416	416	416
MINFREQ	Pearson Correl.	-.234(**)	.100(*)	.137(**)	-.444(**)	-.233(**)	1
	Sig. (2-tailed)	.000	.042	.005	.000	.000	.
	N	416	416	416	416	416	416

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Although there are statistically significant intercorrelations in the table, none approach the $r > 0.7$ threshold and there appears to be little threat of multicollinearity effects. An initial run of the regression procedure produced a scatterplot that showed a normal distribution of residuals; in other words, the regression equation represented a line of best fit around which the ‘errors’, or residuals, were distributed in the expected fashion. There were no outliers, defined as cases with standardised residuals of less than -3.3 by Tabachnick and Fidell (1996: 139). Other aspects, such as linearity and the independence of residuals, did not appear to present any problems. Linearity means that a one-unit increase in the value of an independent variable is associated with a one-unit increase in the value of the dependent variable, regardless of whether the increase is at the lower or higher end of the range of values (Cohen 2003: 194). The ‘independence of residuals’ assumption refers to the necessity for residuals to be independent of each other, without any clustering effects caused by, for example, certain methods of measurement (Cohen 2003: 120).

The overall viability of the regression procedure was thus indicated. First of all, a linear regression analysis using the six speaker variables as the independent variables and overall acceptability as the dependent variable was employed. In this procedure, only eight of the original twelve speakers were included. As mentioned above, the two female speakers could not be included because their pitch range measurements were significantly different. A decision was made at this point to exclude Speaker 5. It was felt that the inclusion of the error relating to the missing plural ‘s’ in the word *applications* might risk conflating grammatical and phonological error. Although there were other errors that might be classed as grammatical, because this speaker appeared to be reading from a script there was a possibility that the error occurred during the preparation, rather than the delivery, of the speech. In addition, this was the only example of plural marking being affected by final CCR in the data. Speaker 11, the native speaker, was also excluded from this analysis in order to focus on the Hong Kong speakers. The SPSS output tables showing the model summaries and the coefficients are shown in Table 5.9 below.

Table 5.9. Model summary and coefficients for regression analysis of the six speaker variables.

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.500(a)	.250	.239	.85680	1.681

a Predictors: (Constant), SYNCOMP, INTON, LEXCOMP, ERROR, MINFREQ, SPEED

b Dependent Variable: AVEBF

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	5.851	1.260		4.644	.000		
	ERROR	-.052	.007	-.481	-7.707	.000	.471	2.125
	INTON	-.003	.002	-.067	-1.065	.287	.462	2.166
	SPEED	.000	.003	.004	.048	.962	.248	4.040
	LEXCOMP	-.014	.007	-.125	-2.206	.028	.574	1.743
	MINFREQ	.001	.006	.018	.221	.825	.282	3.548
	SYNCOMP	-.022	.011	-.094	-1.973	.049	.803	1.245

a Dependent Variable: AVEBF

The model summary table shows the extent to which the variance in the dependent variable is explained by the regression model. The adjusted R square value is 0.239, meaning that 23.9% of the variance in the acceptability scores is explained by the selected independent variables. Adjusted R square values of between 0.11 and 0.30 indicate a model with a ‘modest fit’, according to Cohen *et al.* (2007: 538). Nevertheless, given the authentic nature of the samples and the potential for extralinguistic factors to affect the results, the results suggest that tentative conclusions about the relative effects of the different factors can be drawn. Confirmation of the independence of residuals is given by the Durbin-Watson statistic of 1.681; Savin and White (1977) give lower and upper bounds of 1.613 and 1.735 for a data set of these dimensions.

Turning to the coefficients, in the ‘Sig.’ (significance) column there is only one statistically significant independent variable at the $p < 0.0001$ level, shown in SPSS by a ‘.000’ value. This variable was ERROR, with a standardised beta coefficient of -0.481 . Standardised beta coefficients and significance values can be used to

compare the contributions of independent variables, and they show that ERROR makes a significant and unique contribution to the equation that predicts the acceptability scores. In other words, accuracy was more important than any other measured category. As expected, the coefficient is negative (as phonological errors increase, acceptability scores decrease). The unstandardised beta value of -0.052 can be interpreted as meaning that for every unit of increase in ERROR, acceptability ratings decrease by 0.052 points on the six-point scale. The overall conclusion of this stage of analysis is that amongst all the variables selected for measurement, accuracy, or perceptions of phonological ‘error’, played the most important role in influencing acceptability ratings. This was also suggested by the inter-item correlations listed in section 5.2.3, where item B (relating to accuracy) correlated relatively strongly with both direct and overall acceptability. The next most important variables appeared to be lexical and syntactic complexity, although neither achieved statistical significance.

As with any regression analysis, the possible effects of other, unmeasured variables should also be considered. The factors above measure only a small part of the spectrum of variables; other prosodic factors such as rhythm were not measured. There is also the possibility that variables may interact in unknown ways, and regression analysis may obscure combined or individual effects. It could be argued that lexical and syntactic complexity should be seen as belonging to an overall factor of ‘complexity’, rather than being two separate factors (although Table 5.8 shows that they are not correlated with each other). In addition, there is also the strong possibility that some errors classed as ‘phonological’, such as consonant cluster reduction, also have a grammatical component. The greater effects of phonological error are also likely to be due to this being the only factor directly assessed by the students, and to the survey method asking them to make some kind of connection between phonological features and acceptability scores. But working within these limitations, the effects of phonological error on the acceptability scores were always greater, no matter how the variables were combined and entered into the analysis.

5.4 Student error codings

It is not particularly surprising that students should prefer samples with relatively fewer errors. The aim of the next part of the data analysis was therefore to investigate which phonological features or ‘errors’ affected the acceptability ratings the most. A linear regression procedure was again employed, in order to compare the effects of different types of error. The data came from Part 2 of the survey form, in which the students were asked to listen to the accent samples again and mark the transcripts to show the features that influenced their acceptability judgments. The students were instructed to confine their codings to ‘negative’ features. In Chapter 4 it was noted that the term ‘error’ will be used temporarily to denote all student error codings, regardless of their effects on the acceptability ratings. At a later stage of the study, the possible differences between ‘error’ and ‘feature’ will be clarified.

Investigating the relationship between these error codings and the acceptability scores therefore involves the assumption that the students were able to make a connection between their impressionistic and relativistic assessment in Part 1 and perceived features of the speaker’s pronunciation in Part 2. Although this process may have been conceived of in different ways by different students, both the high intraclass correlations and the significant observed effects of ‘error’ on the ratings suggest that this was a reasonable assumption. The students were asked to mark no more than three features, although some marked more; in these cases the features were still entered into the analysis. These assumptions and outcomes do not seem to threaten the general construct validity of ‘accuracy’ in the methodological approach.

5.4.1 Error codings and categories

The aim of the next regression analysis was to assess the effects of different categories of error on the acceptability scores. To avoid making presuppositions about the types of error that existed in the samples, the students were given the option to mark several kinds of features in Part 2. However, a consequence of this ‘open-ended’ approach was that decisions had to be made about which codings to include, and in which categories they should be placed. Some student error codings

were excluded from the analysis. Part 2 of the survey form included codes for intonation and connected speech, but these were excluded because it was often unclear what the students were marking at a suprasegmental level, or how they interpreted these terms; some of them appeared to conflate intonation with sentence stress, for example. All other error categories were included in this part of the analysis. A full list of the error categories used in this study is provided in Table 5.10 (section 5.4.2). As has been mentioned, Speakers 5 and 11 were excluded from the error analyses. It was felt that Speaker 5's sample contained an error which may have been perceived as more grammatical than phonological. Speaker 11, the British native speaker, was excluded in order to focus on the effects of the Hong Kong accent features.

The overall aims of the error coding and analysis procedure were:

1. to achieve satisfactory construct validity for the categories, so that they were internally consistent and represented distinct phonological features;
2. to remain faithful to student codings, where possible, while ensuring an acceptable level of accuracy;
3. to use categories that are pedagogically relevant, wherever possible; and
4. to avoid an unnecessary proliferation of categories, which might conflict with 3) and reduce the scope for meaningful statistical analysis.

The achievement of construct validity involved both 1) and 2). The use of phonologically meaningful categories, the consistent assignment of individual codings to these categories and the accuracy of these codings were all important considerations. In initial analyses, codings were assigned to the basic categories given in part 2 of the survey form. At a segmental level the error codings were usually unambiguous, but on some occasions it was necessary to make inferences about the type of error involved. For example, if a student had underlined the first part of the word *these* without specifying the type of error using the expected 'C' code, it was usually assumed that this represented an instance of TH stopping. The likelihood of such assumptions being made was increased if other students had marked the same error in a more specific way. On the other hand, some codings were

excluded from the analysis because of irrecoverable ambiguity. For example, some students marked a ‘V’ (vowel) error by underlining the word *cards* thus:

cards

In this case it is possible that the perceived ‘error’ arose because of the speaker’s non-rhotic accent, and that some students expected to hear an ‘r’ sound (the spelling may also have influenced perceptions). However, the vast majority of codings were unproblematic. The occasional lack of clarity about the type of error reflects a weakness of the data collection instrument, but this was an inevitable consequence of not controlling in advance for the type of ‘errors’ or ‘features’ that existed; such an approach would have contravened the student-centred, non-judgmental approach taken by this study. A consideration of alternative methods suggests that there will almost always be problems in interpreting the intentions behind codings. An online questionnaire with categories developed during previous analyses might be a worthwhile avenue to explore, but even then there may be uncertainties about how the categories are actually interpreted and applied by the listeners.

After the initial analysis, the categories and the criteria for inclusion or exclusion were further developed. The choice of categories was clearly very important to the study. Having too many categories would tend to make the data unwieldy and possibly introduce unwanted intercorrelations, thus reducing the chance of identifying significant categories. Equally, having too few categories could lead to inaccurate attributions of influence, or missed identifications of important features. As mentioned above, it was desirable to use categories with pedagogical relevance wherever possible. These include the features identified as inconsequential for intelligibility (Jenkins 2000), as they may be strong candidates for removal from teaching syllabi; as has been mentioned, He and Li (2009) make a case for the acceptance of dental fricative substitutions in China English. Other relevant features and categories include L vocalisation and final cluster reduction (final CCR), although this category is somewhat problematic because it occurs very widely and is conditioned by the surrounding phonological environment, among other factors. It was also decided to introduce new categories into the analysis whenever assignment

to an existing category seemed problematic, and where there were significant numbers of codings. Student comments sometimes helped to interpret the intention behind the codings; an example is provided by the word *political* in Speaker 6's sample, where noticeable vowel reduction was accompanied by comments relating to speech rate. Codings, comments and repeated listening led to the creation of a new category SYLL (for 'syllabic modification', representing excessive vowel reduction or consonantal modification as a result of a rapid speech rate). The final list of error categories is explained in more detail in section 5.4.2 below.

When the categories had been established with clear criteria for inclusion and lists of tokens for each had been prepared, the codings were analysed again to establish consistency. Before doing so the codings were also assessed for accuracy to confirm that they did represent actual instances of the relevant feature category. All word tokens assigned to error categories were reviewed for their accuracy by listening to the relevant sample, in order to exclude doubtful codings. Particular attention was given to the most frequently occurring word tokens, as these would constitute the essential nature of the category in subsequent statistical analyses. Equally, word tokens with only one report were scrutinised and, in some cases, removed. The accuracy of the codings was thus verified partly through having multiple raters, and partly through the checking of both frequently-occurring and infrequently-occurring tokens. The number of instances of removal varied according to the phonetic similarity of the sounds concerned. For example, no word tokens representing /v/ or /r/ substitution were excluded, but several tokens of TH-fronting were rejected as not being sufficiently clear examples. It may be the case that seeing 'th' spellings, and being aware of the difficult nature of these sounds, prompted students to over-report cases of dental fricative substitution. In general, tokens representing intermediate forms were excluded, in order to maximise the internal consistency and construct validity of the error categories. The initial analysis of the codings showed nine word tokens identified as containing TH fronting, but in the end only one (the word *forthcoming*, used by Speaker 6), was retained. However, in the case of TH stopping twelve of the original eighteen identified word tokens were retained. The actual features of the speakers' utterances are discussed in more detail in section 5.5, and this discussion will further assess the accuracy of the student error codings.

A possible problem with this error checking procedure is that the acoustic conditions may have been different between the classroom listening, where a PA system was used, and the verification listening, which involved headphones. It is therefore possible that some error codings were excluded when they did in fact represent tokens of the error categories. However, this would seem to be preferable to including doubtful tokens. In general, the coding and analysis procedures showed that the great majority of the individual codings could be satisfactorily assigned to clearly-defined categories, and the accuracy checks supported the construct validity of these categories. The error categories are explained and illustrated in the following sections.

5.4.2 Error categories

The error categories used in subsequent analyses are shown in Table 5.10 below, which also lists the most frequently occurring tokens of each category. The underlined portions of words indicate the segments or syllables that were underlined most frequently. Words that are completely underlined indicate a predominance of non-specific markings for these word tokens. The VOWEL and PHONSUB categories were divided into two subcategories, in case these subcategories exerted significant independent effects on the acceptability scores.

Table 5.10. Error categories used in the study.

Category code	Description of category and subcategories	Frequently occurring, confirmed tokens (speaker numbers in parentheses)	Total number of identifications by students	
			<i>Total</i>	<i>Sub-total</i>
SYLL	Syllabic modifications, or alterations to syllable structure, probably a result of excessive vowel reduction linked to rapid speech (usually non-specific marking, but often accompanied by comments about clarity and/or speed)	<i>political</i> (6) <i>accredited</i> (9) <i>leadership</i> (6)	57	
VOWEL	Vowel modifications (marked V)		37	
	a) VOWEL SUB: vowel substitutions, possibly transfer influenced b) FULL VOWEL: use of a full vowel (non-reduction) in unstressed syllables	<i>want</i> (7), <i>maintain</i> (4), <i>leadership</i> (6) <i>production</i> (9), <i>confused</i> (12), <i>standards</i> (9)	27	10
PHONSUB	Consonantal substitutions, probably transfer-related		47	
	a) PHONSUB-V: /v/ substitution b) PHONSUB-R: /r/ substitution	<i>advantage</i> (1) <i>reason</i> (6)	30	17
TH-STOP	Substitution of /ð/ with [d]	<i>that</i> (1) <i>the</i> (6)	30	
TH-FRONT	Substitution of /θ/ with [f]	<i>forthcoming</i> (6)	4	
L-VOCAL	The vocalisation or deletion of postvocalic /l/	<i>people</i> (4)	2	
OTHER C-SUB	a) devoicing of final consonants or consonant clusters in plurals or verbs	<i>cards</i> (1), <i>aims</i> (9)	18	
	b) devoicing of final consonants	<i>have</i> (1), <i>because</i> (1)	2	
	c) C substitution	<i>built</i> (pronounced as [d]) (2), <i>department</i> (marked as sounding like [b]) (7), <i>department</i> (glottalised)	31	
	d) C deletion	<i>continued</i> (2), <i>accredited</i> (9), <i>have</i> (10)	3	
			54	
CCRF	Final consonant cluster reduction		237	
	a) CCRF-PV: in prevocalic or prepausal position	<i>confused</i> (12) <i>relaxed</i> (3) <i>privileged</i> (8)	113	
	b) CCRF-PC: in preconsonantal position	<i>found</i> (12) <i>suggests</i> (4)	124	
CS-LINK	Linking phenomena in connected speech	<i>by it</i> (3)	24	
		(Total)	493	

Figure 5.3 below shows the distribution of identified errors across all error categories.

Figure 5.3. The distribution of identified errors across error categories, ranked according to frequency.

From an overall perspective, the frequencies of occurrence shown in Figure 5.3 are somewhat different to those represented by the implicational scales in Chapters 3 and 4. For example, L vocalisation is much less frequent in the above ranking. However, this is to be expected: the implicational scale deals with the presence or absence of features, and the listeners' aim was to establish this by repeated listening. Figure 5.3 includes only those features that were actually marked, and the student listeners were influenced by the frequency of occurrence of features, the relative noticeability of the features, their relative importance, or some combination of these. Low frequency of occurrence, or low noticeability, may account for the infrequent marking of some features, such as L vocalisation. There is also the possibility that some features were noticed by the students, but were not actually noted on the survey forms because of perceived non-severity. The uncertain effects of frequency of occurrence, noticeability and severity present something of a methodological problem for the study, and this will be considered in the following sections and in later chapters.

The feature categories are further explained in the following sub-sections. The most frequently occurring tokens identified in each category are discussed in more detail, and bar charts showing the overall frequencies of occurrence of tokens are also

provided. In these sections it can be assumed that the student identifications were confirmed by the subsequent accuracy checks, unless stated otherwise. There were relatively few instances of missed identifications, where errors went unmarked by the students, and significant instances are discussed in the individual speaker analyses in section 5.5.

5.4.2.1 *Vowel-related errors: SYLL, VOWEL*

The analysis of the codings suggested that an additional category, that of ‘syllabic modification’ or SYLL, was needed. One of the most frequent error codings was derived from the underlining of a word or a part of a word with no specific indication of the type of error. Many of these related to a particular word (*political*) used by Speaker 6, in which there is an absorbed vowel in the first syllable. The speaker pronounces this word so rapidly that the vowels are very considerably reduced or even absorbed into neighbouring sounds. A possible transcription would be [p^hlɪtəkl̩] if the initial syllable is analysed as containing a voiceless vowel (Shockey 2003: 26). A syllabic consonant in the final syllable would be expected with many speakers, but the syllabic modification represented by the loss of a vowel in the first syllable seems to have triggered many error codings.

Even though schwa absorption (see Wells 1982: 434) is a common connected speech process in conversational English (Shockey 2003: 22), the students evidently thought there was something ‘wrong’ with this word. Their comments help to pinpoint the reasons for this perception; one wrote that ‘the word “political” is very unclear’, and two others appended the comments ‘confusing’ and ‘could be slower’ to this particular word. Some students attributed the problem to an excessively rapid speech rate (general comments included about this speaker included ‘speaks too fast, pronunciation sounds strange, not clear enough’; ‘he speaks too fast’; ‘some words are too fast’; and ‘bad pronunciation in key words’). Student comments are presented in full in the speaker-by-speaker analysis in section 5.5.

While the lack of a vowel in the first syllable might suggest a VOWEL coding, the word *political* seems to illustrate a more specific conceptual category involving a departure from the expected syllabic structure of a word. The other frequently-

marked word token was *accredited* (Speaker 9). Again, the codings seemed to be triggered by an overly rapid delivery of this word, and student comments also suggested this. One student supplied a ‘pronunciation spelling’ of *agreded* to represent the speaker’s production of this word, neatly representing the nature of SYLL errors as syllable structure modifications. General comments about this speaker included ‘the consonants are not accurate’, and in this case it appears that consonant modification, as well as vowel reduction, was also a trigger for the error codings. The speaker’s actual pronunciation could be transcribed as [ɛ'krədədəd], but there is considerable reduction of schwa in the antepenultimate and penultimate syllables (as in the *political* example). These and other features can also be explained as being due to ‘overly rapid’ production, or less judgmentally as being too far towards the ‘hypo-’ end of the hypo- to hyper-articulated speech continuum. For example, the onset of the final syllable is here produced as [d], rather than [t], perhaps as a result of voice assimilation in intervocalic position. This is also a feature of many NS, especially North American, accents. SYLL errors may have a ‘deviation from spelling’ aspect in terms of their perception, and this in turn may be related to listener unfamiliarity with certain accent features.

VOWEL errors can be generally defined as vowel substitutions, or realisations that differ from the expected parameters. The actual qualities of these vowels do not really matter here, as it is the fact of deviation (noted by students) that is of interest. The first subcategory of VOWEL was termed VOWEL SUB, which includes all vowel substitutions apart from the non-reduction of full vowels in unstressed syllables (this subcategory was called FULL VOWEL). The most frequently identified token of VOWEL SUB was *want* (Speaker 7), where an interesting example of feature co-occurrence can be heard. The speaker elides the final /t/ of the word, but this instance of final cluster reduction would not be particularly unusual in this inter-consonantal position (i.e., between the /n/ of *want* and the /k/ of the following ‘word’, *KCR*). However, under the influence of the velar plosive /k/ there appears to be some assimilatory velarisation of /n/ towards a velar nasal [ŋ], and this is further accompanied by a nasalised vowel in *want*. The other most frequently identified instance of VOWEL error was in the word *maintain* (Speaker 4), where there is shortening and only weak diphthongisation in the first syllable; in the general

comments, one student gave *maintain* (without the first 'i') as a possible pronunciation spelling. Another example of VOWEL SUB occurs in the first syllable of *leadership* (Speaker 6), where vowel shortening is apparent. This is an example of the merger of KIT and FLEECE (/ɪ/ and /i:/) in HKE.

Another type of vowel substitution, termed FULL VOWEL in this study, occurred in words such as *production* (Speaker 9) and *confused* (Speaker 12). Here, full (or non-reduced) vowels were used in the initial, unstressed syllables, instead of being reduced to schwa. This commonly occurs in both HKE (see, for example, Bolton and Kwok 1990: 152) and many other NVEs. The case of *democracy* (Speaker 12), where an [ɛ] vowel is used in the first syllable, may also appear to be an instance of FULL VOWEL. However, it was assigned to the VOWEL SUB category because there is also the likelihood of a full vowel [ɪ] in some varieties, including RP. The pedagogical interest of these substitutions and their possible differences from other types of vowel substitution were the main reasons for the introduction of this subcategory. The overall category of VOWEL was retained, however, in case the creation of these subcategories exerted undesirable effects on the statistical procedures (for example, an excessive number of independent variables).

Finally, there were relatively few instances of word stress error, marked as such by the students. The only retained token of word stress (WS) error was the word *information* (Speaker 5), but as Speaker 5 was excluded from the subsequent statistical analysis it is discussed here purely for general interest. Both specific error codings and general comments suggested that this was seen as a word stress problem; one student wrote 'good pronunciation on most words (except word stress 'information')'. The actual pronunciation shows a full vowel, rather than a reduced one, in the second syllable, but the main trigger for the WS coding by the students seems to be the roughly equal stress given to the first and penultimate syllables, compared with the stress pattern normally found in standard varieties (penultimate syllable stress, with secondary stress on the first syllable). It could be argued that FULL VOWEL errors also affect perceptions of word stress, but as has been discussed in Chapter 2, it seems that vowel quality is not among the most important markers of word stress.

Frequency charts for tokens of SYLL, VOWEL SUB and FULL VOWEL identified by the students and included in the analysis are given in Figures 5.4 to 5.6 below. The frequently-occurring tokens are discussed in more detail in the speaker-by-speaker analyses in section 5.5.

Fig. 5.4. Identified word tokens for SYLL.

Fig. 5.5. Identified word tokens for VOWEL SUB (subcategory of VOWEL).

Fig. 5.6. Identified word tokens for FULL VOWEL (subcategory of VOWEL).

5.4.2.2 Consonantal features: PHONSUB, OTHER C-SUB

These consonantal categories were distinguished from each other mainly on the basis of specificity of operation. PHONSUB consists of /r/ and /v/ substitution, both of which only occur in onset position. The most frequently identified tokens in the data were *risky* (/r/ substitution, Speaker 1) and *advantage* (/v/ substitution, also Speaker 1); these were clearly marked by the students, often with specific reference to the [w] substitutions used. The fact that a considerable number of these substitutions (47 in total) were identified also suggested that subdivision was needed, and as with the VOWEL category it was divided into two subcategories, PHONSUB-R and PHONSUB-V. The frequency charts for PHONSUB show that there were relatively few word tokens, but that the number of students coding them was quite high, suggesting that these errors were both noticeable and noteworthy from the students' point of view (they were asked to mark the features that affected their acceptability judgments). The charts also show that /v/ substitution was marked more frequently than /r/ substitution, but this may have been due to frequency effects (there were more possible contexts in the samples). An additional observation about /r/ substitution can be made, based on its occurrence in *granting* (Speaker 6). Here it occurs within an initial consonant cluster, and this may be influenced by the

existence of a co-articulated labio-velar stop [g^w] in Cantonese (Chan and Li 2000: 69).

OTHER C-SUB includes several processes which may apply to various consonants. Firstly, as shown in Table 5.10, there is final obstruent devoicing, which in the data can apply to single-consonant codas in words such as *have* and *because*, to both members of biconsonantal clusters in words such as *cards*, and to the second member of such clusters in words such as *aims*. Secondly, there is non-systematic consonant substitution, appearing in the words *built* (where a final [d] was used) and *department* (where the medial /p/ was heard by several students as a [b], perhaps because of its weakly aspirated quality). Thirdly, there is consonant deletion, such as of the final /d/ in *continued*. The case of final consonant deletion in *continued* (from Speaker 2's sample) again raises the issue of the extent to which such deletions can be classed as phonological, rather than grammatical. However, there was only one identification of this feature, so it is unlikely to exert much influence on the results (the deletion of the plural 's' in the word *applications* by Speaker 5 was noted by 17 students, and led to the decision to remove this speaker from the analysis).

The frequency chart for OTHER C-SUB shows that the four most frequently marked tokens account for more than half of the total. The use of [d] instead of the final /t/ in *built* by Speaker 2 seems idiosyncratic, although assimilatory effects would tend to make this easier to pronounce in this intervocalic context (especially if the two-part verb *built up* is analysed as one phonological 'word'). The final cluster in *cards* (Speaker 1) shows a combination of deletion and devoicing, as the /d/ is weakly articulated and devoiced, and the following fricative is voiceless. It is worth noting that the plural or third person 's' in bimorphemic final clusters containing /t/ or /d/ after voiced plosives (as in *cards*) is prone to devoicing in HKE because of the effects of progressive assimilation. In this case, the /d/ would often show devoicing even if not followed by another consonant, and this in turn predicts a voiceless realisation of the final 's'. The word *aims* (Speaker 9) shows devoicing of the final 's', but obviously without any assimilatory effects.

The fact that many of these features occur in other varieties of English, combined with the above observation about /r/ substitution in *granting*, suggest that there is a possible qualitative distinction between PHONSUB and OTHER C-SUB: the former category is more closely related to transfer from the L1, while the latter category consists of developmental or universal features. Of course, it is not always possible to neatly separate the two. Final obstruent devoicing in HKE phonology may also be related to the absence of voiced final consonants in Cantonese, although in cross-linguistic comparisons it is often cited as an example of a developmental process related to language universals (for example, see Major 2008: 76).

Frequency charts for PHONSUB-R, PHONSUB-V and OTHER C-SUB are given in Figures 5.7 to 5.9 below.

Fig. 5.7. Identified word tokens for PHONSUB-R.

Fig. 5.8. Identified word tokens for PHONSUB-V.

Fig. 5.9. Identified word tokens for OTHER C-SUB.

5.4.2.3 Other consonantal features: TH-STOP, TH-FRONT, L-VOCAL

These features are well-known phonological categories and are widely distributed across varieties of English. TH stopping in initial position is found ‘practically all around the globe’ according to Schneider (2004: 1123). TH fronting has a rather more restricted distribution, being found in New Zealand English (Gordon and Maclagan 2004) and in Estuary or south-eastern British English. Their presence in

Hong Kong English is a topic worthy of further investigation; Deterding *et al.* (2008) consider acoustic similarity as an explanatory factor but also note that this pronunciation is found in London, raising the possibility of contact influences. L vocalisation is listed by Schneider (2004: 1125) as occurring fairly generally in some dialects of American English, variably in Australian and New Zealand English, and regionally in British English. All of these features have pedagogical relevance, and are identified in Jenkins (2000) as substitutions that do not affect international intelligibility. Some indication of their acceptability would therefore be useful in order to round out the evaluation picture and begin to make pedagogical recommendations.

The construct validity of these categories is therefore robust, in terms of their phonological distinctiveness, and their pedagogical relevance is clear. However, the difficulty with these features is that they tend to be difficult to distinguish from ‘standard’ variants in many cases, because of acoustic similarity. The summary of the data analysis for the implicational scale presented in Chapter 3 (Table 3.5) shows that even during careful listening, the levels of agreement between the two raters were lowest for these three features, ranging between 75.6% for L vocalisation and 90.4% for TH fronting. The accuracy of identification by the student listeners in this study therefore requires some attention. Eight of the original nine tokens of TH fronting marked by students were excluded from the analysis as they were not thought to represent sufficiently robust examples of the feature in question. For example, the most frequently marked token of TH fronting was *think* (Speaker 12, 10 mentions), but it was decided that it did not represent a clear enough example of the category. In the final analysis the only retained token of TH fronting was the word *forthcoming* (Speaker 6), where it occurs in word-medial position.

In the case of TH stopping, more tokens were retained. The most frequently identified instances were in initial position in the words *the* (Speaker 6 and Speaker 9) and *that* (Speaker 1), and in word-medial position in *bothered* (Speaker 3). As the speakers in samples 1 and 6 were actually the same person, it is worth noting that TH stopping was not categorical in this case. It was not used in the third instance of *they* in the Speaker 1 sample, where a slight pause before the word may indicate increased

attention to this sound or to the following clause in general. Figure 5.10 shows the distribution of identified word tokens of TH stopping.

Fig. 5.10. Identified word tokens for TH-STOP.

In the case of L vocalisation or deletion, very few instances were marked by the students. The only retained token was *people* (Speaker 4), which received two codings. This probably reflects the inherent difficulty of distinguishing a vocalised /l/ from its non-vocalised, consonantal variants. It may also suggest a more specific inability on the part of the students to differentiate between the two variants. Although missed identifications will mainly be discussed in the speaker-by-speaker analysis in section 5.5, it is worth considering those relating to L vocalisation as there were more cases than in other categories. L vocalisation seems to occur in the words *itself* (Speaker 3) and *children* (Speaker 8), where the preconsonantal context may mask the vocalisation, and in the word *local* (Speaker 9). Several students marked Speaker 5's pronunciation of *impossible* as an example of L vocalisation, but this speaker was excluded from the analysis.

The relevance of these missed identifications is that the error analysis may fail to reflect the importance of certain features. On the other hand, if certain features pass unnoticed by a majority of listeners, then the low number of codings can be said to reflect this low noticeability. The lack of statistical significance, which is a likely

result, can also be seen as an indication of a low level of evaluative significance, with these samples and listeners.

5.4.2.4 Consonant clusters: CCRF, CCM-I

The category of initial cluster modification, or CCM-I, was easily dealt with; there was only one clear example in the data (in the word *progress*, by Speaker 12), and this was only noted by one student. It was therefore decided to exclude CCM-I from the analysis, while noting that the low number of identifications suggested difficulties in perception. Speaker 12's utterance also contained a number of other features, and this may have reduced the rater attention available for each type.

The categorisation of final consonant cluster reduction (final CCR) was more problematic. The initial analysis showed that it was the most frequently identified feature, with a total of 237 recorded error codings in a range of words. However, given the widespread occurrence of final CCR in all varieties of English there would seem to be little point in including it as a category. As Schreier (2009: 62) observes, in itself final CCR 'offers little qualitative information on dialect distinctiveness and has no diagnostic value whatsoever'. In the present study it was decided to focus on the possible qualitative distinctiveness of final CCR in HKE by examining the phonological contexts in which it occurs. Final CCR is constrained by two main factors, morphemic status and the following phonetic segment (Schreier 2009: 60). The latter constraint predicts that CCR is less likely to occur in prevocalic contexts, so that /t/ deletion in *next* is more likely to occur in *next day* than it is in *next afternoon*. A basic division can therefore be made between prevocalic final CCR (less frequent in standard varieties of English) and preconsonantal CCR (frequent and widespread in standard varieties). Schreier (2009: 68) observes that prevocalic CCR is more prominent in non-native varieties of English, but assuming that both types will be present in speakers' utterances, the prevocalic/preconsonantal distinction provides an initial categorisation. Included in the prevocalic category are deletions occurring in prepausal position, as final CCR also seems to be less likely before a pause. The difficulty here is that 'pause' is hard to define operationally, and in this study 'prepausal' is taken to mean utterance-final position.

Looking at the CCR codings as a whole, there was a tendency for prevocalic and prepausal CCR to be marked more often. The most frequently identified token of prevocalic final CCR, or CCRF-PV, was the word *confused* (Speaker 12), which occurs in prepausal or utterance-final position and is possibly more salient because of this. The other frequently marked tokens included *relaxed* (Speaker 3) and *privileged* (Speaker 8). The word *found* (Speaker 12) was the most frequently occurring token of preconsonantal CCR, or CCRF-PC, but there is also a pause before the next word. The marking of final CCR showed a high level of accuracy, and very few tokens were excluded. No missed identifications were discovered. Figures 5.11 and 5.12 below show the distribution of identified final CCR word tokens in the categories of CCRF-PV (prevocalic) and CCRF-PC (preconsonantal). All the tokens of CCRF-PV relate to the deletion of /t, d/, while two tokens of CCRF-PC relate to /k/ deletion in the word *think*.

Fig. 5.11. Identified word tokens for prevocalic or prepausal final CCR (CCRF-PV).

Fig. 5.12. Identified word tokens for preconsonantal final CCR (CCRF-PC).

It is immediately noticeable that there are three past participle forms among the four most frequently-marked CCRF-PV tokens. More generally, the bar charts show that five out of seven tokens of CCRF-PV are bimorphemic, while only three out of 14 CCRF-PC tokens are. The data above may thus conflate the effects of prevocalic context and morphemic status, and a more detailed analysis would be required to fully assess the effects of different types of final CCR. Schreier (2009: 60) notes that final CCR is generally less likely in bimorphemic clusters, and the greater noticeability of such reductions in prevocalic position may have resulted in a larger number of identifications. There is again the possibility that these reductions were seen as being grammatical, as well as phonological, in nature. A more detailed study would also need to take account of whether the syllables containing CCR are stressed or not, as this also affects their salience for the listener. The first two examples of CCRF-PV (*confused* and *relaxed*) both feature final CCR in stressed syllables, as do the first six examples of CCRF-PC (if monosyllabic words are classed as ‘stressed’).

5.4.2.5 Connected speech: CS-LINK

The only suprasegmental feature retained in the analysis referred to linking phenomena in connected speech, marked ‘CS’ by the students. Most of the 22 codings in this category referred to Speaker 3, in which the words *by it* are linked

with an ‘r’ sound, rather than the expected [j] glide normally found in such a hiatus (Britain and Fox 2009: 179). The other example of CS-LINK involved the opening words of Speaker 9’s utterance, *the accredited*. Many speakers would use the prevocalic allomorph of the definite article, /ði/, triggering a glide [j] between it and the following vowel, although the use of the preconsonantal allomorph /ðə/ in all contexts is becoming more common in some varieties of English (Britain and Fox 2009). Speaker 9 uses the preconsonantal allomorph /ðə/ (actually [də], with TH stopping), and this was noted as a CS problem by two students. The fact that this feature is noticeable in many varieties again suggests that lack of accent familiarity may explain some of the codings.

Fig. 5.13. Identified word tokens for connected speech linking phenomena (CS-LINK).

5.4.3 The effects of the features on the acceptability ratings

Table 5.10 and the bar charts above show the occurrence of different types of error codings, and the next stage of the analysis was to measure the effects of the different error categories on the acceptability ratings. The student error codings were entered into an SPSS data table as binary values, in which 1 indicated the presence of an error coding and 0 its absence. The categories of phonological error were then used in a linear regression analysis with overall acceptability as the dependent variable. As

with the analysis of the speaker variables carried out earlier, the aim of the regression analysis was to compare the effects of the independent variables, in this case the error categories, on the acceptability scores.

The priority of this analysis was thus to identify the relative effects of the error categories, while acknowledging the fact that other variables may be present. A general goal of regression analysis is to identify the smallest set of independent variables that will predict a substantial and independent component of the variability in the dependent variable (Tabachnick and Fidell 1996: 132). The earlier regression analysis has already suggested that among the measured speaker variables such as speech rate, complexity and pitch span, phonological error had the greatest influence on the acceptability scores. This analysis was therefore designed to identify the components of 'phonological error' that had the greatest effect. In this part of the study, the construct validity of both the dependent and the independent variables has been verified through measurement and careful categorisation.

The assumptions of regression analysis mentioned in section 5.3.2 above will first be examined again for this analysis. Firstly, sample size; assuming that all twelve error categories and subcategories in Table 5.10 were to be employed in the analysis, according to the formula provided by Tabachnick and Fidell (1996) a sample size of at least 116 cases would be required, and this is comfortably exceeded by the 520 cases (52 listeners, 10 speakers) in this part of the study. The SPSS program gives multicollinearity statistics as part of its regression output, and as an initial regression run using the maximal set of twelve independent variables did not reveal any such intercorrelations, no variables were excluded on this basis. Tabachnick and Fidell (1996: 133) advise that extreme cases, or outliers, 'have too much impact on the regression solution and should be deleted, rescored or the variable transformed'. A scatterplot revealed only two cases with standardised residuals of less than -3.3 (the definition of an outlier, according to Tabachnick and Fidell 1996: 139). As this represented less than 0.5 percent of the cases, and as Pallant (2001: 144) state that it is not unusual for a few outliers to appear with large samples, no remedial action was taken.

Regarding the distribution of residuals, the normal probability plot indicated a normal distribution. A scatterplot showing the predicted values of the dependent variable against the standardised residuals revealed no linearity problems but indicated some heteroscedasticity. An assumption of regression models is the homoscedasticity of residuals, meaning that the variability in scores for the dependent variable is roughly the same at all values of the independent variable or variables (Weiner *et al.* 2003: 128; Berry and Feldman 1985: 73). This may be due to higher acceptability being associated with fewer errors; thus, there is a greater measurement error when there are fewer error codings. However, as Tabachnick and Fidell (1996: 138) point out that heteroscedasticity does not invalidate a regression analysis, but merely weakens it, no attempts were made to transform any of the variables. Finally, the Durbin-Watson statistic of 1.753 in the model output summary (Table 5.11) suggests that the independence of residuals is within the acceptable range. Savin and White (1977) give lower and upper bounds of 1.55 and 1.801 for samples with 200 cases and 12 variables.

Tables 5.11 and 5.12 below show the SPSS output tables for the regression analysis with the full set of twelve independent variables. Space limitations in the character field mean that some feature names are shortened in the tables; the category OTHER C-SUB is shown as 'OTHERCSB', and FULL VOWEL is 'FULVOWEL', for example.

Table 5.11. The SPSS model summary output table.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.470(a)	.221	.203	.86596	1.753

a Predictors: (Constant), CSLINK, VOWELSUB, LV, FULVOWEL, PHONSUBV, THS, OTHERCSB, THF, PHONSUBR, CCRFPV, CCRFPC, SYLL

b Dependent Variable: AVEBF

Table 5.12. The SPSS regression coefficient output table.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4.238	.057		74.264	.000		
	THS	-.425	.167	-.101	-2.540	.011	.979	1.021
	CCRFPV	-.493	.092	-.218	-5.374	.000	.932	1.073
	CCRFPC	-.307	.097	-.130	-3.163	.002	.915	1.093
	LV	-.091	.615	-.006	-.148	.882	.995	1.005
	THF	-.960	.442	-.087	-2.172	.030	.967	1.034
	PHONSUBR	-.300	.223	-.055	-1.345	.179	.918	1.089
	PHONSUBV	-.766	.157	-.195	-4.869	.000	.955	1.047
	SYLL	-.684	.132	-.219	-5.188	.000	.864	1.157
	VOWELSUB	-.819	.173	-.188	-4.740	.000	.981	1.019
	FULVOWEL	-.367	.283	-.052	-1.297	.195	.955	1.048
	OTHERCSB	-.463	.127	-.147	-3.634	.000	.938	1.066
	CSLINK	.169	.194	.035	.874	.382	.950	1.053

a Dependent Variable: AVEBF

In Table 5.11, the overall adjusted R squared value of 0.203 is within the ‘modest fit’ range given by Cohen, Manion and Morrison (2007: 538). This indicates that 20.3 % of the variability in the acceptability scores can be explained by these variables, a fairly low proportion but one which is unsurprising given the authentic data and the possible effects of unmeasured extralinguistic variables (as with the earlier regression analysis). The ‘Sig.’ (significance) column shows that five of the features were significant at the $p < 0.0001$ level (these are shown in bold type). Ranked in terms of their standardised beta coefficients, these were:

- SYLL;
- CCRF-PV;
- PHONSUB-V;

- VOWELSUB; and
- OTHER C-SUB.

CCRF-PC was also significant at the much lower threshold of $p < 0.01$. The discovery of significant variables indicates that ‘it is useful to proceed with the analysis, as it contains important results’ (Cohen *et al.* 2007: 538). The effects of the features will be considered in more detail in Chapter 6, but it is worth noting even at this stage the close relationship between the acceptability effects of phonological features and their likely intelligibility effects. All of the five features found to be significant in Table 5.11 would also be expected to reduce intelligibility, according to the list of core features presented in Chapter 2 (with the possible exception of VOWELSUB, depending on whether quantity or quality modifications were involved). Conversely, only one of the non-significant factors (PHONSUB-R) would be expected to reduce intelligibility. In other words, deviant core features, in the terminology of Jenkins (2000), were those which also reduced acceptability. However, modifications to non-core features, such as dental fricative substitutions, did not affect the acceptability scores significantly.

The limitations of the procedure must also be considered. As has been mentioned, the effects of feature frequency, noticeability and severity on the results are not entirely clear. In some cases the low number of tokens probably explains the low level of significance achieved by, for example, LV and TH-FRONT. In the case of LV this may be related to low noticeability, but in the case of TH-FRONT it is probably also a result of the relative rarity of contexts for its occurrence; as explained in Chapter 2, the voiceless dental fricative simply occurs less often. In both cases, if there had been more tokens, the results may have been different.

On the other hand, frequency of identification (whether a result of natural frequency patterns, or noticeability) is clearly not the only determining factor. While PHONSUB-V achieved statistical significance with 30 identifications or codings, the same number of identifications did not result in significance for TH-STOP, suggesting a difference in the perceived severity of these error types. The category with the largest number of identifications (CCRF-PC) also did not achieve the

$p < 0.0001$ level of significance. Therefore, it appears that the survey procedure and data analysis were able to distinguish to some extent between frequently-marked, but non-severe features (such as CCRF-PC, and to a lesser extent TH-STOP) and less frequently-marked, but more severe features such as VOWELSUB.

When considering the frequency of identification, it should also be noted that many errors were marked by relatively few students. Although no detailed measurements of detection rate were made it is clear from the bar charts above that only two errors – SYLL in *political* and CCR-PV in *confused* – were noted by more than half of the students. It is uncertain whether this meant the other students did not notice the errors, or whether they thought they were unimportant. However, a more general explanation lies in the open-ended nature of the survey form. Students were advised to mark no more than three errors, while clearly there were many possible choices in the samples. This suggests that students varied in their perceptions of what was important, although once again the use of 52 student raters appeared to be sufficient to attain significance for some of the error categories.

Another factor that complicates the interpretation of the data is the likelihood of feature co-occurrence, as suggested by the implicational scale in Chapter 3. This may also explain why certain features that were marked relatively less frequently (such as PHONSUB-V and VOWELSUB) were able to exert significant effects on the acceptability ratings: they may have been accompanied by other features, the combined effects of which increased the tendency to give lower scores. However, it is important to distinguish between the effects of a feature on the acceptability ratings of a speaker and the effects of that feature on the ratings as a whole. The regression procedure involves isolating the effects of each case of feature marking and then combining them into an overall ‘best fit’ equation, reducing the chance of their being influenced by speaker-specific phenomena. Even so, it is also likely that in terms of cause and effect, it would be an oversimplification to say that the features influence the ratings in a mechanical fashion. A more accurate interpretation might be that the features are indexical of some overarching attribute (such as proficiency, for example), and that it is perceptions of this quality that are the ultimate

determinant of the ratings. Nevertheless, this is how human perception and attribution often work, more generally as well as in the field of language use.

Still another intriguing possibility is that as well as the acceptability ratings being influenced by feature use, feature non-use may also have played a part. The successful negotiation of a complex sound sequence, or the use of a preferred variant in a particular context, may have increased acceptability ratings without being reflected in the codings. But once again, the regression procedure does not seem to be unduly compromised by this, and would take account of feature absence being associated with higher ratings.

These methodological and interpretational problems are almost inevitable with the use of authentic speech data. While there are no easy solutions to these problems, the general position adopted by this study is that the error codings and acceptability scores represent the combined effects of factors such as the number of possible or actual contexts, the noticeability of the error, and the perceived severity of the error. By abstracting the features from their contexts and analysing how their occurrence affects acceptability on a case-by-case basis, the variation between cases is minimised. Statistical significance can thus be seen as something of a portmanteau, but one that has real-world relevance: statistically significant features are those that have general evaluative significance within the speech community under consideration.

Finally, as in all regression analyses, the possible effects of other important but unmeasured variables should not be ignored. The earlier regression analysis showed that phonological error was able to predict a larger proportion of the variance in acceptability scores than any other measured variable, and taken together the two analyses provide an indication both of the relative importance of such errors and of the differential importance of particular types of error. However, behind both of these conclusions lies the caveat that other variables are certainly at work. A possible example of an unmeasured extralinguistic factor is related to the ‘authority’ dimension of accent studies (e.g. Bayard *et al.* 2001); it is possible that tone and timbre of voice, combined with topic and a certain threshold level of language

proficiency, may interact to produce overall impressions of authority, calmness, confidence and so on. The effects of suprasegmental factors such as the degree of syllable- or stress-timing are also ignored by the above analysis. Nevertheless, the findings show that there are significant differences between the error categories in terms of their relative effects on the acceptability scores.

5.5 Student comments and individual speaker analysis

In addition to the quantitative data in Part 1, student comments in Part 2 of the survey form provided some qualitative comments for use in triangulation. The comments were analysed and allocated to four category headings: ‘Sounds’ (meaning segmental features and comments about specific words), ‘Connected Speech and Intonation’, ‘Fluency’ and ‘Other’. Comments about suprasegmental features were frequent, but this does not necessarily mean that these were more important than the other categories; it is probable that intonation was seen as a global characteristic which was less amenable to local marking than segmental features. In the figures below, comments are also subdivided into positive, negative and neutral categories. Comments referring to the speakers’ employment status or accent strength were assigned to the ‘neutral’ category, as were mixed comments such as ‘good pronunciation on most words, except word stress on “information”’. There were several comments relating to possible employment status (for example, ‘LegCo (Legislative Council) member’), but nearly all of these were made by the same student, and it does not appear that presumed employment status was a significant factor.

For each of the twelve speakers in the following sub-sections, a visual representation of the transcript shows the frequency with which certain errors were marked by the students. Only the errors that were submitted to subsequent analyses have been included. These errors are signalled by a larger font size in the relevant part of the transcript. The default font size is 10, but errors have been shown using a font size of 15 as the baseline; thus one error is shown by a font size of 16 and ten errors are shown by a font size of 25. The total number of identifications is also shown by a

superscript. The advantage of this is that it enables comparisons to be made between speakers, as well as showing the words and parts of words that were marked most often. However, it should be borne in mind that not all of the errors were equally significant, as shown by the statistical analysis. The number of student comments in each category is shown in a bar chart, which provides a visual summary of the distribution of positive and negative comments across the areas mentioned above (sounds, connected speech and intonation, fluency, and other comments). The positive, negative and neutral comments received are shown in table form for each speaker. There is also a commentary that attempts to link the different sources of data and explain the acceptability ranking received by the speaker. While this section thus focuses on the speakers, including the effects of feature use, the features themselves will be considered in more detail in the following chapter.

As the identified error tokens have already been checked for accuracy, and the ambiguous or misidentified tokens removed, it can be assumed that the remaining tokens fairly represent the categories they have been assigned to. Although there were few cases of missed features, and while the focus of the study is on students' perceptions, when these features were thought to be significant they are mentioned in the sub-sections below. The speakers are considered in numerical order, except in the case of Speakers 1 and 6, who were actually the same person and are considered first.

5.5.1 Speaker 1

Figure 5.14. Transcript of Speaker 1's sample showing the location of errors and the number of identifications.

Figure 5.15. The distribution of Part 2 positive and negative comments for Speaker 1.

Table 5.13. Students' Part 2 comments regarding Speaker 1.

Category	Positive comments	Negative comments	Neutral comments
Sounds		- "don't" should be pronounced longer	
Connected speech & intonation	- good sentence stress / rhythm	- poor intonation - weird stress, pause - should have more intonation - very strange intonation	
Fluency	- fluent, quite professional		
Other	- good English pronunciation		- Head of Monetary Authority?

This speaker was ranked 11th in both acceptability rankings and 12th. Speaker 1 and Speaker 6 were the same person; the fact that the two samples received similar ratings (overall acceptability scores were 3.20 and 3.29 respectively) could be taken as evidence of the reliability of the evaluation procedure. The Speaker 1 sample has the highest speech rate of all, at 222 wpm, but unlike in the Speaker 6 sample, there are no SYLL errors. The overall error density is the second highest, if measured in terms of errors per word (4.33). The most frequently marked errors were:

- PHONSUB-V or /v/ substitution in *advantage*, with 18 student markings in Part 2 of the questionnaire. The speaker's actual rendition is [ɛ'wɑ:ntɪdʒ], with an elided /d/ in the initial cluster (this was noted by one student). The /v/-substitution resembles Hung's postulated 'underlying representation' of *advertise* in HKE as [l'ɛdwɜtaɪz] (Hung 2000: 350), although it appears as the onset of a stressed syllable in this case.

- OTHER C-SUB, in this case devoicing, in *cards* received 15 student mentions. Both of the final consonants in *cards* show devoicing, the second perhaps undergoing progressive assimilation under the influence of the first.
- CCRF-PC or preconsonantal final CCR occurred in *don't* (10 mentions), where there is also vowel shortening ([don]; this was noted by the student who wrote “‘don’t’ should be pronounced longer”).
- PHONSUB-R in *risky* (6 mentions).
- TH stopping in *they* (2 mentions) and *that* (4 mentions).

Given the overall statistical significance of /v/ substitution, the marking of this feature by a large number of students goes some way towards explaining the low acceptability score. Among the other features, OTHER C-SUB was also significant and CCRF-PC was significant at a lower level ($p < 0.01$). In general, there is congruence between the various parts of the error analysis, in terms of their effects on acceptability; there is a relatively large number of errors, and a large number of significantly acceptability-reducing errors are present within these.

The comments regarding this speaker are not very detailed, but there are three positive comments. One was of a general nature (‘good English pronunciation’) and two referred to suprasegmental features (‘good sentence stress/rhythm’ and ‘fluent, quite professional’). However, intonation was not seen as favourably (comments included ‘poor intonation’, ‘should have more intonation’ and ‘very strange intonation’). There is no obvious reason for this in the pitch span data, and it is likely that the placement of stress on particular words was seen as ‘strange’. For example, in the phrase *they don't have to* there is a noticeably syllable-timed rhythm and no obvious tonic stress placement, except for a falling tone on *to*. As has been mentioned, the fact that there are several comments about intonation is most probably a result of the difficulty of marking intonational errors, and suprasegmental errors in general. Nevertheless, these comments suggest that such features are definitely among the unmeasured variables that also account for the variance in acceptability scores.

5.5.2 Speaker 6

Figure 5.16. Transcript of Speaker 6’s sample showing the location of errors and the number of identifications.

Figure 5.17. The distribution of Part 2 positive and negative comments for Speaker 6.

Table 5.14. Students' Part 2 comments regarding Speaker 6.

Category	Positive comments	Negative comments	Neutral comments
Sounds		<ul style="list-style-type: none"> - the word 'political' is very unclear - could be slower [re 'political'] - confusing [re 'political'] - bad pronunciation in key words 	
Connected speech & intonation	<ul style="list-style-type: none"> - fluent - overall: fluent 	<ul style="list-style-type: none"> - poor word stress - very flat tone 	<ul style="list-style-type: none"> - intonation / accents [sic] very local
Fluency		<ul style="list-style-type: none"> - speaks too fast, pronunciation sounds strange, not clear enough - he speaks too fast - too fast in speaking - speed - some words are too fast 	
Other	<ul style="list-style-type: none"> - (university professor?) Good English pronunciation 	<ul style="list-style-type: none"> - strange sound!!! - strange pronunciation 	

As mentioned above, Speaker 6 and Speaker 1 were actually the same person; both samples received similar overall acceptability ratings (3.06 and 3.00 respectively), placing them in the lower quartile (10th and 11th in the rankings). In the Speaker 1 sample the error density is above average, while in the Speaker 6 sample it is slightly below average. The speech rate is above average in both cases, but is closer to the norm in Speaker 6's utterance.

Speaker 1's low acceptability score was thought to be a result of the above-average error density and the presence of significant errors (/v/-substitution and OTHER C-SUB, a consonantal modification) within the sample, but in the case of Speaker 6 the most likely explanation of the acceptability ratings is the large number of SYLL markings (this was also a significant category). The word *political* was marked by 29 students, making it the second most frequently marked word among all the samples. As has been explained in section 5.4.2.1 regarding SYLL errors, the rapid production of this word, including schwa absorption in the first syllable, appears to have been the trigger for these codings. The general comments in the table above also lend support to the idea that 'syllabic modification' is linked to overly rapid speech, and three mention this particular word.

Another significant feature was the VOWEL error represented by vowel shortening in the first syllable of *leadership*. This could also be seen as a consequence of rapid speech, but was marked as a vowel problem by the four students who noticed it. Among the non-significant errors there was TH stopping, noted in *the* (seven mentions) and *There's* (one mention), TH-fronting within *forthcoming* (four mentions) and /r/-substitution, marked in the words *reason* (nine mentions) and *granting* (two mentions). The /gr/ cluster of *granting* is produced with a sound resembling the labial-velar approximant [g^w] of Cantonese.

Once again it is the presence of salient, significant features, noted by many students, that provide a likely explanation of the acceptability ratings. Of course, it is also possible that the students realised this was the same person, and gave similar ratings to both samples. The fact that the features were different in the cases of 1 and 6 (/v/-substitution and consonantal modification *versus* syllabic modification and vowel modification) does not weaken the features-based explanatory approach, as significance in the regression procedure derives from a consideration of all instances of the feature, across all cases. On the other hand, it is noticeable from the bar charts (Figures 5.4 and 5.8) that two of the four identified word tokens of SYLL and one of the two identified word tokens of PHONSUB-V occurred within this speaker's utterances. The regression procedure does not distinguish between speakers, but rather between cases involving features. While this may have reduced the ability of the regression procedure to separate the effects of the two features, the fact that both attained statistical significance suggests they were important overall determinants of the acceptability scores.

However, as mentioned in section 5.4.2, the possibility that there were co-occurring but unmarked features should not be discounted; the concept of significant features being indexical may be relevant here. The implicational scale suggests that /v/- and /r/-substitution would tend to be accompanied by a range of other features. With this speaker, at the phonemic level this is true of TH fronting and TH stopping, and possibly also of initial CCM, if the /r/-substitution in *granting* were to be analysed as the modification of an initial cluster. An additional linguistic explanation of the low acceptability could thus be that there are many minor errors, including those at a subphonemic level, that were not measured by the analysis. These features may have

influenced the acceptability ratings by affecting perceptions of overall accentedness. An additional non-linguistic explanation could be that once certain features have been noticed by the listener, other features will tend to be rated more harshly. It may not only be the significant features themselves that lead to low acceptability scores for the speaker in question, but also their effects on how other features are perceived. This ‘association effect’ may also partly explain why Speaker 1 and Speaker 6 received similar ratings.

5.5.3 Speaker 2

Figure 5.18. Transcript of Speaker 2’s sample showing the location of errors and the number of identifications.

You can see the words commitment^{t⁹}, sustainability and pragmatism. In the pas^{t¹⁵} year the economy has continued^{d¹} to perform well and we have buil^{t²¹} up a considerable surplus

Figure 5.19. The distribution of Part 2 positive and negative comments for Speaker 2.

Table 5.15. Students' Part 2 comments regarding Speaker 2.

Category	Positive comments	Negative comments	Neutral comments
Sounds			
Connected speech & intonation	- some linking, e.g. 'past year' - very good pronunciation and intonation	- no intonation - intonation X - more weak forms/linking needed - intonation not strong enough	
Fluency			
Other	- good pronunciation - satisfactory		- not like a foreigner - very common in HK - Financial Secretary?

Speaker 2 is in the middle of the acceptability rankings (6th place, under both measures). The error density is slightly below average, as is the speed of delivery. Among the other speaker variables, the minimum frequency is the highest; this sample was recorded during a public address and there is probably a tendency to avoid sounding inappropriately relaxed in these situations. The effects of the increased volume used in public speaking may also explain this. Pitch span is the second lowest, perhaps as a result of the high minimum frequency. In this case, some student comments also referred to intonation (for example, 'intonation not strong enough').

For speakers in this range of the acceptability rankings, there may be a mixture of significant and non-significant error categories and it is difficult to assess their combined effects. Four words, representing two error categories, were marked:

- OTHER C-SUB in *built* (21 mentions), where the final /t/ is realised as a [d], and in *continued* (1 mention), where the final /d/ is deleted. The first appears to be an idiosyncratic error, although voice assimilation is perhaps natural in this intervocalic context. The two-part verb *built up* is pronounced quite rapidly, and the presence of [d] can perhaps be attributed to a medial /t/ voicing process such as might occur in words like *better*. The deletion of /d/ in singleton codas is quite unusual, and may be related to the consonantal onset of the following word.

- CCRF-PC in *past* (15 mentions) and *commitment* (9 mentions). The fact that *commitment* occurs at the end of a tone unit perhaps makes the CCR more noticeable, and more akin to a prepausal context. Equally, the word *past* is given considerable stress by this speaker, as is often the case with content words. This may have increased its salience for the listeners.

Although OTHER C-SUB was a significant feature in the regression analysis, the mixed nature of this category reduces its explanatory and predictive usefulness. In this case, although there were a considerable number of markings, the effects on acceptability do not seem to have been too severe. This raises the possibility that not all instances of OTHER C-SUB were viewed in the same way. In general, however, there were a relatively large number of markings of particular words (*built* was the most frequent example of OTHER C-SUB, and *past* was the third most frequent example of CCRF-PC). This suggests that these errors were quite noticeable, although other aspects of the sample may have compensated for their effects.

5.5.4 Speaker 3

Figure 5.20. Transcript of Speaker 3's sample showing the location of errors and the number of identifications.

Figure 5.21. The distribution of Part 2 positive and negative comments for Speaker 6.

Table 5.16. Students' Part 2 comments regarding Speaker 6.

Category	Positive comments	Negative comments	Neutral comments
Sounds		- CC [consonant clusters] could be better	
Connected speech & intonation	- shows connection between words	- odd intonation - HK-style linking, too exaggerated	
Fluency	- native, fluent - very fluent	- speaking too fast - too fast - can speak slower	
Other	- quite natural presentation - good English pronunciation - quite natural - generally quite good	- some minor pronunciation errors - unclear - not clearly pronounced	

Speaker 3 came fourth in the acceptability rankings. The overall error density is the second lowest among the samples. The speech rate is slightly above average at 185wpm, and there are two low-frequency lexical items (*heterogeneous* and *identities*). All of these factors may have helped to create an impression of competence and proficiency. Pitch span is the second highest, mainly as a result of the emphasis given to words such as *bothered*. The speaker was almost certainly not seen as a native speaker, despite being ranked eighth on the 'sounds like a Hong Kong person' question and receiving the comment 'native, fluent'. Negative comments (9) outnumbered positive ones (7), and reflected some divergent opinions. Two students were impressed by his fluency, while three thought he could speak

more slowly. One thought he ‘shows connection between words’, while another saw this as ‘HK-style linking, too exaggerated’.

In terms of the marked errors, the word *relaxed* was noted by 22 students as an instance of final CCR because of the deletion of the final /t/. Despite its prevocalic position (*relaxed attitude*) and the fact that it occurs in a stressed syllable, this error does not appear to have been unduly penalised by the students in this instance, despite the fact that CCRF-PV was one of the significant error categories, in overall terms. The triconsonantal status of the cluster may have reduced its perceived importance, as may its bimorphemic status (although this usually inhibits final CCR in World Englishes; see Schreier 2008: 211). Preconsonantal final CCR occurs in the words *don’t* (five mentions) and *think* (one mention). The unusual linking of *by it*, where there is an /r/-like approximant, rather than a [j] glide, was noted by 22 students. TH-stopping in *bothered* was noted by four, while the SYLL error in *national*, where the final syllable is not pronounced, was noted by only two.

Listening to the recording in more detail, there are some features that passed unnoticed by students. The final /d/ is missing in *attitude*, and in fact the word *are* is not pronounced at all. L vocalisation occurs in *local*. It seems likely that the students realised there were more shortcomings, compared with the higher-ranked speakers (the general comments included ‘some minor pronunciation errors’ and ‘not clearly pronounced’). These somewhat vague comments, and the failure to notice the sound changes above, may be related to the relatively rapid speech rate; this recording suggests that some sound modifications may become less noticeable as speech rate increases, masking the effects on acceptability. It also suggests that certain features occurring in coda position, especially in unstressed syllables, are barely perceptible to these listeners. In general, this speaker displays an ability to use complex language accurately and fluently, and his pronunciation generally shows the avoidance of salient errors (although there are many non-standard features on closer listening). That a relatively high ranking was achieved despite the presence of a significant error (CCRF-PV) may indicate mixed effects for this category.

5.5.5 Speaker 4

Figure 5.22. Transcript of Speaker 4's sample showing the location of errors and the number of identifications.

Well I think² that the very concep⁸t of one country, two systems sugges¹⁶t s the people² of Hong Kong should try to at leas¹²t mai⁶ntain some of their own attribut³es

Figure 5.23. The distribution of Part 2 positive and negative comments for Speaker 4.

Table 5.17. Students' Part 2 comments regarding Speaker 4.

Category	Positive comments	Negative comments	Neutral comments
Sounds		- the word 'concept' isn't clear - 'maintain' seems like 'mantain'	
Connected speech & intonation	- good stress, rhythm	- weird pauses, don't help understanding	
Fluency	-fluent (x 2)	- too fast	
Other	- positive		- it seems that he is able to express in English but seldom does so - LegCo member - like a foreigner! Educated! [sic] - politician - politics - sounds like an NS

Speaker 4 appeared in fifth position in the acceptability rankings. The error density is below average and the speech rate somewhat above average, at 200wpm (one student thought this was ‘too fast’). There are unfilled pauses after the words *people* and *should*; these were termed ‘weird’ by one student. In general, however, this speaker appears to have the attributes of a competent speaker in terms of accuracy and fluency (two comments describe the speaker as ‘fluent’). Pitch span is near the average, but maximum frequency is slightly above average.

There are several error types, but few are in significant categories (with the exception of the VOWEL error in the first syllable of *maintain*, where vowel shortening occurs; this was noted by six students and also appears in the comments). The most frequently marked error is the preconsonantal final CCR in the word *suggests*, with the final /ts/ being deleted (16 mentions). As with the word *relaxed* in Speaker 3’s utterance, this is a triconsonantal cluster occurring in a stressed syllable. CCRF-PC also occurs in *least* (11 mentions) and *think* (2 mentions), while CCRF-PV is found in *concept* (prevocalic; 8 mentions) and *attributes* (prepausal, with deletion of the penultimate /t/; 3 mentions). One of the few identified tokens of L vocalisation occurs with this speaker, in the word *people* (2 mentions).

As mentioned above, speakers in the middle range of the acceptability rankings (5th to 8th position) can be thought of as having contradictory attributes; in this case the speaker profile and student comments suggest a high level of proficiency and some authority (there are comments referring to sounding native speaker-like, as well as to educational level and presumed occupation (a politician). However, there are also errors in significant categories (VOWEL and CCRF-PV). There is frequent simplification of final clusters (all but one of them in this sample are simplified in some way).

5.5.6 Speaker 5

Figure 5.24. Transcript of Speaker 5's sample showing the location of errors and the number of identifications.

Figure 5.25. The distribution of Part 2 positive and negative comments for Speaker 5.

Table 5.18 Students' Part 2 comments regarding Speaker 5.

Category	Positive comments	Negative comments	Neutral comments
Sounds		- missing consonants	- good pronunciation on most words (except word stress 'information')
Connected speech & intonation		- intonation X - poor intonation - no stress, rhythm - no word stress - intonation - no intonation at all - very flat	- heavy HK accent
Fluency			
Other	- clear - easy to understand		- government official - typical HK accent - sounds like a HKer, Cantonese accent

As explained before, Speaker 5 was excluded from the second regression analysis because it was felt that the error in the word *applications* may have been seen as more grammatical than phonological (although the difficulty of separating the two,

especially with final CCR, has been noted). Nevertheless, as the speaker was included in the initial acceptability rankings, an account of the features marked by students is provided here. Speaker 5 appears in the middle to lower levels of the acceptability rankings, in eighth or ninth position. The error density is slightly above average, and most other speaker variables are close to the average; this was one of two public addresses, both of which feature a slightly slower speech rate of around 140wpm. Many student comments refer negatively to the intonation, although there is no evidence for this in the pitch span measurements. There is a possibility that the speaker's tendency towards syllable-timed rhythm influenced these comments, and this may also explain those referring to a 'typical' or 'heavy' Hong Kong accent.

A particular instance of this tendency occurs in the word *information*, which was marked by 16 students as containing a word stress error. The reason for this perception and marking is probably the full vowel in the second syllable, as the overall stress pattern (primary and secondary stress on the third and first syllables, respectively) is close to a standard version. 17 students noted the missing plural 's' in *applications*, and although this instance of final CCR occurs prevocally (CCRF-PV, a significant factor) it may have been more salient simply because it represents a noticeable deviation from orthography; subject-verb disagreement is also noticeable when the verb *are* is used a short time later. (An alternative explanation of the error is that the verb form *are* is incorrect, and that *application* was in fact the intended form.) Other errors in significant categories are the /v/-substitutions (PHONSUB-V) in *diverse* and *varied* (five mentions each). The connected speech feature between the words *it* and *is* (where the /t/ of *it* is glottalised) received ten markings. The remaining errors are in the words *impossible* (L vocalisation, three mentions), *and* (CCRF-PV, one mention), and *varied* (OTHER C-SUB, one mention; there is devoicing of the final consonant).

It therefore seems likely that prevocalic final CCR and the consequent absence of plural marking in *applications*, and /v/-substitution in *diverse* and *varied*, explain the below average acceptability scores to some extent, but the contribution of non-specific suprasegmental problems is probably also significant.

5.5.8 Speaker 7

Figure 5.26. Transcript of Speaker 7's sample showing the location of errors and the number of identifications.

The question we need to ask is: does the public w¹⁰a¹⁴n⁷t⁷ KCRC run like a government⁷ department? MTR run like a government de⁹p¹art¹³men¹t[?] ?

Figure 5.27. The distribution of Part 2 positive and negative comments for Speaker 7.

Table 5.19. Students' Part 2 comments regarding Speaker 7.

Category	Positive comments	Negative comments	Neutral comments
Sounds		- missing /t/ sound	
Connected speech & intonation		- strange intonation - strange to stress all the words - stresses are strange - strange sentence stress - intonation is strange - intonations [sic] are strange	- intonation. Foreigner?
Fluency	- fluent, have intonation		
Other	- very good and clear	- pretend to speak like a foreigner - strange	- Government official. Careful - sound like an Indian - Indian politician - sounds Indian - accent sound like a foreigner - sounds formal with the stress sound

Speaker 7 appeared in the lower middle range of the acceptability rankings, coming ninth in terms of overall acceptability. This came as something of a surprise as the speaker appears to be typical of the carefully-enunciated, British-influenced accents often found among senior civil servants in Hong Kong (the chief executive, Donald Tsang, is another example). The error density measures are above average, and the speech rate slightly below average; however, despite being from a location interview this sample appears to have a semi-scripted, speech-like quality. The pitch span measurement, on the other hand, is the highest among all the samples (this again probably reflects the speaker's choice of language, which makes use of rhetorical questions). Several comments referred to the 'strange' intonation, so it is possible that students were unfamiliar with, and did not respond well to, this manner of speaking.

Looking at the error markings, there were several problems with final consonants. The missing /t/ in *want* had fourteen mentions; another instance of CCRF-PC was *government* with seven mentions (the word occurs twice and there is CCR in both, but the above transcript shows the seven mentions in the first occurrence only). CCRF-PV occurs in the second *department*, and received thirteen mentions. Other consonantal features noted were the medial /t/ in *department*, which is glottalised, and the /p/ of the same word, which was perceived and marked as a [b] by nine students; there is indeed some reduced aspiration here. This idiosyncratic instance of OTHER C-SUB may be related to the comments above about sounding 'Indian', as unaspirated voiceless plosives are a feature of Indian English (Collins and Mees, 2003: 169). There was one VOWEL marking, relating to the word *want* where some anticipatory nasalisation is present (the vowel segment was marked by ten students). From an overall perspective, it seems that the presence of significant error types (CCRF-PV and VOWEL), plus the unfamiliar intonation patterns, resulted in the below-average acceptability rankings.

5.5.9 Speaker 8

Figure 5.28. Transcript of Speaker 8's sample showing the location of errors and the number of identifications.

Th¹ere are many children who are not as privilege^d¹⁹ as we would like to think they should be

Figure 5.29. The distribution of Part 2 positive and negative comments for Speaker 8.

Table 5.20. Students' Part 2 comments regarding Speaker 8.

Category	Positive comments	Negative comments	Neutral comments
Sounds			
Connected speech & intonation		- no intonation - intonation too flat	
Fluency	- fluent	- the pause is a bit strange	- smooth, but unusual structure
Other	- a very clear speech - positive - clear, easy to understand - clear pronunciation, but doesn't sound like a NS - clear pronunciation - a very clear speech - sounds OK		- sounds professional and old - a very typical HKer - LegCo member

Speaker 8 came third in the acceptability rankings. In terms of overall acceptability rankings this speaker received a very similar rating to the native speaker (Speaker 11), despite probably not sounding like a native speaker to most students (she was ranked 7th out of 12 for 'sounds like a Hong Kong speaker', while Speakers 10 and

11 were clearly considered to be the least Hong Kong-like according to this item). One of the student comments above also provides corroboration of this ('clear pronunciation but doesn't sound like a NS'). Other positive comments often feature the word 'clear' (five instances, for example 'clear, easy to understand'; 'a very clear speech'). Negative comments relate to suprasegmental features ('no intonation'; 'intonation too flat'; 'the pause is a bit strange').

Speaker 8 also has a below-average error density (2.89 errors per word, ranked sixth), but the speech rate is close to the average. No measurements of pitch span were taken for the two female speakers. Of the errors noted on the transcripts, CCRF-PV involving the elision of the final /d/ in *privileged* was marked by 19 students, with TH stopping in *There's* being the only other feature (noted by one student). As CCRF-PV emerged as a significant feature in overall terms, the case of *privileged* raises the interesting possibility that while final cluster simplification is clearly common among Hong Kong speakers, and while this does not always follow NS patterns (here, the prevocalic context would generally be considered to inhibit reduction), Hong Kong listeners do not always see this as a stigmatised feature. This may of course depend on the type of simplification and the phonological context. Among the possible factors that might explain why this was so in this particular case, the most promising appears to be occurrence in an unstressed syllable. The bimorphemic status of the word may also be implicated. The word *relaxed* in Speaker 3's utterance contained another prevocalic, bimorphemic cluster whose reduction did not seem to have an adverse effect on the acceptability ratings, even though it occurs in a stressed syllable (unlike in *privileged*). As *relaxed* and *privileged* are both past participle forms, this suggests that there may also be differences between bimorphemic contexts (for example, loss of plural marking may be seen as being more severe).

5.5.9 Speaker 9

Figure 5.30. Transcript of Speaker 9's sample showing the location of errors and the number of identifications.

Th⁶ **accredited**²² fish farm scheme aim⁶ S⁶ at assisting the
local fish farmers to enhance th² eir operation and pr⁵ oduction standa² r d¹² s

Figure 5.31. The distribution of Part 2 positive and negative comments for Speaker 9.

Table 5.21. Students' Part 2 comments regarding Speaker 9.

Category	Positive comments	Negative comments	Neutral comments
Sounds		- enhance/operation/ production s/ʃ - the consonants are not accurate	
Connected speech & intonation		- flat intonation - very flat - maybe more intonation	
Fluency	- speak smoothly		
Other		- a bit hard to follow	- sweet voice (x 2) - sounds like a typical govt. official - sounds like Hong Kongers - student making presentations

Speaker 9 was ranked seventh out of twelve in both acceptability rankings. The error density is above average, and one of the comments above (‘the consonants are not accurate’) suggests a possible reason for this. The speech rate is slightly below average, but as with Speaker 7, this appears to be a semi-scripted utterance with some of the qualities of monologic discourse (this was also a location interview). There are several comments about ‘flat’ intonation; as this is a female speaker, no pitch span measurements were taken.

The most frequently marked word was *accredited*, with 22 mentions, and the nature of the markings suggested that the error belonged to the SYLL category. As has been mentioned, some local transcript comments alluded to the fact that not all of the syllables were clearly pronounced (one student appended the ‘pronunciation spelling’ of *agreded* to the transcript). The comment above about being ‘a bit hard to follow’ may also relate to this word. However, there is no sign from the speech rate that the overall rate of speaking was too fast. Given that SYLL was found to be a significant factor overall, this problem is probably the most important influence on the acceptability ratings. By this stage of the speaker analysis, it appears that the presence of SYLL errors is a reliable predictor of low acceptability. There is also utterance-final CCR in *standards* (classified as CCRF-PV), where the /d/ is omitted (12 mentions), OTHER C-SUB in the form of the devoicing of /z/ in *aims* (6 mentions) and an unreleased, weakly articulated final consonant in *accredited*, and vowel substitution in the opening word, where a schwa is used instead of the prevocalic allomorph /ði/. The non-significant errors include TH stopping in *The* (6 mentions) and *their* (2 mentions), and the use of full vowels in the first syllable of *production* (6 mentions) and the second syllable of *standards* (2 mentions).

5.5.10 Speaker 10

Figure 5.32. Transcript of Speaker 10's sample showing the location of errors and the number of identifications.

Actually I ha¹ve been with the party for a long time, ten years to be exac¹²t, but I have been serving mostly as central committee member and¹ standing committee member

Figure 5.33. The distribution of Part 2 positive and negative comments for Speaker 10.

Table 5.22. Students' Part 2 comments regarding Speaker 10.

Category	Positive comments	Negative comments	Neutral comments
Sounds			
Connected speech & intonation	<ul style="list-style-type: none"> - did an excellent job in intonation and rhythm - very natural! The intonation is good too - good intonation - good linking, fluent - good stress and intonation 		
Fluency	<ul style="list-style-type: none"> - fluent - fluent and good speech 		
Other	<ul style="list-style-type: none"> - quite good (x 2) - good! - good pronunciation - positive - excellent! - excellent - Clear pronunciation, I like the way that this speaker sound very natural - natural 		<ul style="list-style-type: none"> - a bit flat, but sounds fluent and native - American accent - high education standard - sounds like a foreigner - a foreign speaker, well [sic] pronunciation - sounds like a foreigner, professional, confident - political party member quite strong American accent - professional - HK people speak like a foreigner - he doesn't receive any education in HK?

Speaker 10 appeared at the top of the acceptability rankings. This was one of the two 'native speaker' accents, but Speaker 10 has L1 Cantonese and was largely educated in Canada. However, he moved away from Hong Kong at around fifteen years of age, and has a perceptible North American accent. It is not entirely clear how the students pictured him in terms of ethnicity and language background. There were no negative comments, and the positive comments often referred to suprasegmental aspects such as fluency (four comments), intonation, stress and rhythm. The 'naturalness' of his speech was also noted by two students, although it is not entirely clear what they meant by this.

The error density is the lowest among all the samples and the error codings mainly refer to the final cluster simplification in the word *exact* (12 mentions). However, on careful listening it seems that the final /t/ of *exact* is not actually deleted, but has a very gentle release so that it is almost inaudible; this is a common feature of North American English accents (Collins and Mees 2003: 136). Other features noted were

the virtually inaudible final consonant in *have* (in both instances) and the final CCR in *and*. Both of these only received one mention. An interesting fact about this speaker is that although the speech rate was the highest of all, at 222wpm, this was achieved without any adverse comments about clarity or markings of SYLL errors. Perhaps this balance between speed and clarity is what the students mean by (for example) ‘[c]lear pronunciation, I like the way that this speaker sound very natural’ and ‘fluent and good speech’. The importance of controlling salient segmental features at higher speech rates is demonstrated by the low rating given to Speaker 6.

In fact, in this extract there is almost nothing to distinguish his accent from that of a native speaker. Among the noteworthy (but unmarked) features there is TH-stopping in *the (party)*, but the negotiation of *with the party* using a dental, rather than interdental, fricative in *with* shows a nativelike connected speech process. The /t/ in *mostly* has almost entirely disappeared, as it would with many speakers. There is rhotic /r/ in *for* and *party*. The rhythm is noticeably stress timed, especially in phrases such as *standing committee member*. Connected speech processes including weak forms and linking are used in *but I have been* (weak form of *but*, linking in *but I*, near-elision of the final consonant in *have*) and *as central* (weak form of *as*, with the consonant being co-articulated with the following /s/ of *central*). The use of pitch, volume and pausing to accentuate key words such as *with*, *party*, *time* and *serving*, as well as other content words, is effective and probably contributed to the student perceptions of both comprehensibility and naturalness. Comments such as ‘did an excellent job in intonation and rhythm’ and ‘good stress and intonation’ suggest that suprasegmental features were a positive influence here.

From the speaker’s words it is clear that he is a member of a political party and this may have increased his prestige for the students, but the importance of linguistic factors is also indicated. Apart from the high level of accuracy and fluency, the effects of pitch span also may also help to explain the popularity of Speaker 10. His low minimum pitch (around 85 Hz) was the lowest recorded, and perhaps this created an overall impression of calmness and authority (‘professional’ and ‘confident’ were among the comments).

The popularity of this speaker tends to confirm the prevailing view regarding Hong Kong students' preferences for exonormative models. In reality, few people in Hong Kong have access to the kind of international education and exposure that Speaker 10 has received, but at the moment this does not stop it from being seen as an appropriate model for pedagogical applications. While it could be argued that the ratings arise from the possible perception of him as a native speaker, and the stereotypical associations it may have, there may also be linguistic principles associated with this preference (for example, the speaker's 'errors' occur in non-salient contexts). Also, given that the students had access to the transcript during Part 2 of the survey procedure, their perceptions of 'clear' speech may be related to the absence of noticeable departures from the written forms (bearing in mind, of course, that students at this level have acquired some knowledge of sound-spelling correspondences in English).

5.5.11 Speaker 11

Figure 5.34. Transcript of Speaker 11's sample showing the location of errors and the number of identifications.

The quality migrant³ attraction scheme seeks to attract talented⁴ people and also talented people to bring their families with³ them

Figure 5.35. The distribution of Part 2 positive and negative comments for Speaker 11.

Table 5.23. Students' Part 2 comments regarding Speaker 11.

Category	Positive comments	Negative comments	Neutral comments
Sounds	- clear consonants pronounced /t, l, s/		
Connected speech & intonation	- placing good stress and unstress - clear and good stress - with stress patterns - good intonation and pronunciation - has weak forms - good intonation (x 2)	- maybe more linking - no weak forms	
Fluency	- fluent - fluent and exact	- too fast, not enough pausing	- too quick but very fluent and native - British English - sounds like a foreigner - high social status - government official and foreigner (but permanent resident)
Other	- clear pronunciation - quite good - excellent! - clear! - good overall - good (x 3)	- pretend to speak like a foreigner - tend to pretend foreigners [sic]	- I think he has a British accent with very clear ending sounds, especially the word 'people' - sounds like a foreigner - don't like his voice

As in the case of Speaker 5, Speaker 11 was not included in the second regression analysis but as a native speaker with a south-eastern British English accent he provides an interesting comparison with the other speakers. Once again, the apparent exonormative preferences of the students are confirmed by the high ratings given to this speaker (he appeared in second position in the rankings). It seems likely that many students identified him as a native speaker; he received the lowest ranking on the 'sounds like a Hong Kong speaker' question, and two students actually identified his place of origin. Interestingly, while another student thought he 'sounds like a foreigner', two others were less certain about his identity ('pretend to speak like a foreigner'; 'tend to pretend foreigners' [sic]). The error density is below average, as is speech rate, and pitch span is slightly below the norm. As with Speaker 10, the comments praise both clarity and fluency (for example, 'fluent and exact'), although two found that he spoke too quickly.

While there were few errors marked (the speaker ranked second lowest in terms of error density, as measured by the original student markings), the word that attracted most attention was *talented* (four mentions, but the word appears twice). One student provided the transcription /tɪt/ to show his or her impression of it in the first occurrence. The final syllable has an unreleased final consonant, a linking phenomenon which is very likely in this context because the next word begins with a consonant (*talented people*). There is also considerable vowel reduction in this unstressed syllable, and the final /d/ is devoiced. Although the speaker's accent might seem quite 'standard' to a British audience, it is likely that all of these features contributed to an unfamiliar-sounding pronunciation for the students. Four students underlined the ending of *migrant*, where there is glottalisation of /t/. Three students noted TH stopping in *them*; both of these are mentioned as aspects of Cockney by Trudgill (1984: 57), but can be seen as features of the London and south-eastern British English accent continuum (Altendorf 2003).

Of the features not noticed by the students, once again the vocalisation of dark /ɪ/ in *people* (especially the second instance of this word) is prominent, lending further support to the idea that this feature is likely to be inconsequential in terms of listener reactions. In fact, one student wrote 'I think he has a British accent with clear ending sounds, especially the word 'people''. While there may be some properties of his realisation that influenced this comment, once again the vocalisation of postvocalic /ɪ/ does not seem to be either noticeable or significant for these students.

5.5.12 Speaker 12

Figure 5.36. Transcript of Speaker 12's sample showing the location of errors and the number of identifications.

Figure 5.37. The distribution of Part 2 positive and negative comments for Speaker 12.

Table 5.24. Students' Part 2 comments regarding Speaker 12.

Category	Positive comments	Negative comments	Neutral comments
Sounds			
Connected speech & intonation		<ul style="list-style-type: none"> - no obvious stress in sentence - no intonation - very flat 	
Fluency			
Other		<ul style="list-style-type: none"> - generally poor in pronunciation - low intelligibility - lazy speaker - many pronunciation mistakes 	<ul style="list-style-type: none"> - sounds like HK speaker - LegCo member - strong HK accent - a typical HK speaker - HK people

Speaker 12 appeared at the bottom of the acceptability rankings. The high error density is almost certainly the main reason, as indicated by the above comments, although suprasegmental features may also have contributed. The pitch span is actually close to the average, despite the comment about 'flat' intonation, and the speech rate is also close to the average. As shown by the transcript above, the final CCR in the word *confused* (categorised as CCRF-PV) was the most frequently marked error among all the speakers and tokens. Its salience may have been increased by the combined effects of its utterance-final position and its occurrence in a stressed syllable; other examples of final CCR in this study suggest that its bimorphemic status was less important, although this is in need of further investigation. Once again, the question of whether this was seen as a phonological or a grammatical error needs to be considered. In terms of the students' perception of

the error, it seems likely that there were elements of both. However, from this perspective both types of error have in common the fact of being deviations from an expected form. With some types of error, for example final CCR, acceptability can thus be seen as being affected by both phonological and grammatical deviation; perceptions of one may reinforce the other. This does not appear to be a serious problem for the present study, although the influence of perceived grammatical deviation must also be acknowledged.

Another significant feature, the PHONSUB-V or /v/ substitution in *virtually*, was noted by 13 students. Vowel substitution was marked in the first syllable of *democracy* by three students; it has a ‘spelling pronunciation’ with a vowel closer to [ɛ] than [ɪ] or [ə]. Among the non-significant features, there is a full vowel in the first syllable of *confused*. Preconsonantal final CCR occurs in *found*, as noted by 19 students, and in fact the remaining /n/ sound is not clearly articulated. Finally, there is TH stopping in *there’s*, noted by two students. A prominent feature that passed unnoticed is the initial cluster simplification in *progress*, pronounced as [ˈpouɡɪes].

5.6 Summary of the findings

To conclude this chapter, and before proceeding to an explanatory discussion in the next chapter, the findings of the main study will be briefly summarised. The measures of inter-rater reliability showed that there was a high level of rating consistency among the 52 student listeners. The intraclass correlation coefficients (ICCs) indicated that this consistency was lower among students with a lower pronunciation skills self-rating, but neither this nor the other measured student variables (gender and major field of study) had a marked effect on the rating consistency, and all of the ICCs were above the 0.7 threshold. An analysis of the inter-item correlations between the questionnaire items indicated that the internal consistency of the scale was high, and that the five items used to obtain the ‘overall acceptability’ measure represented aspects of the same general attribute.

Turning to the speaker variables, the first regression analysis showed the relative importance of phonological error when compared with other measured speaker variables (in the general categories of speech rate, syntactic complexity and prosodic factors such as pitch span). However, the predictive ability of the regression model was modest, and the influence of other, unmeasured variables should not be discounted. For example, an unmeasured linguistic variable was the degree of syllable- or stress-timing in the samples. Similarly, the regression analysis of the error codings indicated the significance of certain phonological features, but the influence of other linguistic features must be acknowledged. The features that exerted significant and unique effects on the acceptability scores were, in order of correlation, SYLL (syllabic modification), CCRF-PV (final consonant cluster reduction in prevocalic or prepausal contexts), PHONSUB-V (a consonantal substitution), VOWELSUB (vowel substitution) and OTHER C-SUB (other consonantal modifications).

Despite the modest fit of the regression model, it provides useful data about the relative acceptability of different phonological features. However, a useful complement to this analysis is a simple ranking of the speakers according to overall acceptability scores, noting the presence or absence of the significant features in the student error codings (Table 5.25, below). Not all of the significant features were effective ‘predictors’ of low acceptability, and the three that pattern most consistently are SYLL, PHONSUB-V and VOWELSUB. None of these features appeared in the upper quartile. PHONSUB-V appears to be an effective predictor of low acceptability scores; it appears three times and is present in the two lowest-ranked samples. VOWEL SUB is also more prominent in the speakers in the lower part of the table, and the four lowest-ranked speakers all have this feature.

Table 5.25. The distribution of three significant errors according to acceptability rankings.

Speaker (ranked by overall acceptability)	Number and type of significant errors (PHONSUB-V, SYLL, VOWELSUB)
10	0
11	0
8	0
3	1 SYLL
4	1 VOWEL SUB
2	0
9	2 VOWEL SUB, SYLL
5	1 PHONSUB-V
7	1 VOWEL SUB
6	2 SYLL, VOWEL SUB
1	1 PHONSUB-V
12	2 PHONSUB-V, VOWEL SUB

Thus none of the speakers in the upper quartile showed any of these three significant features, while all of the speakers in the lower quartile had at least one. The other two significant features, CCRF-PV and OTHER C-SUB, did not pattern in quite the same way. Speaker 8's sample featured CCRF-PV (in the word *privileged*), but this did not prevent her from obtaining the third-highest overall ranking. This indicates that not all of the error categories in the study had a similar level of internal similarity, and that some tokens of these categories had characteristics that meant they did not affect acceptability in the expected way.

The distribution of positive and negative comments in Part 2 of the survey form reflected the acceptability scores to a large extent. The nature of the comments helps to explain some of the results for individual speakers, both in general terms and with regard to particular features. The comments suggest that suprasegmental features such as intonation were seen as important, although it is not clear whether the number of comments about this was related to the survey design, which made it easier to mark local, segmental features. Suprasegmental features were thus not included in the second regression analysis, and these may be another unmeasured linguistic factor that affected the acceptability scores.

CHAPTER 6

DISCUSSION

6.1 Introduction

This chapter aims to provide an explanation of the findings. As one of the ultimate aims of the study is to make pedagogical recommendations, a thorough consideration of the possible reasons for the acceptability ratings is necessary. A challenge for any explanation of language variation, and the effects of this variation, is to integrate diverse but potentially complementary perspectives, such as synchronic *versus* diachronic, individual *versus* social, and internal *versus* external factors. The explanatory approach taken here seeks to achieve this partly by integrating the findings of both the preliminary and the main study. The preliminary study can be seen as a synchronic indication of the way certain phonological features tend to pattern within and between speakers, as well as providing an overview of the relative frequency of occurrence of these features. The implicational scale also has a diachronic dimension, in the limited sense of reflecting developmental processes (such as those outlined by Hansen 2006), rather than its usual sense of historical change. The patterns of variation may indicate typical paths of development for speakers, and show the occurrence of transfer and developmental features at different points along these paths. (As mentioned in Chapter 3, an additional explanation is provided by intra-speaker or stylistic variation, and the constraints that shape this variation.) The main study can also be approached from these different angles, as it shows how phonological features that seem to represent different stages of development (or different types of stylistic variation) are perceived in real time by listeners – listeners who are themselves participants in similar developmental and variational patterns.

In general, the explanatory approach taken here will seek to combine insights from various perspectives, following Bailey's (1996: 378) position that 'deep explanation

and prediction are possible only by investigating and understanding how structures and other phenomena have developed into what they have become'. Briefly, the acceptability data will first be examined from a synchronic, individual, language-internal perspective, in order to see whether factors such as phonetic difference can explain the acceptability ratings. This is then expanded into a broader consideration of how acceptability relates to language variation. This in turn involves a more diachronic, social and functional perspective that focuses on the nature of Hong Kong English as a variable system, one that encodes the paths of development and preferences of its users. Evidence from the structural properties of other varieties of English will also be used to inform and strengthen the explanatory model. Given the multiplicity of factors that affect the development of varieties, an 'ecological' metaphor (Mufwene 2001) appears to be the most appropriate. For the purposes of this study, it also allows the discussion to be related back to the phonological features. By conceptualising a 'feature pool' in which multiple interacting factors decide which features tend to persist, an ecological approach allows for the synthesis of various perspectives without privileging any particular one (in line with Bailey's proposed approach to synchrony and diachrony).

6.2 Towards an explanatory model

The initial question that needs to be answered can be put simply as: why were certain phonological features rated more negatively than others? The answers to this apparently simple question require an explanatory model of some complexity. The first approach to the question will be to draw attention to the apparent correspondence that exists between the intelligibility and acceptability characteristics of the phonological features examined in this study.

6.2.1 The intelligibility-acceptability correspondence

The apparent intelligibility-acceptability correspondence is not only of interest in itself, but it also establishes the need for a multidimensional or ecological perspective

on the data and on the research topic in general. In Chapter 5 it was concluded on the basis of the statistical data that the five phonological features with significant effects on the acceptability scores were:

- SYLL;
- CCRF-PV;
- PHONSUB-V;
- VOWELSUB; and
- OTHER C-SUB.

Applying the criteria for intelligibility derived from empirical studies (e.g., that of Jenkins 2000) and listed in Chapter 2, it is clear that all of these features would be expected to affect international intelligibility. SYLL is an error category developed in this study, and in Chapter 5 it was defined as ‘excessive vowel reduction or consonantal modification as a result of a rapid speech rate’. On this basis, and considering the actual examples encountered in this study, it is arguable that SYLL errors represent noticeable vowel and/or consonant modifications and hence have consequences for intelligibility. Student comments regarding these errors (for example, ‘the word “political” is very unclear’) seem to confirm their intelligibility-reducing (henceforward, IR) nature. This also suggests that another source of variation is the hypo- to hyper-articulated continuum, from rapid (but less clear) speech to slower (but clearer) speech. This may be negotiated in different ways by different speakers. It is possible that native and non-native speakers may differ in this regard, but the data from this study suggest that certain types of connected speech processes may reduce acceptability for some audiences, regardless of whether they occur in NS or NNS speech.

CCRF-PV would also be predicted to reduce intelligibility, on the basis of these criteria. Jenkins (2000) distinguished between acceptable (native speaker) patterns of final CCR and their non-acceptable counterparts, suggesting that CCRF-PC is less likely to reduce intelligibility than CCRF-PV. These are problematic distinctions, not least because patterns of final CCR are similar across all varieties of English, regardless of whether they are ‘native speaker’ varieties or not (see, for example,

Schreier 2009). However, if it is accepted that prevocalic and prepausal environments generally inhibit final CCR in most varieties of English, the findings of this study regarding the acceptability of CCRF-PV also coincide with the claimed intelligibility characteristics of this type of ‘non-native speaker’ final CCR; low acceptability (in the sense of having significant effects on the ratings) coincides with low intelligibility (in the sense of being deviant core features). However, it is not entirely clear why CCRF-PV should affect intelligibility more, as this aspect of final CCR is seemingly unrelated to information-carrying considerations. The reduction of bimorphemic clusters would, on the other hand, be expected to reduce intelligibility by reducing the amount of grammatical information. It seems probable that Jenkins’ distinction between NS and NNS patterns actually encapsulates a range of intelligibility characteristics. While CCRF-PV does not appear likely to reduce intelligibility merely by virtue of being prevocalic, its presence may be symptomatic of a particular positioning on one or more continua of variation. This positioning may also include IR features such as the reduction of bimorphemic clusters.

PHONSUB-V is also an IR feature, according to Jenkins’ data, by virtue of being a consonantal modification (or substitution, if replacement with [w] actually occurs). The reasons for this are probably related to information-carrying factors such as functional load, as argued in Chapter 2, section 2.6.7, and to the phonetic difference of the substitution. Moving down the list of significant features, vowel errors such as VOWELSUB also appear to be IR features. In Jenkins (2000) it was suggested that vowel length modifications were more important than vowel quality modifications. Two of the three tokens of this error category actually did involve length modifications (the vowel shortening in *maintain* and *leadership*). The quantity/quality distinction has already been noted as problematic, but here the significance of the VOWELSUB category supports the general case for the importance of vowel errors. Vowel modifications involving the non-reduction of full vowels in unstressed syllables (FULL VOWEL) were not found to be significant, on the other hand. This would be seen as a non-core feature in the LFC.

Finally, OTHER C-SUB comprises several types of error that involve consonantal modification such as devoicing, deletion, and substitution. Jenkins’ (2000: 159) list

of core features includes the stipulation that ‘certain approximations are not permissible...where there is a risk that they will be heard as a different sound from that intended’. The most frequent tokens involve either non-systematic, apparently idiosyncratic substitutions or realisations ([d] for /t/ in *built up* and unaspirated [p] for /p/ in *department*), or devoicing and hence substitution in final clusters, as in *cards* and *aims*. These two examples do not involve deletion, however, and generally the OTHER C-SUB errors in this study seem unlikely to affect intelligibility. The devoicing that occurs in pronouncing *aims* as [eims] rather than [eimz] occurs in word-final position and is thus likely to be less important for intelligibility; the widespread occurrence of terminal devoicing in many varieties also suggests that this feature may have different intelligibility and acceptability characteristics when compared with other substitutions. As was mentioned in Chapter 5, section 5.6, OTHER C-SUB is a diverse category and its constituents may have different effects on acceptability, as well as being potentially variable in terms of their effects on intelligibility.

In many cases, despite the intelligibility criteria it seems unlikely that there would be actual reductions in intelligibility. But disregarding this caveat for the time being, and acknowledging the need for further and more detailed research, in all five of these categories low acceptability equates with low intelligibility. Turning to the phonological features that did not have significant consequences for acceptability, the same correspondence is visible. In this case, being inconsequential for acceptability equates with being inconsequential for intelligibility. If the features are ranked in descending order of the standardised beta coefficients in Table 5.12, the following list is obtained:

- CCRF-PC (preconsonantal final CCR; beta = - 0.130, sig. = 0.002);
- TH-STOP (TH stopping; beta = - 0.101, sig. = 0.011);
- TH-FRONT (TH fronting; beta = - 0.087, sig. = 0.030);
- PHONSUB-R (/r/-substitution; beta = - 0.055, sig. = 0.179);
- FULVOWEL (full vowels in unstressed syllables; beta = - 0.052, sig. = 0.195);

- CSLINK (linking phenomena in connected speech; $\beta = 0.035$, $\text{sig.} = 0.382$); and
- LV (L vocalisation or deletion; $\beta = -0.006$, $\text{sig.} = 0.882$).

All but one of these features would be expected to be inconsequential for intelligibility, according to Jenkins' criteria. The exception to the intelligibility-acceptability correspondence is PHONSUB-R, which would be classed as an IR feature along with PHONSUB-V, but did not significantly affect acceptability in this study. Among the possible reasons for this are the lower number of codings (17, compared with 30 for PHONSUB-V) and the fact that all the identified tokens occurred with the same speaker (in the Speaker 1 and Speaker 6 samples, representing the same person). While this would not necessarily limit the ability of the regression procedure to detect patterns across cases, it might have affected rater attention. In the Speaker 6 sample, the token of PHONSUB-R competes for attention with other noticeable errors such as SYLL.

A general correspondence between the intelligibility and acceptability characteristics of features can thus be observed. Before discussing the reasons for this in more detail, an initial explanation that relates to the nature of the questionnaire items needs to be considered. Included in the 'overall acceptability' measure is item C, relating to 'ease of understanding'. It could be contended that the construct of 'acceptability' already includes intelligibility, or at least perceived intelligibility, as part of its make-up, and that measurements of the two characteristics will thus tend to correspond. However, it has already been argued that the concept of 'acceptability' naturally includes elements of intelligibility, especially when (as in this study) the former concept is interpreted as meaning 'acceptability for pedagogical purposes'. There is also an aspect of the survey design that makes this correspondence less likely to be merely an artefact of the methodology: the survey form required students to make global assessments of the accent samples using the Likert scales before they marked the errors that affected their judgements. But even if some errors were perceived as being significant both at a 'macro' level (in Part 1 of the questionnaire, thus potentially conflating intelligibility and acceptability) and at a 'micro' level in Part 2

of the questionnaire, the intelligibility-acceptability correspondence still seems likely to represent an actual relationship between the two concepts, at some level.

It will be recalled that at a feature level, one of the main determinants of intelligibility proposed in Chapter 2 was the functional load of the sounds concerned. However, functional load by itself does not satisfactorily explain why ratings were significantly influenced by intelligibility-reducing (IR) features. The study of Munro and Derwing (2006) involved L1 Cantonese English speakers and native speaker listeners, and found that errors with a high functional load (such as the conflation of [n] and [l]) consistently led to harsher accent ratings and lower perceived comprehensibility scores than those with a low functional load, such as dental fricative substitutions. In an earlier article the study's authors point out that accentedness and intelligibility are 'related but partially independent dimensions' (Derwing and Munro 1997: 2), but the findings of the 2006 study suggest that as errors with a high functional load reduced both accentedness and comprehensibility ratings, accentedness and intelligibility were in fact perceived as being related by the listeners (assuming that the listeners believed 'intelligibility' and 'comprehensibility' to be related). The possible reasons for the greater effect of high functional load errors are not considered in detail by Munro and Derwing (2006). In the present study, it seems unlikely that the students made any conscious linkages between IR features, high functional loads and low intelligibility, and a more plausible explanation must be sought.

6.2.2 Synchronic perspectives: the salience of features

If a synchronic, individual, perspective is taken, one that focuses on the students' reactions to the accent samples and features they heard, the likely explanation is simple and perhaps even obvious: intelligibility and acceptability are linked and partially explained by the same underlying principle, that of salience. Of course, it could be argued that there is no such thing as a purely synchronic perspective, as the language systems of both speakers and listeners represent accumulated language knowledge, and languages are themselves the product of centuries of language use by others. But it is still a useful exercise to imagine the sounds being received and

perceived *de novo*. This is the perspective more likely to be taken by formal, as opposed to functional, approaches to linguistic explanation (for example, the ‘evolutionary’ model of sound change proposed by Blevins (2004), in which the main explanatory factor is the inherent ambiguity that can occur in acoustic auditory signals). This section considers the extent to which salience can explain the acceptability of phonological features, as well as their intelligibility characteristics.

The concept of salience is somewhat difficult to define, however. Kerswill and Williams (2002: 105) note that ‘there are no necessary and sufficient conditions which must be met in order for a feature to become salient’, except for the ‘obvious’ one of the difference between presence and absence actually being noticeable in a ‘psychoacoustic sense’. The same authors (pp. 86-87) identify several criteria for ‘salience’ in the literature, among them being language-internal factors such as frequency (high frequency items are salient; see Cheshire 1996) and prominence (phonetic features in prosodically prominent positions are salient). This is because phonological elements in prosodically prominent positions are ‘more forcefully and fully articulated than elements in prosodically weak positions’, according to Wilson and Keil (2001: 680); prosodic prominence is therefore greater in initial position (Martinez-Gil and Colina 2006: 225) and in stressed syllables (Goldsmith 1996: 4). Kerswill and Williams (2002: 91) report that degree of phonetic difference and involvement in phonological contrast are at the core of the ‘salience notion’, according to Trudgill (1986).

Adopting four components of salience (frequency, prosodic prominence, phonetic difference and participation in phonological contrast) for the purposes of this study, it appears that the low-acceptability HKE features have a relatively higher salience profile. For example, VOWELSUB can be seen as being more salient than FULL VOWEL because it occurs in stressed syllables, increasing its prosodic prominence. It can also generate phonological contrast, whereas FULL VOWEL cannot. Similarly, final CCR in prevocalic contexts (CCRF-PV) is likely to be more salient than preconsonantal final CCR (CCRF-PC), because of the combined effects of prosodic prominence and phonetic difference. If the final /t/ of *post* is elided in *post office* [pəʊs ɒfɪs] the remaining [s] becomes more noticeable than it would be in the

elided form of *post box* [pəʊs bɒks], because of the tendency for final consonants to be ‘carried over’ to the onset of the following syllable.

When considering phonetic difference as a factor affecting listener judgements, it has to be explained in terms of the difference between listeners’ ‘expected’ forms, based on previous encounters with written and spoken forms, and the actual form (the notion of ‘expectation’ is discussed in more detail below). Regardless of how the concept is applied, it is often hard to quantify the degree of phonetic difference (see, for example, Wells 1982: 44; Heggarty, McMahon and McMahon 2005). Intuitively, there is a greater difference between [w] and [v] (voiced labiovelar approximant and voiced labiodental fricative, respectively) than there is between [w] and [r] (the latter being a voiced palato-alveolar approximant). Thus, one would expect /v/-substitution to be more salient than /r/-substitution, as suggested by the results of this study (although the limited number of tokens of /r/-substitution makes a proper comparison difficult). The use of a distinctive features approach may sometimes assist, and in this case the opposite conclusion is indicated: [w] and [r] differ by more features (such as labial, anterior and coronal) than do [w] and [v]. But the quantification of phonetic difference by this method is also problematic, and is directly challenged by Fant (1969: 6): ‘one cannot expect the phonetic difference between any two phonemes to be proportional to the number of features by which they differ’.

In terms of phonological contrast, the number of minimal pairs involving /r/ and /w/ is considerably greater (213, according to Higgins (2008), as opposed to 58 for /v/ and /w/). The combined effects of these characteristics on listeners are hard to assess, and in the end it has to be recognised that such cognitive or psycholinguistic factors interact with each other and rarely lead to any kind of optimisation within language systems (Schneider 2007: 98). Saliency is thus a difficult concept to operationalise, partly because its components are not independent of each other. Nevertheless, looking at the differences between significant and non-significant features, some appeal to the notion of saliency appears to be justified. There may well be other dimensions to ‘saliency’ than the linguistic ones discussed here; features might also act as emotional triggers (see, for example, the ‘somatic theory’ of Robinson (2003), a linguistic parallel of Damasio’s (1994) somatic marker hypothesis). But even so,

there will still be some linguistic or acoustic element – features are less likely to be triggers if they are difficult to hear.

Salience cannot really explain, except in a very general sense, the significance of errors that involve the deletion of sounds, such as SYLL or final CCR. However, the deletion of sounds is by its very nature likely to be psychoacoustically salient for the listener, because it involves a noticeable departure from orthography. One of Trudgill's (1986) other criteria for salience is that 'the variable has a high-status variant reflected in the orthography' (in Kerswill 2002: 686). Trudgill's criteria are mainly concerned with dialectal accommodation, but in environments where much of the language input comes from the written word, as in Hong Kong, these features are likely to be salient because they appear to be departures from the 'expected' form.

The notion of 'expectation', especially in hypothetically 'new' varieties of English, is somewhat problematic. As well as having considerable exposure to written forms, learners are exposed to different pronunciations of words; some may have more characteristically local representations than others. However, it can be concluded on the basis of the data that there are parameters of acceptable variation, beyond which judgments of acceptability may be affected. This idea resembles Sapir's (1921) contention that variation at the individual level is kept from rising to 'dialectic importance' by a silent 'consensus of usage' within the speech community (in Jones and Singh 2005: 9). Again, the notion of a 'speech community' can be challenged in that attitudes are unlikely to be stable or homogeneous, but nevertheless sound deletions seem unlikely to become speech community norms unless they do not interfere with the 'core function' of language, that of 'carrying complex information from the speaker to the hearer' (Schneider 2007: 89). The idea of 'acceptable deviation' also makes sense if the nature of phonological systems is considered from an information theory perspective. If categories overlap with each other, information-carrying capacity is reduced and some adjustments may be needed.

On the other hand, the significance of SYLL errors could also be a result of listeners being overly preoccupied with the written form. They may not appreciate the differences between citation forms and connected speech, and may be unaware that

certain types of deletion are extremely common – in some cases, effectively obligatory – in the speech of native speakers. The pedagogical implications of this possibility are discussed further in Chapter 7, section 7.4.1. However, it should also be noted that the SYLL errors marked in this study may have characteristics that place them outside the normal range of variation for most listeners, even those who are familiar with connected speech processes.

6.2.3 Salience and stigmatisation

At this point it may be worth clarifying the difference between the concepts of salience and stigmatisation, both as theoretical constructs and as factors that need to be considered when interpreting the results. According to Collins and Mees (2003: 142), salient features are those which are known and remarked upon by local members of the community. Although this definition may apply mainly to dialectology, rather than to L2 varieties of English, it still carries the basic sense of noticeability. This study has shown that certain HKE features are ‘known and remarked upon’, at least with these listeners, and with a certain amount of prompting. However, while the patterns of error codings appear to reflect salience, the features may differ in terms of their degree of stigmatisation. A stigmatised accent characteristic is one which has low status, and thus is the subject of social disapproval (Collins and Mees 2003: 142). Stigmatisation thus appears to be closely related to the idea of acceptability, as employed in this study, although the latter term may here include elements of ‘formal acceptability’, or acceptability for pedagogical purposes.

If it is assumed, according to the methodological intention of the study, that features were marked according to their salience and that the acceptability ratings reflected their degree of stigmatisation, to a certain extent, then there are two possible explanations for features being non-significant in terms of their effects on acceptability:

1. Salience was relatively high, but stigmatisation was relatively low: CCRF-PC and TH-STOP appear to fall into this category since while there were

numerous reports, they did not significantly affect the acceptability ratings on an overall basis. The fact that the statistical procedure was sometimes able to distinguish between the frequency of error reports and their severity can be seen as one of the strengths of the approach.

2. Salience was relatively low: it seems that some features were simply less noticeable than others, or that they occurred less often in the data. For example, there were only two cases of LV reported in the data (one word token), and some instances of LV in the samples were missed. The low number of cases would tend to result in a lower level of statistical significance; Tagliamonte (2006: 237) notes that '[a] data set with a large number of tokens will tend to detect more factors to be statistically significant than one with fewer tokens'.

On the other hand, TH-STOP is usually seen as having relatively low salience because of its occurrence in less prominent contexts (i.e. mainly in unstressed syllables). It is possible that because students were required to mark up to three features, they marked even the less salient cases of TH-STOP (because it occurs frequently, in a majority of speakers, because there is a considerable degree of phonetic difference between [ð] and [d], and because this is a 'known' feature of HKE, at least for some listeners). Low acceptability, on the other hand, can be seen as a result of features being both salient and stigmatised. Many students marked them, increasing the chance of significance, and the acceptability ratings reflected the negative effects of these sounds. The case for salience and stigmatisation being partially independent dimensions is again strengthened by the status of VOWELSUB errors; there were fewer cases than for TH-STOP, but only VOWELSUB was found to exert significant effects on the acceptability ratings.

In summary, the relative effects of the different components of salience are difficult to separate, although prosodic prominence appears to be important, and salience may not necessarily lead to stigmatisation. It may be the case that while language-internal factors largely explain perceptual salience and determine whether features are noticed or not, social-psychological factors such as stigmatisation may determine

whether this salience will have an effect on listener evaluations. A fuller explanation of why features appear to differ in terms of their acceptability requires a more comprehensive approach.

6.3 A combined approach: a dynamic or evolutionary model

As has been mentioned, the most effective explanations of the differences between phonological features appear to result from the use of an evolutionary or ecological metaphor. Under this approach, the survival of features within speakers and varieties is seen as result of competition and selection within a common ‘feature pool’ (Mufwene 2001: 197). This process applies to all languages and to all instances of language change, although the factors that influence selection and the outcome of these factors, in terms of selected features, will obviously vary. The use of evolutionary metaphors to explain language change has a long history, but since the advent of Saussurean linguistics there has been a tendency to move away from the evolutionary paradigm and towards purely structural or formal explanatory approaches (Nettle 1999: 450).

Under an evolutionary approach, the acceptability of features can be seen as an indication of selection processes in action; the task for both biologists and linguists is thus to explain these processes. The concept of variation is relevant to both disciplines. In the case of a second-language context such as that represented by English use in Hong Kong, there is variation in terms of the frequency and distribution of features (as shown, for a limited number of features, by the implicational scale in this study). The reasons for this variation may include the user’s stage of language development, as suggested by developmental models such as those of Hansen (2006) and Major (2001). This is in turn related to factors such as transfer and markedness, so that the features which persist for longer in individual speakers’ productions tend to be marked features (defined as those which are rare in the world’s languages and tend to occur later in L1 acquisitional processes; see Thomason 2009: 357). Other instances of variation may arise from speaker innovations and the random, idiosyncratic variation that naturally occurs in speech.

Performance factors, such as speech rate or the speaker's position on the hypo- to hyper-articulated continuum, will also lead to variant forms. Nettle (1999: 448) summarises these other sources of variation:

In linguistic performance a constant stream of variations on canonical forms is produced. Some of these are random errors in production or perception. Others are generated in discourse as novel, sometimes deliberate solutions to particular communicative problems. In addition, the speaker will hyper-articulate forms where necessary to make herself understood, but will otherwise produce the most reduced variants she can as her speech output tends towards a maximal economy of production.

The incorporation of functional imperatives such as 'making oneself understood' in explanatory models is often criticised by formalists, who point out that many instances of language change involve a reduction in the complexity of sound systems and thus an increased danger of homonymic clash. Labov (1994: 190) provides as an example the Northern Cities vowel shift, which has led to the merger of certain vowels. However, it should be remembered that Labov is referring to L1 speech communities; in L2 situations an additional source of variation is incomplete learning, or differences in proficiency level. It therefore seems plausible that intelligibility is among the factors that influence language development, at least for some speakers. Trudgill (1986: 21) discusses 'the need to be understood' by observing that in situations where mutual intelligibility is potentially problematic, speakers rapidly become aware that 'some features are likely to cause interlocutors more trouble than others'. Generally, it can be hypothesised that intelligibility exerts developmental effects, especially on those speakers who need or wish to be understood by speakers from other language backgrounds. While in Hong Kong there are limited opportunities for international communication, especially for students, the 'core function' of language still suggests that intelligibility is among the factors that influence language development.

Another, related criticism of functional approaches is that they tend to be *post hoc* and therefore unprincipled (Lass 1980). If an established sound change or variational

feature leads to reduced intelligibility it can be argued that the reduced effort involved was a more powerful factor; if on the other hand a feature maintains intelligibility then it can be argued that intelligibility, or ‘being understood’, was more powerful. Labov (1994: 549) criticises the implied teleology of functional explanations thus: ‘I find myself inherently suspicious of anything that is inherently good’. However, the existence of counterexamples does not invalidate the functional approach, as pointed out by Nettle (1999: 456): ‘many cases of apparently anti-functional drift may be argued to be functional within a broader, competing-motivations framework which includes the pressure towards simplicity as well as the pressure towards communicational efficiency’. In other words, functional considerations may help to define the constraints that operate on sound changes at some level; within these constraints, other selection pressures may also be operating.

In practice, then, an evolutionary perspective may be unpalatable to some because it tends to result in lists of the possible factors that may exert selection pressures on linguistic variants. Many of these factors tend to compete with one another, so the identification of operative factors is problematic. Nevertheless, it is profitable to consider the constellation of such factors. Among the factors listed by Schneider (2007: 111-112) as playing a role in determining why some forms persist while others vanish, the following are most relevant to this context:

- demography: ‘[f]orms used by a majority of speakers have a greater chance of survival than minority uses’;
- frequency: ‘[f]orms which occur frequently in communication...stand a greater chance of survival than rare ones...this criterion is related to demography’;
- markedness: ‘[u]nmarked forms...are more likely to be selected’;
- salience: ‘[f]orms which are perceptually salient...have a greater chance of survival...this principle may conflict with the previous one, as there may be a certain correspondence with salience and marking’;
- the status of speakers: [f]orms used by high-status speakers will tend to be copied and are thus likely to spread and survive, unlike those typical of low-

status speakers...“status” in this sense need not conform to official status in society’;

- the identity-marking functions of linguistic forms: ‘[s]peakers will copy and adopt forms used by those they wish to accommodate to, and so forms used by popular groups will tend to spread’; and
- the similarity or difference between L1 and L2 forms or patterns: ‘similar or identical forms will tend to reinforce each other’.

Schneider does not make any claims for the completeness or mutual exclusivity of this list, which includes internal factors (salience and markedness), external factors (L1/L2 differences, and frequency if it is considered in functional, interactional terms rather than as part of the inherent structure of a language) and extralinguistic factors (the remainder of the list). If frequency is considered to be part of salience, then the latter category may tend to subsume the former. As Schneider points out, demographic factors also affect frequency in that the number of speakers using particular forms may affect their chances of survival. ‘Frequency’ can thus be seen as having external, as well as internal, aspects. A factor that could be added, again from a functionalist perspective, is that of intelligibility: features or innovations that reduce intelligibility are less likely to survive. However, it seems probable that intelligibility is already coded into language systems to some extent; it is noteworthy that some of the marked features of English (the dental fricatives, postvocalic /l/ and certain types of final consonant cluster) do not play significant roles in maintaining intelligibility.

Some factors appear to involve circularity. Saying that forms stand a greater chance of survival because they occur frequently seems to be no different to saying that forms occur frequently because they stand a greater chance of survival. In the above list, salience is seen as increasing the odds that a feature will survive; this is also the view of salience taken in language-contact studies, for example Kerswill and Williams (2002). However, this study has shown that it is the features with lower salience, such as TH stopping and L vocalisation, that tend to persist in the feature pool. This may reflect the different social realities of L1 and L2 speech communities. In L1 communities language change is thought to be propagated by influential

‘nonconformists’ whose linguistic symbols diffuse throughout the community ‘in a larger pattern of upward social mobility’ (the Constructive Nonconformity Principle; Labov 1994: 516). These features therefore tend to be salient, or they would lack the ability to signal identity. In some L2 situations, such as in Hong Kong, upward mobility is more likely to be associated with the linguistic symbols of existing elites, thus creating a tendency towards conservatism and the gradual avoidance of salient linguistic nonconformity.

Of course, using the above list as a guide to the ‘linguistic ecology’ of the situation, the markedness and the contribution to intelligibility of features are also likely to be implicated. Despite the difficulty of achieving robust explanations that avoid circularity, a survey of the other features examined in this study from an evolutionary perspective may help to illustrate how the factors and processes operate, as well as assisting in a synthesis of the data on variation and acceptability from the two parts of this study. There appear to be some grounds for concluding that low acceptability means a feature is unlikely to persist and become entrenched, at either the individual or the speech community level (although this of course may vary greatly between individuals). Whether it is a result of the factors listed above, or whether it constitutes a separate factor, is somewhat uncertain.

6.4 The phonological features from an evolutionary perspective

Starting with the low-acceptability features, this section will consider the phonological features and the findings of the study using the ‘evolutionary’ factors listed above. In terms of the factors, the order of consideration will be internal, external and extralinguistic. Where appropriate, evidence from other varieties of English, and from language change, will also be used to evaluate the degree of universality of the selection processes involved.

6.4.1 SYLL

As was mentioned in Chapter 5, SYLL errors include various types and permutations of vowel and consonant modifications. In this study they have been linked with rapid speech, and therefore can be seen as sound variations associated with the hypo-articulated part of the speech continuum. While developmental processes such as transfer may explain the prevalence of features such as /v/ substitution in early-stage learners, SYLL errors are performance-related and can potentially occur within any speaker (in fact, they may even be more common in high-proficiency speakers). SYLL errors appeared to be salient in the data, and this would be expected given the nature of the modifications involved. They may also affect intelligibility due to the loss of information involved. Perhaps more importantly, the idiosyncratic nature of these errors makes it extremely unlikely that any would be retained in the feature pool, as no single type will occur frequently enough to become ‘cognitively entrenched’ (Schneider 2007: 110). From the point of view of language change, it might be the case that vowel reduction in English was at one time a SYLL ‘innovation’ that did become part of the feature pool; an additional evolutionary factor that may have contributed to its adoption is the possibility of maintaining intelligibility, even when vowel reduction is employed widely and systematically in function words.

6.4.2 CCRF-PV and CCRF-PC

CCRF-PV (prevocalic final CCR) was found to have significant effects on acceptability in this study, while CCRF-PC (preconsonantal final CCR) was not. While final CCR occurs in all varieties of English, reduction in prevocalic or prepausal contexts seems to be more prominent in non-native varieties (Schreier 2009: 68). The general prevalence of final CCR in HKE phonology can be explained by the lack of consonant clusters in the L1, as well as by the pressures towards simplification experienced by all speakers. As was mentioned in Chapter 5, the salience of prevocalic CCR (CCRF-PV) appears to be greater, and this offers an initial explanation for the significant effects on acceptability it was found to have in this study. The effects of morphemic status have not been separated out in the data,

but it is worth noting that intelligibility considerations and information theory might predict a lower incidence of final CCR in bimorphemic clusters, as they carry more information. As has been mentioned, Jenkins' (2000) findings on the intelligibility of final cluster modifications may be a result of non-native speakers achieving different optimisations of the competing factors involved, and in some cases prioritising simplicity over intelligibility.

Preconsonantal final CCR (CCRF-PC) was a feature that missed the significance threshold used in this study, but its significance at the $p < 0.01$ level suggests it is worthy of further consideration and study. The lower salience of preconsonantal CCR helps to explain its reduced effects on acceptability, compared to prevocalic CCR.

6.4.3 PHONSUB-V

PHONSUB-V or /v/-substitution appeared fairly infrequently in the HKE data. There appear to be several 'selection pressures' that help to explain this. Firstly there is the salience of the /v/ phoneme, which is in turn mainly a result of its relatively high frequency of occurrence and its appearance in prosodically prominent positions. This means that substitutions will also tend to be noticeable, depending on the degree of phonetic similarity or dissimilarity involved. Secondly, inter-variety comparisons suggest that /v/ substitution occurs in relatively few varieties of English (although it occurs widely in Indian English). It may therefore be related to transfer, rather than developmental processes, and this would normally tend to make a feature infrequent in both inter-variety and intra-variety terms. As Hansen (2006) and Major (2001) have shown, transfer features tend to disappear in the speech of evolving L2 users. The likely causes of this are contained in the list of evolutionary factors. The /v/ phoneme does not seem to be particularly marked, in comparison with other features such as the dental fricatives, although if L1/L2 comparisons are made it would be expected to create difficulty for Cantonese L1 speakers; Hung (2000) makes a case for this and other voiced fricatives not forming part of the HKE system. In general terms, it appears that from a developmental perspective the relative difficulty of acquiring the /v/ sound is outweighed by the other factors, such as its salience (for

example, its occurrence in prosodically prominent positions), plus the absence of markedness constraints on its acquisition. There are also its possible consequences for intelligibility, and the combined effect of these factors is to select out transfer-related features.

Turning to the extralinguistic factors, the implicational scale shows that /v/ substitution was only used by a minority of speakers. Of course, this does not explain its low acceptability, which may be seen as either a cause or effect of ‘external frequency’ or demography. It can, however, be hypothesised that speakers who have themselves acquired the /v/ phoneme may be consciously or unconsciously aware of the L1-related origins of substitution and its tendency to occur in the speech of speakers with a lower proficiency level, and may attach stigmatisation to the [w] substitution. This may reflect a general tendency; for speakers who do not possess a phonemic distinction, substitutions or confluations may not be noticeable, but those who have acquired it tend to stigmatise its absence in the speech of others. If this is the case, low acceptability becomes another causative factor that may hasten the acquisition of /v/ by individual speakers and further reduce the chances of survival of the [w] variant, at the speech community level. This raises the question of whether stigmatisation needs to be added to the list of evolutionary factors, in some form. The approach taken in this study is that stigmatisation is at least partly due to the combined operation of some of the existing factors, such as salience, demography and intelligibility effects. There does not appear to be a case for adding it as a separate factor.

Finally, other extralinguistic factors that militate against the entrenchment of [w] substitutions follow on from the above discussion. If /v/ substitution is stigmatised as a low-proficiency variant, then the positive selection pressures for /v/ are enhanced by its presence in both local, high-proficiency accents and in international varieties of English. Similarly, the rarity of the [w] substitution in other varieties of English means that there are few external pressures for its adoption. Childs and Wolfram (2004: 442) note that the alternation of /w/ and /v/ is found ‘in scattered varieties of English throughout the world’, and mention a possible Anglo source in earlier varieties of British English such as Cockney. All in all, the evolutionary model offers

a satisfactory explanation of the distribution of this feature and its acceptability characteristics.

6.4.4 VOWELSUB

In this study, VOWELSUB substitutions such as those occurring in the words *want*, *maintain* and *leadership* were found to exert significant effects on the acceptability scores. As with SYLL errors these substitutions were often idiosyncratic, although Speaker 4's use of a short vowel in the first syllable of *maintain* may be seen as representing a general tendency towards vowel shortening in HKE phonology. Vowel substitutions are certainly likely to be salient, as vowels form the sonority peaks of syllables. Many of them will have consequences for intelligibility, although from an evolutionary standpoint vowel mergers such as those in the Northern Cities vowel shift (Labov 1994) can be interpreted as the victories of simplification over the avoidance of homophony. Nettle (1999: 456) believes that:

If two segments are phonetically so similar as to cause discrimination problems, then their merger could produce a system which is more optimal than their maintenance, since the simplification of signal processing could more than offset the problems of homonymy thus created.

Thus from an evolutionary perspective, phonetic similarity could be invoked to explain a sound merger, while the avoidance of homophony could be invoked to explain the absence of such a merger. The possible circularity of functionalist arguments has already been mentioned, but Nettle (1999: 457, citing Asma 1996) draws on biological parallels to show that adaptation is often imperfect and structurally limited. Perhaps the best way to assess which types of vowel substitution are more likely to survive as emergent or even dominant features is to survey the vowel systems of varieties of English in general; similarities may indicate the parts of the vowel system where selection pressures operate most powerfully. Mesthrie and Bhatt (2008: 120) observe that in 'New Englishes', five-vowel monophthong systems are common and are likely to merge either the TRAP/STRUT (/æ, ʌ/) contrast, or the LOT/STRUT (/ɒ, ʌ/) contrasts. There are no obvious functional

explanations for this, and such mergers are probably the result of the various physiological, acoustic and psycholinguistic constraints operating on speech production and processing. The statistical model of vowel systems designed by Lindblom (1986) works on the assumption that ‘vowel systems evolve in such a way as to give maximum intelligibility for a given type of articulation’ (cited in Nettle 1999: 452). Lindblom’s model can successfully predict the optimal combination of vowels for vowel systems of a given size, suggesting that the constraints and evolutionary pressures are to some extent universal and species-specific to humans, rather than being related to particular languages.

Hung (2000) and Deterding *et al.* (2008) found that the mergers of TRAP/DRESS (/æ, e/) and KIT/FLEECE (/ɪ, i:/) are typical of some HKE speakers’ vowel systems. The first merger did not seem to be present in the samples used in the study, but Speaker 6’s production of *leadership* was probably marked by students because of the shortening of the first vowel, an instance of KIT/FLEECE merger. While the L1/L2 similarity (or dissimilarity) factor could be invoked to explain this HKE feature, it seems that this merger is fairly common across varieties of English. In a global synopsis of phonetic and phonological variation worldwide, Schneider (2004: 1128) concludes that the importance of length in distinguishing phonemes is ‘definitely waning’ in general and is ‘considerably reduced’ even in American English, as well as in African and Asian varieties. KIT and FLEECE therefore tend to merge in ‘quite a number of varieties’, along with certain other length contrasts such as FOOT/GOOSE and LOT/THOUGHT. The merger of TRAP and DRESS also occurs in several varieties, although this may depend on the phonological context (for example, it occurs before /l/ in South Africa, Australia and New Zealand (Schneider 2004: 1122).

On the other hand, the ‘globally predominant’ realisations of KIT and FLEECE are [ɪ] and [i:], respectively (Schneider 2004: 1128), and their merger is classed by Schneider as a ‘distinctive sound realisation that may serve to characterise specific regions’ (Schneider 2004: 1128). In this study, the low acceptability of vowel substitutions suggests that for some listeners, certain distinctive realisations may not be positively evaluated, although there is no indication of the possible differences

between vowel substitutions. Several aspects of salience may explain the effects on acceptability, including frequency and participation in phonological contrast (the KIT/FLEECE contrast is involved in a large number of minimal pairs; see Brown 1991: 82). In general, the complexity and interrelatedness of vowel systems makes it difficult to assess whether changes may be taking place, but a consideration of the vowel mergers above suggests that the combined effect of the evolutionary factors is mixed, with some general tendencies towards merger in certain areas of the English vowel system. More data would be needed to accurately assess the situation in Hong Kong.

6.4.5 OTHER C-SUB

The OTHER C-SUB category represents a somewhat mixed bag of features and it is difficult to draw conclusions about the general factors affecting it. Two of the most frequently identified tokens involve devoicing, and the devoicing of voiced consonants is certainly a characteristic feature of HKE, even with many higher-range speakers. Again, for those speakers who have more reliably acquired the voiced/voiceless distinction, there may be stigmatisation of its presence in the speech of others. The fact that it can occur in a wide range of contexts explains why it occurs frequently as a category, although the chances of it becoming cognitively entrenched in terms of the pronunciation of particular words seems quite low; this feature often has considerable salience. Devoicing may also lead to homophony, in some cases (for example, with *back* and *bag*), and there may be intelligibility effects that tend to select it out of phonological systems.

6.4.6 Non-significant features

As has been mentioned, the non-significance of PHONSUB-R is somewhat surprising, given its apparent salience (in terms of frequency and potential to affect intelligibility; see section this may be due to the relatively small number of reported cases (17, against 30 for PHONSUB-V). The relative salience of this feature, in comparison with PHONSUB-V, is somewhat uncertain (see section 6.2.2 above).

Turning to the other features, the phrase that best explains the overall patterns of variation and acceptability characteristics is ‘the emergence of the unmarked’ (McCarthy and Prince 2004). Neither of the dental fricative substitutions (TH-FRONT and TH-STOP) significantly affected acceptability, although the low incidence of the former must be considered. The English dental fricatives are marked features in that they are quite rare in the world’s languages and tend to occur later in L1 acquisitional processes (Thomason 2009: 357). In terms of salience, the voiced dental fricative tends to occur in unstressed and therefore less prosodically prominent positions, reducing its salience (and thus the salience of substitutions). Dental fricative substitution, then, represents the emergence of the unmarked and the ‘natural’, an evolutionary tendency that appears to be unchecked by competing linguistic or extralinguistic factors. This can be seen as operating at an individual level, as speakers appear to experience few pressures to incorporate the marked variants into their phonological repertoires, but it also has an impact at the speech community level. The implicational scale shows that a majority of speakers used TH-STOP at least once, although it is uncertain how far this represents the actual speech community. These patterns of occurrence may in turn have frequency and demographic effects.

Although dental fricative substitutions are widespread, there is a general perception that in native varieties they tend to be stigmatised (e.g. in British and American varieties; Schneider 2004: 1123). It is probably this stigmatisation that prevents a natural tendency from becoming a permanent change. Blevins (2004: 30) provides the example of the /w, ʍ/ contrast, extinct in many varieties but preserved in others ‘through unnatural means: enforced practice and repetition, combined with spelling pronunciation’. Although both are marked sounds, the dental fricatives may differ from /ʍ/ in terms of frequency and functional load. Even so, it may be that extralinguistic factors are important in determining how far tendencies towards the ‘unmarked’ and the ‘natural’ will actually proceed. There may be differences between native and non-native varieties in this regard. Kortmann (2004: 1) believes that in grammatical domains, ‘regional and social non-standard varieties conform to cross-linguistic tendencies where the relevant standard varieties do not’ (cited in Filppula *et al.*, 2009: 3). However, as well as there being differences in the ecology

of extralinguistic factors, non-native varieties may also be affected by the fact that their listeners will not always perceive differences in the same way. An instance of TH stopping may be highly salient to a native speaker or high-proficiency listener, but less so to a low- or middle-proficiency listener, and this will exert its own effects on language development at the individual and community level.

The same tendencies can be observed in the case of L vocalisation, one of the least significant features in the study (although this was probably also due to the low number of cases). This seems to be a prototypical case of the emergence of the unmarked and the natural, as noted in the survey of its occurrence in British English dialects by Johnson and Britain (2007). As in the case of dental fricative substitutions, there is abundant evidence from cross-linguistic comparisons and from language change to support the seeming inevitability of this phenomenon. However, in the case of inter-variety comparisons, the evidence is less coherent. Schneider (2004: 1125) concludes that L vocalisation occurs ‘fairly generally’ in some American English varieties and ‘variably’ in Australian and New Zealand English, being less common and ‘sociolinguistically conditioned’ in British English. It is common in African varieties of English, according to Mesthrie and Bhatt (2008: 128). Among the possible reasons for this relative rarity, Wells (1982: 259) observes that L vocalisation may have consequences for the rest of the phonological system; vowel realisations may be affected, and some may result in extra vowel phonemes being created. This is in turn a consequence of a proposed general principle of language change, namely that ‘a change that simplifies the system in one place may well complicate it in another’ (Thomason 2009: 358). How the possible problems arising from this are dealt with by language users, and how these functional considerations may affect language development, are questions for further investigation.

In contrast to the low number of cases in the main study, the implicational scale shows L vocalisation to be the most frequently occurring feature among those considered; all but four of the 25 speakers in the mini-corpus used it at least once. The apparent difference between occurrence in the preliminary study and error coding in the main study is probably a consequence of the shorter samples in the

main study providing fewer contexts, as well as the lower detection rate (for example, the instance of L vocalisation in the word *children* in Speaker 8's sample passed unnoticed). The feature also occurs variably, rather than categorically, in HKE, and is probably also conditioned by phonological context. In my own experience of teaching English phonetics to students from Hong Kong, I have been struck by how frequently postvocalic /l/ is transcribed as a vowel. Words such as *able* and *little* are often transcribed as /eɪbəʊ/ and /lɪtʊ/, and these representations of 'pronunciation spellings' often occur in students' spoken English (the transcriptions may be affected by the presence of orthographic 'e', however). On the whole, this feature also appears to be subject to competing influences, with markedness, naturalness, intelligibility and L1/L2 differences increasing the likelihood of L vocalisation or deletion, and other factors preserving the consonantal realisation. These may include the influence of spelling, but it is interesting to speculate whether extralinguistic factors also play a role. While neither L vocalisation nor TH stopping generally occur in the kind of 'prestige' British or American accents Hong Kong students are often exposed to, TH stopping may have gained covert prestige through the influence of youth culture (an animated movie entitled *Round Da Way* was released in 2009). In this case, its sociolinguistic salience is probably greater than that of L vocalisation.

Among the remaining non-significant features, the case of FULL VOWEL (the non-reduction of full vowels in unstressed syllables) is interesting as it contrasts with the VOWELSUB category in terms of its effects on acceptability. This could be predicted by invoking salience, as arguably the degree of phonetic difference between schwa [ə] and a replacement full vowel such as [ou] (in *production*) or [ɒ] (in *confused*) is less than in the case of vowel substitutions. FULL VOWEL also occurs in unstressed syllables and therefore in theoretically less prominent positions, although whether or not the replacement affects perceptions of word stress is another matter. It is also unlikely to affect intelligibility, as discussed in Chapter 2. Word class pairs such as *contract* in its noun and verb forms may be an exception, although the fact of word class difference would tend to make confusability less likely.

6.5 General summary

Within the evolutionary model, linguistic factors appear to explain many of the acceptability findings, and salience (even if limited to degree of phonetic difference and prosodic prominence) is particularly useful. Essentially, the salient features of an assumed ‘standard’ phonology seem to be more likely to appear in HKE phonology, especially as many of the HKE substitutions (e.g. /v/ substitution, and vowel substitutions) are phonetically quite different and therefore have salience of their own, assuming listeners are able to hear them. These features may also affect intelligibility, although the mechanism by which this affects phonological development is somewhat unclear. If a feature reduces information-carrying capacity it may come under negative selection pressure, but the evidence from language change is replete with examples that have involved simplification and a reduction of information (e.g. Labov 1994; Sampson 1980). Functional factors such as intelligibility are likely to form parameters in that they restrict the range and type of possible innovations; within these parameters, a great deal of variation is possible and selection pressures towards simplification may be stronger than those towards information or system preservation.

The non-standard features that are likely to persist in individual HKE phonological systems, thus using their ‘hosts’ to replicate and become speech community norms of some kind, are essentially those that have low salience and are difficult to hear or ‘notice’ in a psychoacoustic sense. Having few consequences for intelligibility may also help a feature to persist, but this could be seen as an aspect of salience. Languages appear to evolve by preferring the use of salient features to make important contrasts, as suggested by Nettle (1999: 448): ‘[p]honological distinctions which are hard to hear or unreliably signalled are unlikely to be represented in a nascent grammatical system’. While this raises the question of possible cross-linguistic differences in terms of salience (the extent to which distinctions are ‘hard to hear’ depends on language background and experience with these sounds), the features considered here appear to be non-salient in universal terms. In addition, the more systematic a feature is, the greater will be its frequency of occurrence and its demographic impact. The prevalence of TH stopping and L vocalisation in the

speaker data increases the likelihood that these may become (or may already be) entrenched features of HKE. Substitutions of the voiceless dental fricative (i.e. TH fronting) appear to be less likely to persist, on the other hand; a salience-related explanation is the greater tendency for this sound to occur in stressed (and thus prosodically prominent) positions. The non-reduction of full vowels in unstressed syllables is also highly likely to occur and may relate to a tendency towards syllable timing in Asian, and World, Englishes (Crystal 1996). However, it appears to be less systematic in terms of phonological context, as well as in terms of intra-speaker variation, and for some speakers may be related to movement along the hypo- to hyper-articulated continuum.

This conclusion – that TH stopping and L vocalisation are the HKE features which appear to be best adapted to their environment and hence most likely to persist – is generally supported by evidence from other varieties of English, and in the case of L vocalisation by examples of historical change in other languages. In the case of TH stopping, the key factors appear to be the low prosodic prominence of most contexts for occurrence, in unstressed syllables), and its limited effects on intelligibility (hence, there is low salience). Furthermore, TH stopping involves the replacement of a marked sound with a less marked sound. This tends to increase its demographic occurrence and frequency, and makes it more likely to appear in the input received by new users (for example, in schools). This in turn creates frequency effects that further increase its chances of replication and persistence. In the case of L vocalisation, the key factors appear to be the phonetic similarity of the sounds concerned, the low prosodic prominence (in coda position) and again, the low contribution to intelligibility and marked nature of the original sound.

These are of course the same selection pressures that have led to these features being widespread in other varieties of English, and this raises the wider question of whether L2 varieties of English are as ‘new’ as some observers have claimed. An initial answer might be that if ‘autonomy and creativity’ (Bolton 2000) exist in HKE, they are more likely to be found in other linguistic realms than the phonological. It appears to be the case that the evolution of postcolonial Englishes can be regarded as a subset of more general processes of language change (Schneider 2007: 97).

Features that are highly distinctive are often related to transfer and tend to disappear as users' phonological development progresses (Major 2001, Hansen 2006). On the other hand, while it is obviously true that there is a great deal of distinctiveness in HKE phonology, this appears mainly at the subphonemic level, for example in vowel realisations, or at the suprasegmental, prosodic level. There is also evidence for the claim that L2 varieties such as HKE represent more 'natural' environments for language change, where the pressures of stigmatisation and conformity are perhaps weaker and the universal tendency towards 'the emergence of the unmarked' finds fuller expression. The priorities of speakers of these varieties may be very different to those of speakers of L1 varieties, for a variety of reasons. These adaptations could be seen as an example of the general creativity of the human language learner, who saves processing capacity for other tasks by not learning (or not using) what is not needed.

The evolutionary model used in this chapter has proved to be effective in explaining the patterns of variation and their consequences for acceptability, although the main criticism of such an approach is that it resembles a 'theory of everything' in which nothing will be found (Chomsky 2000: 70), except for *post hoc* explanations. But as Schneider (2007: 101) notes, although model-building obscures the complexity of processes it is still 'a useful conceptual and heuristic exercise'. While formal approaches may provide useful data about how language-related cognitive mechanisms operate, there still remains the question of how or why such mechanisms should be the way they are (Nettle 1999: 460). The attempt to move from reduction to synthesis, although fraught with difficulties, is still worthwhile.

To conclude this chapter and to help illustrate some of the processes involved, a diagrammatic representation of the process of L2 phonology development is given in Figure 6.1 below. In Chapter 2, a four-quadrant diagram was used to show the important considerations when evaluating features of pronunciation models. With a few modifications to take into account some of the factors discussed in this chapter, the same diagram can also be used to show how linguistic and non-linguistic, internal and external factors may operate in a dynamic fashion at the individual level.

Figure 6.1. Factors affecting individual feature development and use in L2 phonology acquisition.

Of course, if non-linguistic, socially-conditioned factors such as acceptability are included this ceases to be a purely individual model, but it is still a useful depiction of the processes involved. These proceed as follows: initially, feature use is affected by developmental constraints and may include the transfer or influence of L1 phonological features (quadrant 1). As users gain more experience in using the language, linguistic factors such as intelligibility and the composite notion of ‘salience’ influence the way feature use develops (quadrant 2); features related to transfer tend to decrease, perhaps partly as a result of these factors. Non-standard features with lower salience are more likely to be retained than those with higher salience, although, as in many language contact situations, extralinguistic factors such as identity may allow for the retention of more salient features (quadrant 3). Feature use is further mediated by the user’s ongoing evaluation of attitudes and acceptability, for example by accommodating to the speech of groups with which he or she wishes to identify (quadrant 4). Fluent speakers may use features that arise

from prioritising speed over clarity, although these are mainly idiosyncratic and do not form part of the ‘feature pool’. While transfer features tend to disappear, features that have marked and unmarked variants will tend to show the ‘emergence of the unmarked’, or from another point of view the victory of simplification over the avoidance of homophony. The diagram should not be taken as implying universal or neatly sequential processes; there may be considerable inter- and intra-speaker variation within the general pattern. For example, some language users may be capable of approximating or reproducing L1 norms, but may choose not to or may vary their speaking styles according to the situation. The diagram also assumes that speakers have access to NS or ‘standard’ patterns of language use, and as Bolton and Kwok (1990) observe in Hong Kong, this may not always be the case.

Taking an evolutionary perspective, if language features are conceptualised as sequences of genetic code then the biological processes of replication, variation and selection can all be seen to exist in linguistic evolution. In L2 communities, replication occurs when languages are passed on, whether by teachers, parents, peers or outside sources. In second language learning environments replication is more likely to be ‘imperfect’ in the sense of involving less input and being influenced by L1 features and processes. Variation also arises from performance factors. Selection pressures operate via ‘differential adoption’ (Nettle 1999: 449) into phonological systems, under the influence of the various factors discussed in this chapter. These processes determine the ‘reproductive success’ or the overall frequency with which features occur within a speech community, and this in turn influences the input that new ‘hosts’ receive.

While it might be tempting to represent this feedback loop in the form of a diagram, a list such as that given by Schneider is probably equally effective, given the complexity of the possible interactions. The contents of the ‘feature pool’ will vary greatly according to the stage of learning, the medium of education, opportunities for exposure to the L2, and so on. Nevertheless, if the distribution of features shown in the implicational scale is taken as an example, the process can be summarised as follows: selection pressures lead to certain salient, HKE features becoming minority forms, while certain non-salient, unmarked features persist as majority forms (in the

sense of being present in the repertoire of a majority of speakers). These features go on to influence the input that is transmitted to other users in terms of frequency (the rate of occurrence in the input) and demography (the number of individuals who use the features). The features encode social as well as linguistic information.

CHAPTER 7

CONCLUSIONS

7.1 Introduction

After the previous chapter's discussion of the various explanatory factors, this chapter will take a more practical orientation and conclude the study by returning to the research questions. While these are focused on Hong Kong, the possible implications of the study for the teaching of pronunciation in general, and for research in the areas of World Englishes and ELF, will also be considered. Finally, the chapter summarises the limitations and achievements of the study and makes suggestions for further research.

7.2 Research questions

7.2.1 Research question 1

Research question 1 asked whether local students are likely to accept a local model for pronunciation teaching purposes. The results of the survey suggest that they are, as long as the models do not include certain phonological features and are therefore representative of a certain part of the range or continuum of Hong Kong English. This is a significant finding, as several previous studies in Hong Kong (e.g. Forde 1995; Luk 1998; Candler 2001) have indicated that local students are unlikely to accept a local model. While the study of Bolton and Kwok (1990) showed that Hong Kong students were more likely to accept a 'mild' Hong Kong accent than a 'broad' one, the present study has provided far more detail about accent variation and acceptability. The acceptability scores are relative, which makes precise conclusions about 'acceptability' difficult in absolute terms. However, it is interesting that the native speaker, Speaker 11, did not achieve the highest acceptability rating. While

this may have implications for initiatives to recruit ‘native speaker’ teachers in Hong Kong, such as the NET scheme, another interpretation of the data is that teachers need to have native-like accents to maximise acceptability. The important point is that this is fully compatible with having a Hong Kong accent.

Although there was a weak negative correlation between acceptability and perceived accent strength, as measured by the correlation between questionnaire items A and E ($r = -0.464$), it is by no means true that accent strength predicted the acceptability scores. Speaker 8 provides an illustration of this; she was ranked third in terms of both direct and overall acceptability, but came sixth in the ranking of accent strength from weakest to strongest. Student comments also indicated that they were fully aware of her Hong Kong origins (for example, ‘clear pronunciation, but doesn’t sound like a NS’ and ‘a very typical HKer’). Despite this, her overall acceptability score was only slightly lower than the British native speaker’s (4.45, against 4.48). This provides further evidence to support the view that it is not perceived accent strength that affects acceptability, but rather the presence of certain phonological features and the connotations these may have for the listeners. There may of course be some relationship between segmental deviations and accent strength, but research evidence suggests they are at least partially independent (Munro and Derwing 1999). Nevertheless, the present study does not claim that segmental features alone can predict acceptability scores. At several stages it was noted that suprasegmental features were also likely to play a role, although these were not measured by the study. The possibility of co-occurrence also exists, meaning that the segmental and suprasegmental features may exert combined effects on acceptability.

The study’s findings should not be interpreted as suggesting that ‘accent does not matter’. The overall level of phonological accuracy was found to be the most important determinant of acceptability, among the factors tested. The type of error was also important, with some features being noted by a large number of students but not significantly influencing the acceptability scores (for example, TH stopping and certain instances of final CCR). In general, the findings suggest that the more salient the non-standard accent features, the more they tend to reduce acceptability scores. Less salient features, such as TH stopping and L vocalisation, were either noted but

not significantly penalised (in the former case) or not noticed at all (on more than one occasion, in the latter case).

The samples in this study aimed to include ‘high-proficiency’ or ‘higher-range’ speakers, and the student listeners, as English majors, are likely to have better listening skills and therefore to be more critical than, for example, many secondary school students. The answer to this question may therefore depend on the age and level of the students; secondary school students may not have such exacting standards. However, an important finding of this study is that the intelligibility and acceptability characteristics of features and speakers tended to correspond; one way of interpreting the data is to say that the most acceptable speakers were also likely to be highly intelligible, if the external criteria for intelligibility are applied. Equally, lower levels of acceptability will tend to be associated with lower levels of intelligibility, and even if a group of listeners with a lower proficiency level were to rate accent samples as ‘acceptable’, the level of intelligibility may not be sufficient for international communication. There is some evidence in this study that listeners with lower self-ratings of pronunciation skills tended to rate the samples less consistently, suggesting that their auditory discrimination ability or their amount of exposure to some sounds, or both, were also lower. As the survey respondents were mainly female, further research would be necessary to assess the effects of gender.

7.2.2 Research question 2

Research question 2 asked about the effects of phonological features on the acceptability scores. Despite its limitations, the second regression analysis was able to show that certain features reduced acceptability ratings more than others. The five features that had significant effects were syllabic modification, prevocalic final CCR, /v/-substitution, vowel substitution, and other consonantal substitutions. The most reliable predictors of low acceptability appeared to be syllabic modification, /v/-substitution and vowel substitution; in the lower half of the acceptability rankings, all of the speakers had one or more of these features. The features that did not significantly reduce acceptability included several well-known Hong Kong English accent features, such as TH stopping, preconsonantal final CCR and the non-

reduction of full vowels in unstressed syllables (the first two features also occur in many or all varieties of English).

A limitation of the study was that the use of authentic data led to frequency effects. These meant that there were relatively few possible contexts for some features, such as TH fronting. The use of high-proficiency speakers also meant that certain features, such as the conflation or alternation of [n] and [l], did not appear in the data. Another type of frequency effect related to noticeability. Although there were several possible contexts for L vocalisation, not all of the actual cases were noticed by the students. These frequency effects make it difficult to draw conclusions about these features, although the study takes the view that statistical significance relates to what might be called ‘evaluative significance’; features that occur infrequently in normal speech, or are inherently difficult to hear, are less likely to affect judgments of acceptability.

There was generally a close correspondence between the acceptability and intelligibility characteristics of features (the latter being judged according to Jenkins’ *Lingua Franca Core*). Under this ‘intelligibility-acceptability correspondence’, the presence of deviant core features (such as /v/-substitution) significantly reduced acceptability, while substitutions of non-core features (such as the dental fricatives) did not have significant effects. Some of these features are considered in more detail in the next section, relating to research question 3.

7.2.3 Research question 3

Research question 3 concerned the implications for pronunciation teaching, in particular whether any aspects of ‘standard’ models could be omitted from teaching syllabi or testing materials. It will be recalled that ‘simplifying the task’ of pronunciation teaching is one of the potential benefits of research based on NNS, rather than NS, priorities (Jenkins 2007: 27). Features will be discussed with reference to the four-quadrant model introduced in Chapter 2 (reproduced below), but following an integrative approach that attempts to review all the available sources of data.

Fig. 7.1. The four-quadrant evaluation model (from Chapter 2, section 2.4).

Before embarking on a feature-by-feature analysis, some general observations can be made. The model was designed to schematise the possible areas that must be considered when evaluating pronunciation models, whether in terms of the features or the ‘code norms’ (Bamgbose 1998) that relate to particular varieties. It allows for the more detailed evaluation of particular features, and in the following sections some of the alternatives for pronunciation teaching models in Hong Kong will be examined. Given the importance of international communication in Hong Kong, any such evaluation must take account of the intelligibility characteristics of features. However, the study has found that the intelligibility and acceptability characteristics of features tend to correspond. The features that are more acceptable will be those that do not affect intelligibility, and those that are less acceptable will be those that do affect intelligibility; it therefore does not appear to matter which criterion is used. In Chapter 6 it was proposed that salience is a useful concept that goes some way towards explaining the intelligibility-acceptability correspondence. As a result of

evolutionary processes, languages (including English) do not tend to use inherently non-salient forms in order to encode important information. Because of frequency and priming effects, the language learning process also seems to involve an automatic focus on what is most important, namely these salient forms.

In any case, and to briefly restate the overall conclusions of features-based, empirical research into international intelligibility, the most likely candidates for acceptance within teaching syllabi and language testing are dental fricative substitutions (TH stopping and TH fronting) and postvocalic /l/ substitutions such as L vocalisation. The reduction of final consonant clusters is also acceptable, but only if this follows L1 rules of elision (Jenkins 2000). Jenkins (2000, 2007) concludes that variations in vowel quality are unproblematic, but this study has argued that while the precise realisations of vowels can and will vary, the maintenance of most contrasts is important. Thus the merger of /æ, e/ may potentially cause intelligibility problems, and will not be considered here. An additional feature listed by Jenkins (2000) as being ‘non-core’, that of the non-reduction of full vowels in the unstressed syllables of multisyllabic words, will be considered in the light of the study’s research findings. The following sections will look at these features in more detail. These sections frame the issue of ‘the removal of features’ in terms of ‘the acceptance of variants’, but will also consider whether this involves the retention or the removal of standard forms from teaching syllabi.

7.3 Possible candidates for acceptance

7.3.1 TH stopping

TH stopping was the second most widely attested HKE feature in the mini-corpus used in the preliminary study, with 76% of the speakers using it at least once in their utterances. In the main study it was marked less frequently than more salient features, but there were sufficient cases (30) to make a strong case for its non-significance in terms of its effects on acceptability. The reasons for the frequent occurrence and inconsequential nature of TH stopping have already been discussed;

taken together, are these signs that the voiced dental fricative /ð/ need not be taught? This is hardly a new question; Brown (1974) suggested that considerations of time and difficulty could reasonably lead to the omission of both dental fricatives from teaching programmes.

Quadrant 3 of the evaluation model, relating to identity and personal goals, can provide evidence for either position. It can be argued that ‘identity’ includes ‘global’ as well as ‘local’ aspects, and thus there are arguments for retaining some coverage of the sounds that are used internationally, in both native and non-native varieties. While stigmatisation is dismissed by some ELF researchers as principally an NS phenomenon, it is difficult to predict the pathways that learners will take, and thus the attitudes they will encounter. Brown (1991: 77) describes dental fricative substitutions as evoking perceptions such as lack of status, education or proficiency in English, although this may vary greatly according to the usual variables such as geographical area and the age and socioeconomic status of the listeners. One conclusion is that if a learner’s personal goals include living or working outside Hong Kong, they might be well advised to acquire, or develop some awareness of this sound and its typical substitution.

It should also be pointed out that amongst both the speakers in the mini-corpus and in the main study, TH stopping was not categorical. In other words, while most speakers used it some of the time, very few speakers used it all of the time, suggesting it is a feature that varies according to factors such as phonological context and speech style. This is a powerful argument against intervening in the teaching and learning process. In fact, proposals for omitting features from teaching syllabi on the grounds that they are inconsequential for intelligibility appear to contain some problems of circularity. The patterns of language use upon which general judgments of ‘intelligibility’ are made are a result of several factors, including exposure to various types of input during the learning process. If these patterns are altered because of changes to teaching syllabi, the intelligibility effects of feature use or non-use might also change. Furthermore, the choice of whether to learn and use these features would seem to be an individual one that is determined ‘by the social needs and benefits that are associated with the options’ (Leitner 1992: 227).

One solution might be to incorporate an awareness of issues such as acceptability and speaking styles into courses covering speaking skills. This reflects the reality of native speaker verbal interactions, in which speakers make use of a repertoire of linguistic features which they apply variably (Valdman 1989: 262). There is also evidence that non-native speakers make use of style shifting, for example in Singapore (Pakir 1991; Ho and Platt 1993). Of course, if time is short then the reality is that it is not worth prioritising these sounds; it would be better to acquire other contrasts that do affect intelligibility and the ability to understand spoken discourse. This recalls Trudgill's (2005: 226) proposal for 'the maximum number of readily attainable contrasts'. If there is more time, and as proficiency levels increase, syllabi should cover international intelligibility and sociolinguistic issues. Such stratified or 'dynamic' principles of syllabus design are proposed by Valdman (1989: 276). At lower levels, learners should be exposed to models that offer 'the most regular target language patterns...those variants will be selected that are most easily processible by second-language learners'. This might appear to be an argument in favour of teaching dental fricative substitutions, but Valdman (1989: 276) goes on to say that the pedagogical norm must increasingly take into account sociolinguistic considerations as instruction progresses. Using an analogy that may well be appropriate in Hong Kong, Valdman (1989: 268) also states that learning a foreign language can be viewed as 'an economic investment whose value would be depreciated if the variety mastered contained stigmatized features'.

However, a conclusion of this study is that TH stopping was not stigmatised to a significant degree by the listeners involved. It could therefore be argued that it should not be penalised if it occurs in oral production during speaking tests, and in this case, native speaker perceptions of stigmatisation are probably irrelevant. In fact, descriptors for speaking tests often combine consonantal substitutions under the same heading. A draft set of descriptors for a speaking test in which I have participated includes the following:

English pronunciation is impaired by Cantonese L1 mispronunciations of English...[t]ypically these will be consonant sounds such as /th/ sounds, /l/ and /n/ and the /v/ and /w/ sounds.

As the implicational scale and the associated distributional statistics suggest, it is probably inappropriate to include dental fricative substitutions, especially TH stopping, in the same category as /v/ substitution. If the descriptor is adhered to, most of the speakers in the mini-corpus could receive a ‘borderline’ score for pronunciation on the grounds that they use TH stopping; in practice, however, there are other criteria that can be used to corroborate raters’ decisions. Another consideration is that it is not merely the presence of TH stopping that matters, but how often it is used. Although its use by several speakers did not significantly reduce acceptability in the present study, this is no guarantee that different patterns of TH stopping (for example, categorical substitution in the speech of a low-proficiency user) might not have different effects.

7.3.2 L vocalisation

At first glance there appear to be few reasons to object to this feature. The linguistic or language-internal evidence (quadrant 1) includes the absence of /l/ in coda position in Cantonese, but the tendency towards L vocalisation is so widespread that language universals also supply explanations. According to Altendorf (2003: 164), who draws on Lutz (1991) and Labov (1994), this tendency may be part of a universal preference for weakening complex syllable codas in order to approach a CV syllable structure, one that is ‘universally preferred’. This, and the low functional load involved (quadrant 2), explains why L vocalisation is widely observable both synchronically and diachronically. The evidence from language change suggests that this is an active site of change in English.

As Altendorf (2003: 164) also notes, language-internal factors are not sufficient to predict language variation and change; rather, they act as pointers to show where such variation is probable given that the external factors are favourable. In the case of varieties such as Estuary English in the UK these external factors may include ‘acts of identity’ (LePage and Tabouret-Keller 1985) in which the speaker performs a balancing act by using or avoiding features according to an assessment of their prestige (whether overt or covert) for a given audience. In Hong Kong it seems unlikely that L vocalisation acts as an identity-marking or stylistic feature, and its

apparent acceptability is likely to be at least partly due to the difficulty of perceiving it; there were only two reported cases, and this almost certainly reduced its chances of attaining significance. In comparison with TH stopping, L vocalisation may be less likely to affect perceptions of the speaker, simply because it occurs less often and in less salient phonological contexts. However, and as mentioned earlier, more research is needed to inform decision-making in this area.

The favourable results of intelligibility and acceptability considerations could be taken as a signal to stop insisting on the production of postvocalic /l/. But in reality, there seems to be little point in taking a strong position on a feature that is inconsequential and hard to perceive. Time constraints suggest that there will be more important things to cover in a pronunciation syllabus. There is also the interesting possibility that insisting on the use of a feature may actually discourage young people from using it, as noted by Altendorf (2003: 156) in the case of T Glottalling in the UK. All in all, it seems preferable to attempt to raise awareness of it as a variable feature, but this is limited by the sheer difficulty of auditory discrimination. Altendorf (2003: 67) notes that many British NS teachers were unaware of the possibility of what was non-technically described as ‘L dropping’ in their students’ speech. This also raises the question of whether examiners are likely to notice such substitutions. All in all, L vocalisation appears to be one of the most likely candidates for acceptance in English language teaching and testing, as well as being a continuing site of change in many varieties of English. But an evolutionary perspective suggests that this (and other) changes can be allowed to happen naturally. It is likely that many teachers in Hong Kong already use this feature, at least some of the time; it was the most frequently occurring feature in the preliminary study, with 84% of speakers using it at least once. The case for pedagogic intervention is unconvincing, and most language test descriptors do not seem to refer to this feature. There seems to be little point in making recommendations about a feature that very few people can hear.

7.3.3 Consonant cluster simplification

L vocalisation is an example of a universal tendency to weaken complex codas, and the near-universality of final consonant cluster simplification (Schreier 2005) can be seen as another example of this tendency. That Hong Kong speakers appear to extend the scope of final CCR is also unsurprising, given the absence of such clusters in Cantonese. In terms of intelligibility the deletion of /t, d/ in the codas of words such as *privileged* and *relaxed* is unlikely to be problematic, word recognition being well advanced by the time the final consonants are heard (see Schreier 2005: 219). However, a possible caveat in these cases might be the loss of grammatical information involved. Other examples, such as in *government* and *department* (Speaker 7) also appear to follow the intelligibility ‘rules’ by appearing word-finally and being only a small part of the whole word. This principle also applies to singleton coda consonant deletion (e.g. in *attitude*), another instance of the ‘simplify codas’ principle.

The evidence from the acceptability data showed, however, that some types of final CCR had a significant effect on the ratings. Dividing CCR into two categories in order to investigate the salience hypothesis indicated that prevocalic CCR (CCRF-PV) had a greater effect on acceptability. Once again, the greater salience of prevocalic CCR is a likely explanation for the acceptability effects, and suggests that the optimal balance ‘between efficiency and clarity’ (Schneider 2007: 110) may be rather different for speakers, as opposed to listeners. The data also suggested that there were also probable effects arising from morphology. The most frequently marked tokens of CCRF-PV also involved bimorphemic clusters. Further research would be needed to attempt to separate the effects of phonological context and morphological status, and this prevents firm conclusions from being drawn in this area. On the other hand, as final CCR in many cases does not seem to affect intelligibility, one might expect it to extend its range in some varieties of English. The interest of such non-native innovations is that they may point to future developments within a local variety, however conceived, and to directions for language change in general.

7.3.4 Full vowels and the avoidance of schwa

While vowel substitutions in general were found to have a significant effect on acceptability, dividing them into two subtypes (FULL VOWEL, or the non-reduction of full vowels in unstressed syllables, and other types of vowel substitution) revealed that the former subtype did not significantly affect acceptability scores. However, as there were only ten reported cases (the third most infrequent category, after L vocalisation and TH fronting), these results must also be interpreted with care. An initial conclusion is that there are no grounds for penalising this when it occurs, as it seems to be a non-intelligibility threatening, acceptable feature of HKE phonology.

7.3.5 Other possibilities

Among the other possibilities, it is necessary to consider substitutions of the voiceless dental fricative (TH fronting). Although the dental fricative substitutions are often grouped together in discussions of intelligibility and acceptability, the evaluation criteria developed in this study suggest that TH stopping and TH fronting may have rather different characteristics. As has been observed, the distribution of the /θ/ phoneme means that TH fronting is more likely to occur in stressed syllables and therefore to be more prosodically prominent and noticeable. At the same time, it occurs less frequently than the /ð/ phoneme. Unfortunately there were insufficient cases of TH fronting in the data to draw any firm conclusions. In the implicational scale, the frequency of occurrence of TH fronting was much lower than that of TH stopping in terms of the number of speakers using the feature (it was used by six, as opposed to 19, of the 25 speakers).

This suggests that there may be different selection pressures operating upon it, but further research would be needed to clarify this. In terms of intelligibility TH fronting appears unlikely to cause problems, but there are common minimal pairs such as *three* and *free* (as in *free weeks*, an occasional source of confusion in educational settings). TH fronting is also more likely to be salient as it occurs in frequently-occurring monosyllabic (and therefore often stressed) words such as *think*, *thank* and *three*. Another important difference between TH stopping and TH fronting

is that the former occurs widely across varieties of English, whereas the latter has a more restricted distribution (see Chapter 5, section 4.2.3). Looking at the evidence as a whole, while neither of the dental fricative substitutions are therefore likely to be problematic, teaching and learning attention should perhaps be focused on the voiceless TH sound /θ/, if there is sufficient time.

7.4 Recommendations and implications

7.4.1 Teaching pronunciation

Making pedagogical recommendations on the basis of empirical studies is made problematic by variations in areas including classroom conditions, the language proficiency levels of teachers and the attitudes of stakeholders such as parents and principals, as well as the learners themselves. It is perhaps for these reasons that Bolton and Kwok (1990: 170) decide it would be ‘downright dangerous’ to offer any itemised prescriptions for language education in Hong Kong. An initial problem with recommendations that involve ‘simplifying’ the phonological system is the possibility of unintended effects over the longer term, as mentioned above; even though instruction appears to be ineffective in the case of some sounds, removing them from the syllabus may make matters worse. Brutt-Griffler (1998: 387) raises the possibility that ‘the most active agents of the spread of English are actually non-native users, the *teachers* of the language’. If this is so, then a reduced attention to certain sounds may affect the acquisitional patterns of learners, who will form the next generation of teachers, and so on. An evolutionary perspective suggests that changes to syllabi are akin to experiments in artificial selection, and that these may have unforeseen results if the complexity of language systems (or ecosystems) is not taken into consideration.

I have already mentioned the ‘valid truism’ in historical linguistics, that a simplification in one part of the system may lead to a complication in others (Thomason 2009: 358). Thomason mentions the example of the demise of pharyngeal phonemes in Montana Salish, a simplification that has possibly added

one or more new vowel phonemes to the phonemic inventory. Even the seemingly innocuous features identified in this study may be hiding complex resultant changes; Wells (1982: 313) notes that L vocalisation leads to associated vowel neutralisations, so that *doll* becomes a homophone of *dole*. Decisions about teaching pronunciation need to be very carefully considered, because the deletion of seemingly unnecessary features changes their frequency of occurrence in the input and leads to changes elsewhere, changes that may have their own consequences for intelligibility.

This raises the broader question of how far insights from phonology (theory and knowledge about how the sound system of a target language works) should be incorporated into the teaching and learning of pronunciation (the practice and meaningful use of phonological features in speaking; Burgess and Spencer 2000: 191). Similarly, should language teaching aim to reflect language attitudes and language change, or should it aim to be a cause of attitudes and change? Transformationalists such as Kirkpatrick (2006) seem to take the latter position, but this is a question for wider discussion. Once again, whether proposed changes arise from linguistic or political considerations, or a combination thereof, they may have unforeseen consequences. It is possibly unfair to place teachers and students in the role of being agents of language change, as ‘artificially selected’ language systems may not be the best equipped to survive in complex, evolving and unpredictable environments.

From an evolutionary perspective, there may be grounds for intervening as little as possible, but rather ensuring that students are exposed to as much authentic input, from as wide a range of sources and contexts, as they can handle at their stage of development. In the phonological realm, this would involve exposure to a wide range of accents, including those of speakers with the same L1 as the students. This should be done without too much selection for ‘accentedness’, but in a way that bears some relation to the type of accents that students are most likely to encounter in their academic, employment and leisure environments. This view implies a certain faith in language systems as being self-regulating, a position taken by Brutt-Griffler (1998: 388):

Given its process of development, a world language no more needs explicit regulation than a national language. The very sociohistorical processes, the econocultural functions, that called it into being serve to ensure the mutual intelligibility of the language, just as national languages have managed to exist whether consciously regulated by language academies or not, and despite their “actualization” as diverse dialects and registers.

Such a view also implies, however, that too concentrated a focus on ensuring intelligibility may also be misplaced, and that ‘[w]hoever needs to be internationally intelligible...is likely to acquire this skill before too long’ (Schneider 2007: 315). Trusting in the self-regulating power of language and in the ability of language learners to make the necessary adaptations may mean adopting a cautious, ‘precautionary principle’ approach. However, on the other hand it must be pointed out that many language learning environments, including those in Hong Kong, are highly unnatural and are far removed from the sociohistorical and econocultural influences that Brutt-Griffler has in mind. The maintenance of minimum standards of intelligibility appears to be a reasonable goal for pronunciation teaching, then.

Efficient teaching may not necessarily involve spending time on correcting those accent features that have little effect on communication, as in the case of TH stopping and L vocalisation. But it seems almost superfluous to point this out – as has been explained, it is their lack of salience that explains the inconsequential nature of these sounds and suggests their presence or absence may not even be noted, much of the time. The likely absence of certain important features and contrasts in learners’ speech is far more worthy of pedagogical attention; in Hong Kong, priorities may include the lack of voiced/voiceless consonantal contrasts, the conflation of the KIT/FLEECE and TRAP/DRESS vowel phonemes, missing final consonants, even in singleton codas, and the modification of initial clusters. For a learner with these problems, the ‘reduced code’ of the LFC offers little or no reduction in the learning burden.

Another technical indication of this study is that learners may not appreciate the differences between citation forms and connected speech (the hypo- to hyper-articulated continuum, in other words). The student listeners marked many features that are actually common connected speech processes (CSPs), such as syllabic modification, final cluster reduction and final devoicing. These occur ubiquitously in native as well as non-native speech (see Shockey 2003). Although the data analysis suggests that the acceptability of these features depends on how speakers actually employ them, and thus on their degree of salience and noticeability, students may not be aware of the prevalence and naturalness of CSPs.

When planning courses and syllabi, it may be necessary to address non-linguistic factors and consider the learners' attitudes towards varieties of English. They may have mistaken views about the relative intelligibility of native and non-native varieties, or be unaware of the variation that exists within both their own variety and so-called 'standard' varieties. One of the most important implications of this study is that learners in Hong Kong – and perhaps elsewhere – will accept a local accent that does not have salient, intelligibility-reducing features. These accents may even be more acceptable than those of some native speakers, a possibility which calls into question the automatic privileging of native speakers in English teaching programmes such as Hong Kong's NET (Native-speaking English Teacher) scheme.

7.4.2 Testing pronunciation

In the area of speaking tests there is a case for examiners to make their requirements more explicit. The effects of language tests on learning are arguably less direct than those of language teaching, although they may exert a considerable influence on students' attitudes towards different varieties. Tests that are designed to assess a candidate's suitability for academic study, such as IELTS, presumably prioritise international intelligibility. But unless they are provided with guidelines, examiners may be influenced by their own culturally-specific notions of acceptability and downgrade candidates who use certain substitutions that are inconsequential for intelligibility (as suggested by the test descriptor quoted in section 7.2.1 above).

Tests that assess candidates' suitability for teaching (such as the LPAT in Hong Kong) arguably need to set a high standard for pronunciation, as they will partly determine the type of L2 input that students are exposed to. Descriptors should be worded in a feature-specific way so that examiners are not unduly influenced by the use of certain substitutions, and do not tend to assess 'accentedness' in an impressionistic fashion. The possibility of speakers having local accents, but being examples of intelligible and acceptable models, should be emphasised.

7.4.3 The models debate

Given that the study began with the 'models debate', it seems to be necessary to consider its possible implications for the vexed question: which pronunciation model? An important finding of the study has been that student listeners, even fairly proficient ones, do appear to be willing to accept certain types of local accents as pronunciation models. This also implies that these are the type of HKE accents which are 'beginning to be regarded as a positively evaluated source of identification' (Bolton 2000: 277). It should be noted, however, that not all HKE accents were positively evaluated by the listeners in the study. In this sense, the study provides evidence both for and against the contention that native-speaker models no longer provide the most appropriate models for most learners of English. Some local models have been shown to be both acceptable and intelligible, but only insofar – according to one interpretation – as they resemble certain types of native speaker models.

At this point it is necessary to explain another of the study's implications, namely that the use of terminology such as 'native speaker' and 'variety' needs to be reassessed. On the problematic term 'native speaker', Radwanska-Williams (2008: 155) argues that 'we should approach the question of language acquisition not from the perspective of its inception, but of its outcome, and focus on a closer scrutiny of the dynamics of linguistic expertise'. This is in line with the proposals of Kachru (2005: 12), who makes a distinction between 'genetic nativeness' and 'functional nativeness'. There is no need to have a native speaker teacher in order to safeguard intelligibility, and the evidence from this study suggests that high-proficiency local

speakers could serve as intelligible and acceptable models. Whatever the other advantages and disadvantages of local versus non-local, non-native speaker versus native speaker teachers may be, it seems that there are grounds for redefining or replacing the term ‘native speaker’.

However, on the other hand the study’s findings suggest that in order to secure pedagogical acceptability, local teachers need to approach ‘native speaker’ or ‘standard’ norms; indirect evidence from intelligibility studies also supports this conclusion. But the position taken by this study is that there are linguistic explanations for this, and that it need not be seen as a forced convergence on irrelevant native speaker norms. As well as the evolutionary factors already mentioned, an additional centripetal force is created by the lexicon of English, as words form the common currency of intranational and international communication. Without the restraining tendency of spelling, the gravitational pull towards naturalness and ‘the emergence of the unmarked’ would probably be even more pronounced. In an age of increasing literacy, education and information dissemination, channels such as the internet may even increase this centralising tendency. Of course, this is not to say that words need to be pronounced in the same way, merely that the pronunciations need to be within the processable limits of variation for most speakers.

Miscommunication can occur between speakers from the same L1 background, as exemplified by the encounter between Ugandan colleagues related by Abbot (1991: 233):

A: It was impossible because of the [dis'trækʃən].

B: Do you mean ['distrakʃən] or ['destrakʃən]?

Speaker B was unsure of whether Speaker A meant ‘distraction’ or ‘destruction’; apparently both were feasible in Uganda at that time. Interestingly, this intelligibility problem could be due to a tendency towards the merger of TRAP and STRUT (/æ/ and /ʌ/, in RP) in East African English (Mesthrie and Bhatt 2008: 120). The nature of language systems, in this case vowel systems, is that most contrasts need to be maintained if they are to continue to be able to make lexical distinctions. It is

sometimes argued that L2 English speakers with the same L1 background will be able to understand each other's substitutions (Jennifer Jenkins, personal communication), but there are few empirical or theoretical justifications for this view. If a speaker merges vowels, they will be merged for all listeners, unless there are some compensatory disambiguating features.

A counter-argument is that making phonemic systems less complex provides benefits that are greater than the costs of having to occasionally engage in the type of lexical disambiguation shown in the above example. But by and large, L2 varieties of English in Africa and Asia share a number of phonological similarities (Mesthrie and Bhatt 2008: 129). This suggests that a certain amount of convergence is the natural outcome of the interaction of the various evolutionary factors, among them the need to be understood, that operate on varieties of English. Thus there would seem to be greater utility in focussing on what needs to be included in teaching syllabi, rather than on what might conceivably be excluded.

One way to frame the 'models debate' is as a choice between native speaker, local or lingua franca models (e.g., Kirkpatrick 2005). However, if the phonological features of these models are evaluated for intelligibility there appear to be relatively few differences between the internationally intelligible versions of these models. At this stage there may appear to be an irreconcilable conflict between the inevitability and the desirability of local variation and the need for intelligibility. One solution is to downplay the importance of intelligibility; Kachru (1976: 39) believes that in 'overemphasising' the spoken form of a language for intelligibility we are 'under the hangover of the structuralist tenets of language pedagogy'. It may therefore be misleading and unnecessary to make pedagogical recommendations on the basis of intelligibility, particularly as there may be a range of other qualities involved in being a successful language teacher.

Naturally, considerations of intelligibility and acceptability also apply to native speaker accents and varieties. But a features-based analysis is essential: the approach taken by this study suggests that intelligibility does not inhere in varieties, but rather in the phonological features of speech (although of course there may be characteristic

features of varieties that do affect intelligibility). It is therefore somewhat unhelpful to claim, as do Kirkpatrick *et al.* (2008: 359), that ‘some native speaker varieties are not as internationally intelligible as has commonly been supposed’, and that ‘many non-native varieties are more intelligible throughout the world than is often believed, being more intelligible than certain native speaker varieties’. It is the speaker, not the variety, that is intelligible.

Just as we need to make more precise the use of the term ‘native speaker’, we also need to clarify what we mean by ‘varieties’. To rephrase the claims of Kirkpatrick *et al.*, some subvarieties of native speaker varieties may not be as internationally intelligible as has been supposed, and some subvarieties of non-native varieties may be more intelligible throughout the world than is often believed. The varietal type, and thus the type of variation, are crucial considerations. This leads on to the inevitably political nature of the models debate. The intention of some participants appears to be to undermine the perceived dominance of native speaker models, partly by showing that not all of their features are necessary for intelligibility. This is uncontroversial, although it is up to future studies to demonstrate the ways in which some ‘non-core’ features such as rhythm may affect intelligibility. The assertion that such features actually reduce intelligibility is not, for the time being, supported by empirical evidence. Rather than attempting to ‘uncouple’ English from its native speakers, a more profitable undertaking would be to separate the phonological features from their speakers and to further investigate the factors that affect international communication.

Making distinctions at a feature level also leads to the effective neutralisation of the native/non-native distinction in terms of providing intelligible models. If there is a plausible case for intelligibility to be seen as residing in speakers, rather than in varieties themselves, then the concentric circles model proposed by Prodromou (2008) appears to be accurate:

Figure 7.2. The concentric circles of English users (adapted from Prodromou 2008: 259; also in Sewell 2010: 267).

The term ‘unilateral idiomaticity’ was coined by Seidlhofer (2002), and refers to ‘the use of idiomatic language by a speaker...that is not understood by the other participants in the interaction’ (Prodromou 2008: 215). In terms of phonology, it could easily be extended to include ‘unilateral intelligibility’, an insufficient command or awareness of key international intelligibility features. The picture that emerges in some of the literature is one of non-native speakers suffering under the hegemony of native-speaker dominance. It may be preferable to consider the possibility of there being more similarities between high-proficiency users across different varieties than between such users and their low-proficiency counterparts, even those using what are ostensibly the same varieties. Brutt-Griffler (1998: 388) opines that ‘a theory of World English...debunks the myth that there is such a thing as core and periphery in World Englishes’. Of course, this does not mean that differences in power, status and access to education do not exist, but rather that the arena for their discussion needs to be extended beyond the ‘native/non-native’ dichotomy, in order to reflect the complex linguistic, social and psychological dimensions of globalised language use. This is of course Foucault’s position on power relations, namely that they do not stand in an external relationship to other forms of relations such as economic processes and relations of knowledge, and that consequently ‘dominant and dominated enter into relations of power which neither of them control in a simple, absolute way’ (in Flyvbjerg 2001: 121).

Pronunciation teaching should therefore take explicit account of variability, and also be informed by attention to international intelligibility, acceptable local features, accommodation skills and so on, within the constraints imposed by time and the learners' stage of development. As learners' needs are unpredictable, they are best prepared and equipped for a changing world by being aware of these competing dimensions. A desirable outcome of such an orientation to pronunciation teaching is the development of a dual or multi-competence, one that allows users to express both their local and global selves.

7.4.4 Implications for the description of NVEs

Much of what has been said above about pronunciation teaching also applies to the description of new varieties of English. The evidence of variation in HKE phonology found in this study, together with its feature-based assessments of intelligibility and acceptability, support the view that if pedagogical acceptability is desired, the description of new varieties should be framed partly 'in pragmatic rather than a priori terms' (Nelson, 1995: 273). Of course, linguistic description aims to capture the nature of variation without inhibiting it or judging it. Descriptions of NVEs such as Hong Kong English should indeed study the variety in question 'on its own terms' (Hung 2000: 354; see also Mohanan 1992). However, as this study has noted, any consideration of frequency will involve comparing variants with other variants, at least one of which is likely to be associated with 'standard' varieties. A comprehensive perspective that considers variation and development and makes cross-varietal comparisons is helpful, in order to establish the status of NVE features and avoid making potentially misleading generalisations. The inclusion of a wide range of features in some descriptions means that some will be related to transfer and may affect both intelligibility and acceptability. Bolton (2008: 11) also notes the limitations of the 'varieties-based' approach to World Englishes, which has been built on 'an extrapolation (and idealisation) from the individual choices made by individual language users'. Bolton believes that current research points less to the characterisation of varieties as 'uniquely-constituted entities' and more to 'an appreciation of the ways in which the structural features of such varieties develop' (2008: 9). Once again the limitations of the existing terminology are visible. There is

a need for a term that avoids the limitations of the ‘native speaker’ label, and the concept of ‘variety’ needs to be reviewed so that it discourages monolithic interpretations.

The disadvantages of such interpretations are visible in the study of Jenkins (2007), who investigated the attitudes of non-native speakers from various L1 backgrounds towards different L2 English accents. No accent samples were provided, so the study was in fact investigating perceptions of these accents (arguably, stereotypical perceptions). Questionnaire items were designed to assess the correctness, acceptability, pleasantness and familiarity of ten different accents, and the results showed that native-speaker accents were consistently rated more highly on all four items, with Japanese English being the lowest rated on the first three items. However, the interpretability of the findings is limited because it is uncertain how the respondents conceptualised ‘Japanese English’. They may simply have been responding to the fact that English proficiency levels in Japan are relatively low, compared to other countries in the survey such as Germany and Sweden, or even China (in 1998-1999, Japan’s TOEFL scores left it ‘well behind its closest neighbours’, according to Nuttall 2000: 57). The issue of proficiency is discussed further in section 7.4.4 below, but it is clear that we need to differentiate between subvarieties and to establish, while paraphrasing Murakami (2008), ‘what we are talking about when we talk about Japanese English’.

In a similar vein, in the *South China Morning Post* (2009b) a local educator stated that:

There is no problem in Hongkongers using Hong Kong English, when speaking with other local people. However, students have to think about whether it can be used in international situations.

Leaving aside the question of why Hongkongers would wish to use English at all when speaking with other Hongkongers, it is certainly true that low-proficiency Hong Kong English would cause intelligibility problems, as well as being less acceptable for other Hong Kong English users. However, given the inevitability of

local students using Hong Kong English of some form or another, educators of various kinds should be increasing students' awareness of different varieties of English, including the local variety in its low-proficiency and high-proficiency manifestations. There should be no problem at all in students using Hong Kong English in international situations, if they have developed an awareness and a command of the features that maximise its intelligibility.

As well as distorting the perceptions of educators and students, a further disadvantage of non-variationist attitudes towards NVEs is that local users will not find the 'variety' captured by many such descriptions to be acceptable. This may explain the tendency for discussion of 'local varieties' to be largely confined to the discourse of linguists, while local people tend not to take such discussion very seriously (see Joseph 1996 on the case of Hong Kong English). If local people can be persuaded that a local accent exists, and that certain of its features do not threaten international intelligibility, then it will be easier to overcome the prevailing belief that the local variety is merely a collection of errors. In fact, the data on acceptability could be seen as supporting the idea that a local standard is already in operation, 'even if that standard has as yet no recognition or status within the local discourse about English' (Joseph 1996: 175).

The preliminary study can also be seen as an initial description of the high proficiency subvariety of Hong Kong English, as called for by Kirkpatrick (2007b). The fact that there are few differences between this and 'standard' models should not be seen as a rejection of local innovation, as a large amount of variation and differentiation is possible in ways that do not affect intelligibility or acceptability. While many phonemic substitutions were found to be problematic, variation at the subphonemic level will obviously contribute to the uniqueness of the Hong Kong accent. At the lexical and syntactic levels there are also features of Hong Kong English that create distinctiveness, for example the lack of distinction between 'count' and 'non-count' noun phrases that gives rise to terms such as *staffs*, *researches* and *vocabularies*, as noted by Joseph (1996). It is also worth bearing in mind that the apparent similarities between 'high proficiency' local varieties and 'standard' varieties need not be an obstacle to the development of local norms.

According to Joseph (2004: 144) ‘if the desire for a distinct language to be recognised is strong enough, the most minor differences will be invested with the ideological value to fill the bill’.

7.4.5 Implications for studies of ELF (English as a Lingua Franca)

The desire to ‘uncouple’ English from its native speakers is visible in some of the ELF literature (for example, Seidlhofer, Breiteneder and Pitzl 2006). Seidlhofer (2010) observes that although ELF is a heterogeneous phenomenon, successful communication does take place ‘in millions of interactions’ and there must therefore be a ‘significant core of lexical, grammatical and phonological elements that make this possible’. At the phonological level, however, the core elements that make this possible appear to be common to most varieties, and the non-core elements are also widely distributed as synchronically variable or diachronically unstable forms. The nature and origin of these similarities is acknowledged by some ELF researchers. Dewey (2005) focuses on lexical and grammatical innovations in ELF, concluding that these indicate emerging patterns and are ‘systematic, frequent and communicatively effective’ (in Seidlhofer *et al.* 2006: 12). Although the present study is not concerned with lingua franca communication, there is some evidence that the phonological features of an ‘emerging’ Hong Kong English (such as TH stopping and L vocalisation) are also those that are systematic, frequent and communicatively effective in that they would not be expected to reduce intelligibility in international communication.

Dewey’s conclusions can be tentatively explained in a similar manner to that employed by the present study: in ecological terms, the ELF environment is likely to have significant differences from L1 contexts. As in L2 contexts, it may be that extralinguistic factors are less important. ELF ‘communities’ will typically be temporary and heterogeneous (see James 2005), meaning that prestigious groups or speakers are less likely to exert lasting pressures. Developments in the language that might have been expected to take place earlier, on such grounds as naturalness and the avoidance of marked forms, are hastened. They could be seen as innovative testing grounds for forms that may well spread to other varieties in due course.

However, except in the case of the ‘systematic, frequent and communicatively effective’ features such as L vocalisation and TH stopping, this study has not provided much support for the notion that L2 speech communities are less exacting in their judgments of phonological variation than are native speakers. Although further studies are needed to investigate the differences between native speaker and non-native speaker perceptions, it seems likely that native speaker judges would evaluate accentedness and ease of understanding in similar ways. While linguistically unsophisticated judgments may differ, an example of proficiency rating is provided by the public version of the IELTS (International English Language Testing System) descriptors for pronunciation (IELTS 2010):

Band 8 (very good user): uses a wide range of pronunciation features...is easy to understand throughout...L1 accent has minimal effect on intelligibility.

Band 6 (competent user): uses a range of pronunciation features with mixed control...can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times.

Band 4 (limited user): uses a limited range of pronunciation features...mispronunciations are frequent and cause some difficulty for the listener.

These bands could be used to characterise the speakers used in the main study, who probably vary between about Band 6 (‘competent user’) and about Band 9 (‘expert user’). This suggests that ‘proficiency’, a concept which has hitherto received little attention in either ELF and World Englishes (see Bolton 2008), is not necessarily an ‘exonormative’ construct but one which has relevance for all users of a language, whatever their L1 background. Of course, this depends on how it is interpreted, and although the descriptors above do not show any inherent bias the ones discussed in section 7.2.1 above are potentially misleading for examiners. Perceptions of proficiency by non-native listeners may depend on the proficiency profile and social background of the listeners, who as human beings will tend to rate more highly those who show similar characteristics.

Finally, although the concept of proficiency does not have to be biased towards particular varieties, it tends to be associated with the SLA paradigm and its implicit assumptions about ‘target language’, ‘interlanguage’ and so on. This is often positioned as being antithetical to the idea of autonomous, emerging varieties. However, the conclusion of this study is that while Hong Kong English is not merely an interlanguage, certain aspects of variation and development can be partly explained from within an SLA framework that considers, for example, the interaction of transfer and developmental features (e.g. Hansen 2006) and the role of extralinguistic factors in determining ‘ultimate achievement’ (Moyer 2004).

7.5 Overall review of the study

7.5.1 Limitations of the study

As with any study of this nature, the present study has several limitations which may limit the interpretability and applicability of its findings. Firstly, the samples in both the preliminary study and the main study represented authentic language data. As has been mentioned in Chapter 4, this involved a compromise between the desirability of authenticity and the need to control extraneous variables. One of the most serious threats to the internal validity of the study is thus the possibility that unmeasured factors in the accent samples in the main study affected the acceptability scores. The research design attempted to minimise this by selecting key variables and measuring their effects, and even though phonological accuracy was found to be the most important factor, the amount of variation explained by the regression equation is fairly low. Among the unmeasured linguistic factors that could affect perceptions of an accent are prosodic factors, such as the degree to which samples were stress timed or syllable timed. As syllable timing is often considered to be a characteristic feature of Hong Kong English, the lack of data in this area is a significant limitation. Unmeasured extralinguistic factors include the topic (some topics may have engendered emotional reactions from students), the age of the speakers and indefinable elements of voice quality that may have served to create or undermine perceptions of confidence and authority, thus masking the effects of linguistic

factors. However, these problems seem to be mainly inherent to the use of authentic samples, and the possibility of unmeasured (or unmeasurable) factors also exists with more controlled samples.

In the main study, the nature of ‘acceptability’ was, as has been discussed, somewhat problematic. The study attempted to deal with this by limiting the term to meaning ‘acceptability for pedagogical purposes’, partly through the use of a questionnaire item (item E) that was specifically directed at this. The study therefore has nothing to say about whether some features might possess covert prestige, in addition to the overt prestige assumed by the term ‘acceptability’. There is thus the danger that the students were assessing the samples merely in terms of their perceived deviation from native speaker norms, rather than as *bona fide* examples of the local variety. The usual statistical procedures were employed to ascertain the degree of inter-item correlation, and these generally supported the construct validity of the term. Nevertheless, the inclusion of items relating to comprehensibility with those relating to acceptability could be criticised, although if acceptability is limited to the pedagogical domain, comprehensibility or ease of understanding seems to be a natural component. However, the consistency of the acceptability ratings suggests either that some kind of common underlying construct was being assessed, or that these components are in fact independent but related concepts.

In Part 2 of the main study, the research procedure makes the important assumption that students were able to accurately process the samples in terms of recognising their features and being able to assign them to phonological categories. In fact, this assumption was largely confirmed by the results; error codings were checked for accuracy, and relatively few codings were rejected. Similarly, relatively few errors were missed (in the sense of not being noticed by any students), and the position of this study is that missed identifications were mainly a result of low noticeability rather than poor listening skills. The noticeability of features is one of the factors that influenced statistical significance, but it is also a determinant of evaluative or social significance (features that are unnoticed by a majority of listeners have little effect on speaker evaluations).

As well as the noticeability of features, there were other, related difficulties of interpretation arising from the nature of the survey design. Although the frequency and accuracy of error identification was sufficient to enable statistically-based comparisons to be made between features, the codings were also thought to be affected by other factors. These included frequency of occurrence (of both possible and actual occurrences of the features) and the severity with which features were viewed by the students. The use of authentic speech data suggests that phonemes occurred at near-natural frequencies, although this was not measured. The frequency of error marking was therefore assumed to be a result of some combination of noticeability (or salience), and perceived severity (or stigmatisation). To some extent the research design had only limited ability to deal with the differences between salience and stigmatisation. Thus it was difficult to decide whether errors were insignificant because they were non-stigmatised, even though salient, or merely because they were not detected or marked often enough to achieve statistical significance. In other words, the attempted separations between non-salient, non-stigmatised error types and salient, non-stigmatised error types had no clear statistical basis, although the number of codings provided a rough indication. For example, as TH stopping received more codings than VOWELSUB (30 against 27) it can reasonably be assumed that its non-statistically significant effects on the ratings were due to it being less stigmatised overall than VOWELSUB; it is unlikely to have been a result of there being insufficient cases in the data.

However, the effects of frequency of marking were not always clear, especially in the case of non-significant items. The low significance of TH fronting, for example, is almost certainly related to the low number of codings (four), and no definite conclusions about its effects are possible. Generally, word tokens containing errors were marked by relatively few students, and many errors were only marked by a minority of students. This may reflect disagreement about the noticeability or importance of errors, but it is also probably related to the open-ended survey design. In this regard the samples were possibly too long to generate consistent patterns of error marking, but from another perspective shorter samples would have been less able to provide a sufficient number of contexts for feature use. Longer samples could have been used to increase the incidence of contexts for TH-fronting, for example,

and remove the uncertainty that resulted from there being few word tokens for this feature. However, there are problems involved with the use of longer samples, such as the cognitive and statistical overload resulting from a large number of errors.

Similarly, while the use of high-proficiency samples addressed one of the limitations of previous studies, it brought with it limitations of its own in that there tended to be fewer instances of certain features amongst the chosen speakers. TH fronting was infrequent, and [n/l] conflation did not occur. On the other hand, certain features of Hong Kong English, such as the devoicing of voiced fricatives, occurred very widely but were not included in the analysis (as explained in Chapter 3, section 3.6.2). The study cannot, therefore, claim to be a comprehensive investigation of segmental features; the neglect of suprasegmental features has already been acknowledged.

The relationship between Part 1 and Part 2 of the study, and hence between ‘global’ ratings of acceptability and the ‘local’ occurrence of features, is somewhat problematic. The survey design assumes that the former was determined to some extent by the latter, and although this was borne out by the regression equations the percentage of variation explained was fairly low. This was obviously due to there being other causative factors, as mentioned above, but was probably also a result of inter-rater variations in error identification and coding. It is also conceivable that the errors were not actually very important to the students, and that their being required to mark them overstates their contribution to the acceptability scores. Nevertheless, the statistical analysis was still able to indicate the relative effects of different errors, even if their importance of their effects in overall terms was somewhat uncertain.

Another difficulty of interpretation relates to the possibility of there being differences between speakers in terms of the ways they use features. For example, although TH stopping did not have significant effects on acceptability, this may have been due to the ways in which the speakers concerned used the feature. Although intra-speaker variation was not addressed by the study, the speakers did not appear to use this feature categorically. It is possible that different patterns of TH stopping might have caused this feature to be significant.

In both the preliminary and the main study, the samples may not have been very representative of Hong Kong English phonology, further limiting the generalisability of the findings. Choosing speakers from television programmes meant restricting them to quite a narrow range of ages and occupations (not to say gender; ten of the twelve samples in the main study were of male speakers). The students may therefore have been evaluating samples that had little to do with the way they actually speak English, and thus with the way that English is generally used in Hong Kong. However, discussions of which speakers 'represent' Hong Kong English are likely to be unproductive, and the need for a variationist approach is again indicated. A related limitation of the study is of course the sampling bias inherent in using a group of similar students to assess the samples. It would be necessary to repeat the surveys with different groups in order to make more confident conclusions about the acceptability of Hong Kong English. Among the factors that might affect acceptability ratings are the age, proficiency level, gender, and educational and social background of the listeners. In this study, most of the listeners were female; Bolton and Kwok (1990) found that female listeners were more likely to prefer standard forms, a finding that has been replicated in several studies. They were of a broadly similar proficiency level. As university students and English majors this level was almost certainly above the average, but they may still have lacked linguistic or sociolinguistic sophistication. The views of teachers would provide valuable data, as they are likely to be among the principal agents of language transmission and change in Hong Kong.

The fact that the samples were chosen by a native speaker, who may have had different criteria to those of Hong Kong English users, could be argued to have introduced a different kind of sampling bias. However, once again the implicational scale proved to be a useful measuring device. The scale ordering is a result of feature occurrence and is thus independent of arbitrary judgments of proficiency. By selecting samples that had more, or fewer, HKE features, the desired range of samples was obtained. Another possible source of native speaker bias occurred at the error coding stage, when some error codings (mainly those relating to suprasegmental features) were ignored as a result of irrecoverable ambiguity. It could be contended that this selection obscured potentially significant types of error and

shoehorned the error codings into preexisting categories; however, the introduction of a new category (SYLL) suggests that the procedure was receptive to patterns in the data. The difficulty of analysing the error codings could be seen as a design flaw, and certainly the analysis would have been simplified by improving the marking procedure in some way, perhaps through the use of an online survey form. The abundance of errors in some samples suggests a more general disadvantage of the use of authentic samples, the length of which must be carefully considered.

A general limitation of the study's approach is that it is mainly quantitative, and although the student comments were useful in understanding the error codings, more qualitative data would have helped to understand the responses to particular errors, some of which may have had culturally-specific or L1-specific aspects that were unknown to the researcher. Finally, the ability of the study to provide pedagogical recommendations was quite limited. For example, the fact that the acceptability scores were relative values makes it difficult to ascertain whether the highly rated samples would actually be acceptable as classroom models, although the use of a native speaker accent acted as a comparison. The study could also be criticised for its lack of consideration of the actual conditions in local educational institutions. While some aspects of the sociocultural and educational context were considered in Chapter 3, the applicability of the study's pedagogical recommendations are limited by the lack of any ethnographic data regarding the attitudes and priorities of local teachers, principals and parents. The differences between EMI and CMI schools were not considered, but these may significantly affect the choice of pronunciation models and priorities. Similarly, the priorities of any teaching syllabus obviously depend on the age, level and motivation of the students.

7.5.2 Achievements of the study

Viewed in its entirety, and despite its limitations, the study can be seen as having broken new ground in the study of new varieties of English. By adopting a features-based, variationist perspective, it has been able to provide more detail about 'Hong Kong English' and show how certain of its features varied in terms of their distribution and acceptability within particular subgroups of the community. By

using authentic speech samples, the study captured the ‘layering of errors, deviations and inconsistencies’ (van den Doel 2006: 304) that occur in some types of L2 speech. The use of broadcast material to create a mini-corpus was both innovative in itself and functionally effective; the overall survey of feature use in the preliminary study enabled some initial observations to be made about the distribution of certain features. This enabled the study to concentrate on frequently occurring forms. The use of broadcast material also led to an automatic focus on speakers of relatively higher proficiency levels. This was both intrinsically useful, as it represented an initial description of English use by proficient bilinguals, and also innovative in studies of Hong Kong English phonology. The associated implicational scale proved to be a useful depiction of feature use, as it showed how features tend to co-occur and provided a rationale for sample selection. A consideration of the reasons behind the implicational patterns laid the foundations for the later, more general explanatory model.

The focus of the main study was on acceptability, and it represents one of the most detailed accounts of the reactions of L2 English users to variable accent features within their own variety. It also discovered that certain types of Hong Kong English accents appeared to be acceptable for pedagogical use, and were not greatly different to the native speaker sample in terms of their pedagogical acceptability (although a fuller evaluation of this issue would require a broader range of accents, in order to avoid the study’s own *bête noire* of making generalisations). Both the explanatory and evaluative aspects of the study involved a multidimensional perspective that integrated considerations of variation, development, intelligibility and acceptability. The study’s focus on acceptability was thus framed within a principled consideration of other important factors, such as intelligibility, that both formed criteria for the evaluation of features and were themselves part of the explanatory model. This allowed the important questions of *what* features were acceptable and *why* they were acceptable to be addressed, although the explanatory approach was characterised as being ‘ecological’ and therefore permitting multiple causative factors.

The study also generated some recommendations for language teaching and testing, although these were not of a detailed nature. In Hong Kong, there seem to be no

obstacles to the use of a high-proficiency version of the local variety as a teaching model. It is hoped that the study will form a constructive contribution to the debate about pronunciation models, as well as providing useful insights for future research into World Englishes.

7.5.3 Directions for future research

As general principles, the present study has emphasised the need for a variationist perspective in World Englishes research. This is best summarised by the contention that it is features and speakers, and not varieties *per se*, that must be considered in evaluations of intelligibility and acceptability. The limitations of certain generalisations, as indicated by the study, also suggest that both terminology and methodology used within the field may be in need of an overhaul. At several points it was found that dichotomies such as ‘native speaker/non-native speaker’, and the undifferentiated use of terms such as ‘variety’, were unable to capture the complexity of World Englishes in terms of structure and attitudes. Studies in these areas should maintain an awareness of these limitations, and where appropriate should work towards overcoming them by developing an improved nomenclature.

In terms of specific avenues for research, the study’s strengths and weaknesses indicate several interesting directions. The interactions between the areas of intelligibility, comprehensibility and acceptability are complex and studies need to be extended to different contexts, using different types of speakers and listeners. One possibility would be to investigate the differences between native speaker and non-native speaker perceptions of the same accent samples, in order to establish the similarities and differences between them. This study did not assess intelligibility directly, and there is a great need for more detailed investigations of the effects of different types of error. While the study has argued for the benefits of using authentic samples, there is also a role for the use of more controlled samples. These could be obtained through various methods, including standardised reading passages and the matched-guise technique, in which case the findings of the implicational scale might be of use in informing the selection of ‘errors’. The inclusion of a wider range of features would also be possible with a matched-guise approach.

One possible extension of the study would be to provide listeners with one-word samples containing errors and either measure detection rates, thus assessing the salience of the errors, or ask listeners to rank the errors in order of perceived severity. In measuring the properties of accent samples, attention to the suprasegmental level would provide a fuller picture of the sources of rating differences; the Pairwise Variability Index (PVI) developed by Low, Grabe and Nolan (2000) offers an accessible measurement of certain prosodic characteristics. Another extension of the study would be to repeat the experiment with listeners of different proficiency levels, possibly also controlling for other variables such as gender. This would allow a more detailed picture of the factors affecting acceptability to emerge. In such a study, it would also be interesting to investigate the extent to which listeners' perceptions and judgments are related to their own patterns of feature use.

Finally, as several observers (e.g. Bell 1984; Eckert 2000) have noted the possibility that patterns of intra-speaker variation tend to be derived from those present in inter-speaker variation, another interesting direction for research would be to investigate how speakers vary in their use of phonological forms according to the situation or context. Patterns of intra-speaker variation could then be compared with the inter-speaker variational patterns shown in the implicational scale. This would add an extra dimension to the evaluation of features, by showing which features play a role in intra-speaker or stylistic variation.

7.5.4 Final words

While the role and status of Hong Kong English is ultimately a matter for its users to decide and shape, it is hoped that the present study will contribute to discussion. One of the study's counsels is 'beware of generalisations', as far as is humanly possible. On the one hand, generalisations are unavoidable. This study is no exception: the implicational scale and the intelligibility-acceptability correspondence both involve a certain amount of generalisation. On the other hand, one of the dissertation's aims has been to 'test out' particular generalisations, such as 'Hong Kong English should be avoided in the Hong Kong classroom', or 'non-native models are less acceptable than native models'. It is not true that all forms of Hong Kong English need to be

avoided, whether in international communication or as models in local classrooms. It is misleading to link qualities such as intelligibility and acceptability with a speaker's status as 'native' or 'non-native'. Such generalisations disadvantage non-native speaker teachers, for example, but they also polarise discussion instead of focussing it on the phonological features that successful users have in common. Insightful research must address the linguistic and social complexities of language use by both types of speakers, within their diverse communities.

REFERENCES

- Abbott, G. (1991) A new look at phonological 'redundancy'. In Brown, A. (ed.) *Teaching English Pronunciation: A Book of Readings*. London and New York: Routledge, 225-234.
- Altendorf, U. (1999) Estuary English: is English going Cockney? *Moderna Språk* XCII 1-11. Available: <http://www.phon.ucl.ac.uk/home/estuary/altendf.pdf>
Accessed: 11 May, 2008.
- Altendorf, U. (2003) *Estuary English: Levelling at the Interface of RP and South-Eastern British English*. Tübingen: Gunter Narr Verlag.
- Asma, S. (1996) Darwin's causal pluralism. *Biology and Philosophy* 11, 1-20.
- Auer, P. and Hinskens, F. (1996) The convergence and divergence of dialects in Europe: new and not so new developments in an old area. *Sociolinguistica* 10, 1-30.
- Austin, J. (1962) *How To Do Things With Words*. Edited by Urmson, J.O. and Sbisà, M. Oxford: Oxford University Press.
- Bacon-Shone, J. and Bolton, K. (1998) Charting multilingualism: Language censuses and language surveys in Hong Kong. In Pennington, M. (ed.) *Language in Hong Kong at Century's End*. Hong Kong: Hong Kong University Press, 43-90.
- Bailey, C. (1982) The garden path that historical linguistics went astray on. *Language and Communication* 2, 151-160.
- Bailey, C. (1996) *Essays on time-based linguistic analysis*. Oxford: Oxford University Press.
- Bamgbose, A. (1998) Torn between the norms: innovations in world Englishes. *World Englishes* 17/1, 1-14.
- Bansal, R. K. (1969) *The Intelligibility of Indian English*. Hyderabad: Central Institute of English and Foreign Languages.
- Bauer, L. (1986) Notes on New Zealand English phonetics and phonology. *English World-Wide* 7, 225-258.
- Bayard, D., Weatherall, A., Gallois, C. and Pittam, J. (2001) Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics* 5/1, 22-49.
- Bayley, R. (2002) The quantitative paradigm. In Chambers, J.K., Trudgill, P. and Schilling-Estes, N. (eds) *The Handbook of Language Variation and Change*. Oxford: Blackwell, 117-41.

- Beinhoff, B. (2005) *Non-native speakers' perceptions of their foreign accent of English: attitudes and identity*. Unpublished MPhil research essay. Research Centre for English and Applied Linguistics, Faculty of English, University of Cambridge.
- Bell, A. (1984) Language style as audience design. *Language and Society* 13, 145-204.
- Berns, M. (2008) World Englishes, English as a Lingua Franca and intelligibility. *World Englishes* 27, 327-334.
- Berry, W.D. and Feldman, S. (1985) *Multiple Regression in Practice*. Thousand Oaks, CA: Sage Publications.
- Bex, T. and Watts, R.J. (eds) (1999) *Standard English: the Widening Debate*. London: Routledge.
- Bickerton, R. (1973). The structure of polylectal grammars. In Shuy, R. (ed.) *Sociolinguistics: Current Trends and Prospects. Georgetown University Roundtable on Languages and Linguistics 1972*. Washington DC: Georgetown University Press, 17-42.
- Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Bolton, K. (2000) The sociolinguistics of Hong Kong and the space for Hong Kong English. In Bolton, K. (ed.) *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 29-56.
- Bolton, K. (2008) English in Asia, Asian Englishes, and the issue of proficiency. *English Today* 24/2, 3-12.
- Bolton, K. and Nelson, C. (2000) Analysing Hong Kong English: sample texts from the International Corpus of English. In Bolton, K. (ed.) *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 241-264.
- Bolton, K. and Kwok, H. (1990) The dynamics of the Hong Kong accent: social identity and sociolinguistic description. *Journal of Asian Pacific Communication* 1/1, 147-172.
- Boomer, D. (1978) The phonemic clause: speech unit in human communication. In Siegman, A. and Feldstein, S. (eds) *Nonverbal Communication and Behaviour*. Hillsdale, NJ: Lawrence Erlbaum, 245-262.
- Bourdieu, P. (1991) *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
- Bott, O. and Radó, J. (2007) Quantifying quantifier scope: a cross-methodological comparison. In Featherston, S. and Sternefeld, W. (eds) *Roots: Linguistics in Search of its Evidential Base*. Berlin: Walter de Gruyter, 53-74.

- Braine, G. (ed.) (1999) *Non-native Educators in English Language Teaching*. Mahwah, NJ: Lawrence Erlbaum.
- Braine, G. (ed.) (2005) *Teaching English to the World: History, Curriculum, and Practice*. Mahwah, NJ: Lawrence Erlbaum.
- Britain, D. and Fox, S. (2009) The regularisation of the hiatus resolution system in British English: a contact-induced ‘Vernacular Universal’? In Filppula, M., Klemola, J. and Paulasto, H. (eds) *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. New York and London: Routledge, 177-205.
- Brown, G. (1974). Practical phonetics and phonology. In Allen, J.P.B. and Corder, S.P. (eds) *The Edinburgh Course in Applied Linguistics Vol. 3: Techniques in Applied Linguistics*. Oxford: Oxford University Press.
- Brown, A. (1989) Some thoughts on intelligibility. *The English Teacher* 18. Available: <http://www.melta.org.my/ET/1989/main4.html> Accessed 21 November, 2007.
- Brown, A. (1991) *Pronunciation Models*. Singapore: Singapore University Press.
- Brown, A. and Deterding, D. (2005) A checklist of Singapore English pronunciation features. In Deterding, D., Brown, A. and Low, E.L. (eds) *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill, 7-13.
- Brown, G. (1974) Practical phonetics and phonology. In Allen, J.B.P. and Corder, S.P. (eds) *The Edinburgh Course in Applied Linguistics Vol.3: Techniques in Applied Linguistics*. Oxford: Oxford University Press.
- Bruthiaux, P. (2003) Squaring the circles: issues in modeling English worldwide. *International Journal of Applied Linguistics* 13/2, 159-178.
- Brutt-Griffler, J. (1998) Conceptual questions in English as a world language: taking up an issue. *World Englishes* 17/3, 381-392.
- Brutt-Griffler, J. (2002) *World English: a Study of its Development*. Clevedon: Multilingual Matters.
- Burgess, J. and Spencer, S. (2000) Phonology and pronunciation in integrated language teaching and teacher education. *System* 28, 191-215.
- Canagarajah, A.S. (1999a) *Resisting Linguistic Imperialism in Language Teaching*. Oxford: Oxford University Press.
- Canagarajah, A.S. (1999b) On EFL teachers, awareness, and agency. *ELT Journal* 53/3, 207-214.
- Canagarajah, A.S. (1999c) Interrogating the “native speaker fallacy”: non-linguistic roots, non-pedagogical results. In Braine, G. (ed.) *Non-native Educators in English Language Teaching*. Mahwah, NJ: Lawrence Erlbaum, 77-92.

- Canagarajah, A.S. (2005) Introduction. In Canagarajah, A. (ed.) *Reclaiming the Local in Language Policy and Practice*. Mahwah, NJ: Lawrence Erlbaum.
- Candler, R. (2001) *An investigation into Hong Kong non-native speakers' recognition of and attitudes towards different accents of English*. MA dissertation. Hong Kong: University of Hong Kong.
- Cargile, A. and Giles, H. (1997) Understanding language attitudes: exploring listener affect and identity. *Language and Communication* 17/3, 195-217.
- Chambers, J. (2002) Studying language variation: an informal epistemology. In Chambers, J.K., Trudgill, P. and Schilling-Estes, N. (eds) *The Handbook of Language Variation and Change*. Oxford: Basil Blackwell, 3-14.
- Chan, A.Y.W. and Li, D.C.S. (2000) English and Cantonese phonology in contrast: explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum* 13/1, 67-85.
- Cheshire, J. (1996) Syntactic variation and the concept of prominence. In Klemola, J., Kytö, M. and Rissanen, M. (eds) *Speech Past and Present: Studies in English Dialectology in Memory of Ossi Ihalainen*. Frankfurt: Peter Lang, 1-17.
- Chia, B.P. and Brown, A. (2002) Singaporeans' reactions to Estuary English. *English Today* 18, 33-38.
- Childs, B. and Wolfram, W. (2004) Bahamian English: phonology. In Schneider, E., Burridge, B., Kortmann, B., Mesthrie, R. and Upton, C. (eds) *A Handbook of Varieties of English*. Vol. 1: *Phonology*. Berlin: Mouton de Gruyter, 435-449.
- Chomsky, N. (2000) *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- City University of Hong Kong (1999) What medium of instruction for Hong Kong? *Bulletin* 19. Hong Kong: University Publications Office, City University of Hong Kong.
- Coetzee-Van Rooy, S. (2009) Intelligibility and perceptions of English proficiency. *World Englishes* 28/1, 15-34.
- Cohen, J. (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Cohen, L., Manion, L. and Morrison, K. (2007) *Research methods in education*. London: Routledge.
- Collins, B. and Mees, I. M. (2003) *Practical Phonetics and Phonology: a Resource Book for Students*. London: Routledge.
- Coniam, D. and Falvey, P. (2002) Selecting models and setting standards for teachers of English in Hong Kong. *Journal of Asian Pacific Communication* 12/1, 13-38.

- Cruttenden, A. (1986) *Intonation*. Cambridge: Cambridge University Press.
- Crystal, D. (1996) The past, present and future of English rhythm. *Speak Out!* 18, 8–13.
- Crystal, D. (1997) *English as a Global Language*. Cambridge: Cambridge University Press.
- Crystal, D. (2003) *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.
- Crystal, D. (2004) *The Stories of English*. London: Penguin.
- Dalton, C. and Seidlhofer, B. (1994a) *Pronunciation*. Oxford: Oxford University Press.
- Dalton, C. and Seidlhofer, B. (1994b) Is pronunciation teaching desirable? Is it feasible? In Sebbage, T. and Sebbage, D. (eds.) *Proceedings of the 4th International NELLE Conference*. Hamburg: NELLE.
- Damasio, A. (1994) *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Putnam.
- Date, T. (2005) The intelligibility of Singapore English from a Japanese perspective. In Deterding, D., Brown, A. and Low, E.L. (eds) *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill, 173-183.
- Dauer, R.M. (2005) The Lingua Franca Core: a new model for pronunciation instruction? *TESOL Quarterly* 39/3, 543-550.
- DeCamp, D. (1971) Toward a generative analysis of a post-creole speech continuum. In Hymes, D. (ed.) *Pidginization and Creolization of Languages*. Cambridge: Cambridge University Press.
- Denes, P.B. (1963) On the statistics of spoken English. *Journal of the Acoustical Society of America* 35, 892-904.
- Deterding, D., Brown, A. and Low, E.L. (eds) (2005) *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill.
- Deterding, D. and Kirkpatrick, A. (2006) Emerging South-East Asian Englishes and intelligibility. *World Englishes* 25 (3-4), 391-409.
- Deterding, D., Wong, J. and Kirkpatrick, A. (2008) The pronunciation of Hong Kong English. *English Worldwide* 29/2, 148-175.
- Derwing, T.M. and Munro, M.J. (1997) Accent, intelligibility and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition* 20, 1-16.

- Dewey, M. (2005) English as a lingua franca in a framework of globalization. Paper given at the Annual Meeting of the British Association for Applied Linguistics (BAAL) at the University of Bristol, UK.
- Dewey, M. (2007) English as a lingua franca and globalization: an interconnected perspective. *International Journal of Applied Linguistics* 17/3, 332-354.
- Dressler, W.U. (1984) Explaining Natural Phonology. *Phonology Yearbook* 1, 29-51.
- Drucker, P.F. (1954) *The Practice of Management*. New York: Harper and Row.
- Dunn-Rankin, P. (1983) *Scaling Methods*. Hillsdale, NJ: Lawrence Erlbaum.
- Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) (2005) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang.
- Eckert, P. (2000) *Linguistic Variation as Social Practice*. Oxford: Blackwell.
- Eckman, F. (1977) Markedness and the contrastive analysis hypothesis. *Language Learning* 27/2, 315-330.
- Eckman, F. (2008) Typological markedness and second language phonology. In Hansen-Edwards, J.G. and Zampini, M. (eds) *Phonology and Second Language Acquisition*. Amsterdam and Philadelphia: John Benjamins, 95-115.
- Edwards, M.L. (1979) *Patterns and processes in fricative acquisition: Longitudinal evidence from six English-learning children*. Unpublished PhD dissertation, Stanford University.
- Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Faerch, C. and Kasper, G. (1983) (eds) *Strategies in Interlanguage Communication*. London: Longman.
- Fant, G. (1969) Distinctive features and phonetic dimensions. *STL-QPSR (Speech Transmission Laboratory. Quarterly Progress and Status Reports)* 10/2-3, 1-18.
- Farrar, K. and Jones, M. (2002) Introduction. In Jones, M. and Esch, E. (eds) *Language Change: the Interplay of Internal, External and Extra-linguistic Factors*. Berlin: Mouton de Gruyter, 1-16.
- Fasold, R. (1990) *The Sociolinguistics of Language*. Oxford: Basil Blackwell.
- Fayer, J.K. and Krasinski, E. (1987) Native and nonnative judgments of intelligibility and irritation. *Language Learning* 37/3, 313-326.
- Feldt, L. (1990) The sampling theory for the intraclass reliability coefficient. *Applied Measurement in Education* 3, 361-367.

- Filppula, M., Klemola, J. and Paulasto, H. (2009) Overview. In Filppula, M., Klemola, J. and Paulasto, H. (eds) *Vernacular Universals and Language Contacts: evidence from varieties of English and beyond*. New York and London: Routledge, 1-16.
- Flyvbjerg, B. (2001) *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again*. Cambridge: Cambridge University Press.
- Forde, K. (1995) A study of learner attitudes towards accents of English. *Hong Kong Polytechnic University Working Papers in ELT and Applied Linguistics* 1, 59-76.
- Foulke, E. (1968) Listening comprehension as a function of word rate. *Journal of Communication* 18, 198-206.
- Foulke, E. and Sticht, T. (1966) Listening rate preferences of college students for literary material of moderate difficulty. *Journal of Auditory Research* 6, 397-401.
- Fraser Gupta, A. (2005) Inter-accent and inter-cultural intelligibility: a study of listeners in Singapore and Britain. In Deterding, D., Brown, A. and Low, E.L. (eds) *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill, 138-152.
- Fry, D.B. (1947) The frequency of occurrence of speech sounds in southern English. *Archives Néerlandaise de Phonétique Experimentale* 20, 103-106.
- Gandhi (1998) *Postcolonial Theory: a Critical Introduction*. New Delhi: Oxford University Press.
- Gass, S. and Varonis, E. (1984) The effect of familiarity on the comprehensibility of non-native speech. *Language Learning* 34, 65-89.
- Gek, H.M. and Deterding, D. (2005) Reduced vowels in conversational Singapore English. In Deterding, D., Brown, A. and Low, E.L. (eds) *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill, 54-63.
- Gimson, A. C. (1980) *An Introduction to the Pronunciation of English* (3rd edn). London: Edward Arnold.
- Gimson, A. C. and Cruttenden, A. (1994) *Gimson's Pronunciation of English* (5th edn). London: Edward Arnold.
- Glenwright, P. (2005) Grammar Error Strike Hard: Language Proficiency Testing of Hong Kong Teachers and the Four 'Noes'. *Journal of Language, Identity and Education* 4/3, 201-226.
- Gnutzmann, C. (2005) 'Standard English' and 'World Standard English'. Linguistic and pedagogical considerations. In Gnutzmann, C. and Intemann, F. (eds) *The Globalisation of English and the English Language Classroom*. Tübingen: Gunter Narr, 107-118.

- Goldsmith, J. (1996) Phonological theory. In Goldsmith, J. (ed.) *The Handbook of Phonological Theory*. Oxford: Blackwell, 1-23.
- Gordon, E. and MacLagan, M. (2004) Regional and social differences in New Zealand: phonology. In Schneider, E., Burridge, B., Kortmann, B., Mesthrie, R. and Upton, C. (eds) *A Handbook of Varieties of English*. Vol. 1: *Phonology*. Berlin: Mouton de Gruyter, 603-613.
- Graddol, D. (2006) *English Next*. London: British Council.
- Gramsci, A. (1971) *Selections from the Prison Notebooks of Antonio Gramsci* (Hoare, Q. and Smith, N., compilers and eds). London: Lawrence and Wishart.
- Grau, M. (2005) English as a global language: what do future teachers have to say? In Gnutzmann, C. and Intemann, F. (eds) *The Globalisation of English and the English Language Classroom*. Tübingen: Gunter Narr, 275-284.
- Guion, S., Flege, J., Liu, H. and Yeni-Komshian, G. (1997) Sentence duration as an overall indicator of proficiency in an L2. Paper presented at the Meeting of the Acoustical Society of America, State College, PA.
- Guttman, L. (1944) A basis for scaling quantitative data. *American Sociological Review* 9, 139-50.
- Hancin-Bhatt, B. and Bhatt, R. (1997) Optimal L2 syllables. Interactions of transfer and developmental factors. *Studies in Second Language Acquisition* 19, 331-378.
- Hansen, J. (2006) *Acquiring a Non-Native Phonology*. London and New York: Continuum.
- He, D. and Li, D.C.S. (2009) Language attitudes and linguistic features in the 'China English' debate. *World Englishes* 28, 70-89.
- Hecht, B.F. and Mulford, R. (1982) The acquisition of a second language phonology: interaction of transfer and developmental factors. *Applied Linguistics* 3, 313-328.
- Heggarty, P., McMahon, A. and McMahon, R. (2005) From phonetic similarity to dialect classification: a principled approach. In Delbecque, N., van der Auwera, J. and Geeraerts, G. (eds) *Perspectives on Variation: Sociolinguistic, Historical, Comparative*. Berlin: Walter de Gruyter, 43-92.
- Held, D., McGrew, A., Goldblatt, D. and Perraton, J. (1999) *Global Transformations: Politics, Economics and Culture*. Cambridge: Polity Press.
- Higgins, C. (2003) "Ownership" of English in the outer circle: an alternative to the NS-NNS dichotomy. *TESOL Quarterly* 37/4, 615-644.
- Higgins, J. (2009) *Minimal pairs for English RP*. Available: <http://myweb.tiscali.co.uk/wordscape/wordlist/> Accessed 10 May, 2009.

- Ho, M.L. and Platt, J. (1993) *Dynamics of a Contact Continuum: Singaporean English*. Oxford: Clarendon Press.
- Holliday, A. (2005) *The Struggle to Teach English as an International Language*. Oxford: Oxford University Press.
- Hung, T.T.N. (2000) Towards a phonology of Hong Kong English. *World Englishes* 19, 337-356. Also in Bolton, K. (ed.) 2002 *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 119-40.
- Hung, T.T.N. (2002) English as a global language: implications for teaching. *ACELT Journal* 6/2, 3-10.
- Hunt, K.W. (1965) *Grammatical Structures Written at Three Grade Levels*. Urbana, IL: National Council of Teachers of English.
- Hyland, K. (1997) Language attitudes at the handover: communication and identity in 1997 Hong Kong. *English World-Wide* 18, 191-210.
- Hymes, D. (1974) *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- International Corpus of English (2010) Hong Kong English. Available online: <http://ice-corpora.net/ice/soundhk.htm>. Accessed April 25, 2010.
- IELTS (2010) IELTS Speaking band descriptors (public version). Available online: http://www.ielts.org/pdf/UOBDs_SpeakingFinal.pdf. Accessed April 25, 2010.
- Ioup, G. and Weinberger, S.H. (eds.) (1982) *Interlanguage Phonology*. Cambridge, MA: Newbury House.
- James, A. (2005) The challenges of the Lingua Franca: English in the world and types of variety. In Gnutzmann, C. and Intemann, F. (eds) *The Globalisation of English and the English Language Classroom*. Tübingen: Gunter Narr, 133-144.
- Jenkins, J. (1997) Changing priorities for successful communication in international contexts. *SIG Selections 1997 Special Interests in ELT (IATEFL)*, 73-79.
- Jenkins, J. (1998) Which pronunciation norms and models for English as an international language? *ELT Journal* 52/2, 119-126.
- Jenkins, J. (2000) *The Phonology of English as an International Language: New Models, New Norms, New Goals*. Oxford: Oxford University Press.
- Jenkins, J. (2005) Misinterpretation, Bias and Resistance to Change: the case of the Lingua Franca Core. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 199-210.
- Jenkins, J. (2006) Points of view and blind spots: ELF and SLA. *International Journal of Applied Linguistics* 16/2, 137-162.

Jenkins, J. (2007) *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.

Jenkins, J., Modiano, M. and Seidlhofer, B. (2001) Euro-English. *English Today* 17/4, 13-19.

Johnson, W. and Britain, D. (2007) L-vocalisation as a natural phenomenon: explorations in sociophonology. *Language Sciences* 29/2-3, 294-315.

Jones, M. and Singh, I. (2005) *Exploring Language Change*. London and New York: Routledge.

Joseph, J. (1996) English in Hong Kong: Emergence and Decline. *Current Issues in Language and Society* 3, 60-79. Also in Wright, S. and Kelly-Holmes, H. (eds) *One Country, Two Systems, Three Languages: Changing Language Use in Hong Kong*. Clevedon: Multilingual Matters, 60-79.

Joseph, J. (2004) *Language and Identity*. Basingstoke and New York: Palgrave Macmillan.

Joseph, J.E. and Taylor, T.J. (1990) (eds) *Ideologies of Language*. London: Routledge.

Kachru, B. (1976) Models of English for the Third World: White Man's linguistic burden or language pragmatics? *TESOL Quarterly* 10/2, 221-239.

Kachru, B. (1990) *The Alchemy of English: The Spread, Functions and Models of Non-Native Englishes*. Urbana and Chicago: University of Illinois Press.

Kachru, B. (1992a) Teaching World Englishes. In Kachru, B. (ed.) *The Other Tongue: English Across Cultures*. Urbana, Illinois: University of Illinois Press, 355-365.

Kachru, B. (1992b) Models for non-native Englishes. In Kachru, B. (ed.) *The Other Tongue: English Across Cultures*. Urbana, Illinois: University of Illinois Press, 48-74.

Kachru, B. (2005) *Asian Englishes: Beyond the Canon*. Hong Kong: University of Hong Kong Press.

Kachru, Y. (1994) Monolingual bias in SLA research. *TESOL Quarterly* 28/4, 795-800.

Kaplan, D. (2004) *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, CA, London and New Delhi: Sage Publications.

Kenworthy, J. (1987) *Teaching English Pronunciation*. Harlow: Longman.

Kerswill, P. (2002) Koineization and accommodation. In Chambers, J.K., Trudgill, P. and Schilling-Estes, N. (eds) *The Handbook of Language Variation and Change*. Oxford: Basil Blackwell, 669-702.

Kerswill, P. and Williams, A. (2002) 'Salience' as an explanatory factor in language change. In Jones, M. and Esch, E. (eds) *Language Change: the Interplay of Internal, External and Extra-linguistic Factors*. Berlin: Mouton de Gruyter, 81-110.

Kirkpatrick, A. (2000) English as an Asian Language: implications for the English language curriculum. In Brown, A. (ed.) *English in South East Asia*. Singapore: National Institute of Education, 58-70.

Kirkpatrick, A. (2002) ASEAN and Asian cultures and models: implications for the ELT curriculum and teacher selection. In Kirkpatrick, A. (ed.) *Englishes in Asia: Communication, Power, Identity and Education*. Perth: Curtin University of Technology, 213-224.

Kirkpatrick, A. (2004) English as an ASEAN lingua franca: implications for research and language teaching. *Asian Englishes* 6/2, 82-91.

Kirkpatrick, A. (2006) Which model of English: native speaker, nativized or lingua franca? In Rubdy, R. and Saraceni, M. (eds) *English in the World: Global Rules, Global Roles*. London and New York: Continuum, 71-83.

Kirkpatrick, A. (2007a) *World Englishes: Implications for International Communication and English Language Teaching*. Cambridge: Cambridge University Press.

Kirkpatrick, A. (2007b) Setting attainable and appropriate English targets in multilingual settings: a case for Hong Kong. *International Journal of Applied Linguistics* 17/3, 376 - 391.

Kirkpatrick, A. and Saunders, N. (2005) The intelligibility of Singapore English: a case study in an Australian university. In Deterding, D., Brown, A. and Low, E.L. (eds) *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill, 153-162.

Kirkpatrick, A., Deterding, D. and Wong, J. (2008) The international intelligibility of Hong Kong English. *World Englishes* 27 (3-4), 359-377.

Kortmann, B. (2004) Introduction. In Kortmann, B. (ed.) *Dialectology meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin and New York: Mouton de Gruyter, 1-10.

Koster, C.J. (1987) *Word Recognition in Foreign and Native Language*. Dordrecht: Foris.

Kuo, I.-C. (2006) Addressing the issue of teaching English as a lingua franca. *ELT Journal* 60/3, 213-221.

- Labov, W. (1966) *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1972) Where do grammars stop? In Shuy, R. (ed.) *Georgetown Monograph on Languages and Linguistics*. Georgetown: Georgetown University Press, 43-88.
- Labov, W. (1973) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1994) *Principles of Linguistic Change*, vol.1: *Internal Factors*. Oxford and Cambridge, MA: Blackwell.
- Labov, W. (2001) *Principles of Linguistic Change*, vol.2: *Social Factors*. Oxford and Cambridge, MA: Blackwell.
- Labov, W., Ash, S. and Boberg, C. (2006) *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter.
- Lambert, W. (1967) A social psychology of bilingualism. In Macnamara, J. (ed.) *Problems of Bilingualism. Journal of Social Issues* 23, 91-109.
- Lass, R. (1984) *Phonology*. Cambridge: Cambridge University Press.
- Laver, R. (1994) *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lee, I. (2005) English Language teaching in Hong Kong Special Administrative Region (HKSAR): a continuous challenge. In Braine, G. (2005) (ed.) *Teaching English to the World: History, Curriculum, and Practice*. Mahwah, NJ: Lawrence Erlbaum, 35-46.
- Leech, G.N., Rayson, P. and Wilson, A. (2001) *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. Harlow: Longman.
- Leitner, G. (1992) English as a pluricentric language. In Clyne, M. (ed.) *Pluricentric Languages: different norms in different nations*. Berlin and New York: Mouton de Gruyter, 179-237.
- LePage, R. (1980) Theoretical aspects of sociolinguistic studies in pidgin and creole languages. In Valdman, A. and Highfield, A. (eds) *Theoretical Orientations in Pidgin and Creole Studies*. New York: Academic Press, 331-397.
- LePage, R. and Tabouret-Keller, A. (1985) *Acts of Identity*. Cambridge and New York: Cambridge University Press.
- Levis, J. M. (2006) Pronunciation and the assessment of spoken language. In Hughes, R. (ed.) *Spoken English, applied linguistics, and TESOL: Challenges for theory and practice*, pp. 245-270. Basingstoke, UK: Palgrave Macmillan, 245-270.

- Li, D.C.S. (2000) 'Hong Kong English': new variety of English or interlanguage? *English Australia Journal* 18/1, 50-59.
- Li, D.C.S. (2002) Hong Kong parents' preference for English medium education: Passive victims of imperialism or active agents of pragmatism? In Kirkpatrick, A. (ed.) *Englishes in Asia: Communication, Power, Identity and Education*. Perth: Curtin University of Technology, 29-62.
- Lindblom, B. (1986) Phonetic universals in vowel systems. In Ohala, J. and Jaeger, J. (eds) *Experimental Phonology*. Dordrecht: Foris, 13-44.
- Lippi-Green, R. (1997) *English with an Accent: Language, Ideology, and Discrimination in the United States*. London and New York: Routledge.
- Llurda, E. (2000) Effects of intelligibility and speaking rate on judgements of non-native speakers' personalities. *International Review of Applied Linguistics in Language Teaching* 38 (3-4), 289-300.
- Low, E.L. (1998) *Prosodic Prominence in Singapore English*. PhD thesis, University of Cambridge.
- Low, E.L., Grabe, E. and Nolan, F. (2000) Quantitative characterisations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43/4, 377-401.
- Luk, J. (1998). Hong Kong students' awareness of and reactions to accent differences. *Multilingua* 17, 93-106.
- Luk, J. and Lin, A. (2006) Uncovering the sociopolitical situatedness of accents in the World Englishes paradigm. In Hughes, R. (ed.) *Spoken English, TESOL and Applied Linguistics*. Basingstoke: Palgrave Macmillan, 3-22.
- Luke, K.K. and Richards, J.C. (1982) English in Hong Kong: functions and status. *English World-Wide* 3, 47-63.
- Lutz, A. (1991) *Phonotaktisch gesteuerte Konsonantenveränderungen in der Geschichte des Englischen*. Tübingen: Niemeyer.
- Magnan, S. and Walz, J. (2002) Pedagogical norms: development of the concept and illustrations from French. In Gass, S., Bardovi-Harlig, K., Magnan, S.S. and Walz, J. (eds) *Pedagogical Norms for Second and Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, 15-40.
- Major, R. (1987) A model for interlanguage phonology. In In Ioup, G. and Weinberger, S.H. (eds) *Interlanguage Phonology*. Cambridge, MA: Newbury House, 101-124.
- Major, R. (2001) *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum.

- Major, R., Fitzmaurice, S., Bunta, F. and Balasubramanian, C. (2002) The effects of nonnative accents on listening comprehension: implications for ESL assessment. *TESOL Quarterly* 36, 173-190.
- Markham, D. (1997) *Phonetic Imitation, Accent and the Learner*. Lund: Lund University Press.
- Marks, J. (1999) Is stress-timing real? *ELT Journal* 53/3, 191-199.
- Martinet, A. (1961) *Elements de Linguistique Generale*. Paris: P. Colin.
- Martínez-Gil, F. and Colina, S. (2006) (eds) *Optimality-Theoretic Studies in Spanish Phonology*. Amsterdam and Philadelphia: John Benjamins.
- McCarthy, J. and Prince, A. (2004) The emergence of the unmarked. In McCarthy, J. and Prince, A. (eds) *Optimality Theory in Phonology: a Reader*. Malden, MA and Oxford: Blackwell, 483-494.
- McGraw, K. O. and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1(1), 30-46.
- McKay, S. (2002) *Teaching English as an International Language*. Oxford: Oxford University Press.
- McKay, S. (2006) *Researching Second Language Classrooms*. Mahwah, NJ: Lawrence Erlbaum.
- McMahon, A. (1994) *Understanding Language Change*. Cambridge: Cambridge University Press.
- Mesthrie, R. and Bhatt, R. (2008) *World Englishes: the Study of New Linguistic Varieties*. Cambridge: Cambridge University Press.
- Milroy, J. (2001) Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 5/4, 530-555.
- Moag, R.F. and Moag, L.B. (1977) English in Fiji: Some perspective and the need for language planning. *Fiji English Teachers' Journal* 13, 2-26.
- Mohanan, K.P. (1992) Describing the phonology of non-native varieties of a language. *World Englishes* 11, 111-128.
- Mollin, S. (2006) *Euro-English: assessing variety status*. Tübingen: Gunter Narr Verlag.
- Moyer, A. (2004) *Age, Accent and Experience in Second Language Acquisition*. Clevedon: Multilingual Matters.
- Mufwene, S. (2001) *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

- Munro, M. (2008) Foreign accent and speech intelligibility. In Hansen-Edwards, J. and Zampini, M. (eds) *Phonology and Second Language Acquisition*. Amsterdam and Philadelphia: John Benjamins, 251-82.
- Munro, M. and Derwing, T. (1998) The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48, 159-182.
- Munro, M. and Derwing, T. (1999) Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 49, 285-310.
- Munro, M. and Derwing, T. (2006) The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34/4, 520-531.
- Munro, M., Derwing, T. and Morton, S. (2006) The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28, 111-131.
- Murakami, H. (2008) *What I Talk About When I Talk About Running: A Memoir*. New York: Alfred A. Knopf.
- Nelson, C. (1995) Intelligibility and world Englishes in the classroom. *World Englishes* 14/2, 273-279.
- Nettle, D. (1999) Functionalism and its difficulties in biology and linguistics. In Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M. and Wheatley, K. (eds) *Functionalism and Formalism in Linguistics* vol.1: *General Papers*. Amsterdam and Philadelphia: John Benjamins, 445-467.
- Nunn, R. (2005) Competence and teaching English as an international language. *Asian EFL Journal* 7/4, 61-74.
- Nuttall, M. (2000) English in Japan. *Japan Railway and Transport Review* 25, 56-57.
- Otago (2002) Evaluating English Accents WorldWide: The Impact of Intonation. University of Otago. Available online:
<http://www.otago.ac.nz/anthropology/Linguistic/Intonation/Intonation.html>.
 Accessed April 27, 2010.
- Pakir, A. (1991) The range and depth of English-knowing bilinguals in Singapore. *World Englishes* 10, 167-79.
- Pallant, J. (2001) *SPSS Survival Manual: a Step-by-step Guide to Data Analysis Using SPSS for Windows (Version 10)*. Buckingham and Philadelphia: Open University Press.
- Pavlenko, A. and Blackledge, A. (2004) *Negotiation of Identities in Multilingual Contexts*. Clevedon: Multilingual Matters.

- Peng, L. and Setter, J. (2000) The emergence of systematicity in the English pronunciations of two Cantonese-speaking adults in Hong Kong. *English Worldwide* 21/1, 81-108.
- Pennington, M. (1998) Perspectives on language in Hong Kong at century's end. In Pennington, M. (ed.) *Language in Hong Kong at Century's End*. Hong Kong: Hong Kong University Press.
- Pennington, M. (1998) (ed.) *Language in Hong Kong at Century's End*. Hong Kong: Hong Kong University Press.
- Perry, F. (2005) *Research in Applied Linguistics: Becoming a Discerning Consumer*. Mahwah, NJ: Lawrence Erlbaum.
- Phillipson, R. (1992) *Linguistic Imperialism*. Oxford: Oxford University Press.
- Phillipson, R. (2002) Global English and local language policies. In Kirkpatrick, A. (ed) *Englishes in Asia: Communication, Power, Identity and Education*. Perth: Curtin University of Technology, 7-28.
- Phillipson, R. (2008) Lingua franca or lingua frankensteinia? English in European integration and globalisation. *World Englishes* 27/2, 250-284.
- Pickering, L. Current research on intelligibility in English as a Lingua Franca. *Annual Review of Applied Linguistics* 26, 219-233.
- Pienemann, M. (1998) *Language Processing and Second Language Development: Processability Theory*. Amsterdam and Philadelphia: John Benjamins.
- Prator, C. (1968) The British heresy in TESL. In Fishman, J., Ferguson, C.A. and DasGupta, J. (eds) *Language Problems in Developing Nations*. New York: Wiley, 459-476.
- Preisler, B. (1999) English in a European EFL country. In Bex, T. and Watts, R.J. (eds) *Standard English: the Widening Debate*. London: Routledge, 239-267.
- Preston, D.R. (2005) How can you learn a language that isn't there? In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 37-58.
- Prince, A. and Smollensky, P. (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. Report No. RuCCS-TR-2. New Brunswick, NJ: Rutgers University Center for Cognitive Science.
- Prodromou, L. (2008) *English as a Lingua Franca: a Corpus-Based Analysis*. London: Continuum.
- Quirk, R. (1990) Language varieties and standard language. *English Today* 21, 3-10.

Radwanska-Williams, J. (2008) The “native speaker” as a metaphorical construct. In Berendt, E. (ed.) *Metaphors for learning: cross-cultural perspectives*. Amsterdam and Philadelphia: John Benjamins, 139-158.

Radwanska-Williams, J. and Yam, J. (2001) The Acquisition of English Plosives by Chinese Learners. *Proceedings of the Phonetics Teaching and Learning Conference* vol. 2001, 32-35.

Rajadurai, J. (2007) Intelligibility studies: a consideration of empirical and ideological issues. *World Englishes* 26/1, 87-98.

Randall, M. (2007) *Memory, Psychology and Second Language Learning*. Amsterdam and Philadelphia: John Benjamins.

Rickford, J. (2002) Implicational scales. In Chambers, J.K., Trudgill, P. and Schilling-Estes, N. (eds) *The Handbook of Language Variation and Change*. Oxford: Basil Blackwell, 142-67.

Ringbom, H. (2007) *Cross-linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters.

Rischel, J. (1962) On functional load in phonemics. *Statistical Methods in Linguistics* 1, 13-23.

Roach, P. (1982) On the distinction between 'stress-timed' and 'syllable-timed' languages. In Crystal, D. (ed.) *Linguistic Controversies*. London: Arnold, 73-79.

Roach, P. (2000) *English Phonetics and Phonology* (3rd edn) Cambridge: Cambridge University Press.

Roach, P., Hartman, J. and Setter, J. (eds) (2003) *Cambridge English Pronouncing Dictionary* (16th edn) Cambridge: Cambridge University Press.

Robinson, D. (2003) *Performative Linguistics: Speaking and Translating as Doing Things With Words*. New York and London: Routledge.

Rosewarne, D. (1994). Estuary English: Tomorrow's RP? *English Today* 10/1, 3-8.

Rubin, D.L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33/4, 511-531.

Sakai, S. and D'Angelo, J.F. (2005) A vision for world Englishes in the Expanding Circle. *World Englishes* 24/3, 323-327.

Sampson, G. (1980) *Schools of Linguistics: Competition and Evolution*. London: Hutchinson.

Sapir, E. (1921) *Language: an Introduction to the Study of Speech*. New York: Harcourt, Brace.

Savin, N. and White, K. (1977) The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica* 45, 1989-1996.

Scheuer, S. (2005) Why native speakers are (still) relevant. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 111-130.

Schneider, E. (2003) The dynamics of new Englishes: from identity construction to dialect birth. *Language* 79/2, 233-81.

Schneider, E. (2004) Global synopsis: phonetic and phonological variation in English world-wide. In Schneider, E., Burridge, B., Kortmann, B., Mesthrie, R. and Upton, C. (eds) *A Handbook of Varieties of English*. Vol. 1: *Phonology*. Berlin: Mouton de Gruyter, 1111-1137.

Schneider, E. (2007) *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.

Schreier, D. (2005) *Consonant Change in English Worldwide*. Basingstoke: Palgrave Macmillan.

Schreier, D. (2008) *St Helenian English*. Amsterdam and Philadelphia: John Benjamins.

Schreier, D. (2009) How diagnostic are English universals? In Filppula, M., Klemola, J. and Paulasto, H. (eds) *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. New York and London: Routledge, 57-79.

SCOLAR (2003) *Action Plan to Raise Language Standards in Hong Kong*. Hong Kong: Government Printing Department.

Seidlhofer, B. (2002) *Habeas corpus and divide at impera*: 'Global English' and applied linguistics. In Spelman Miller, K. and Thompson, P. (eds) *Unity and Diversity in Language Use*. London: Continuum, 198-217.

Seidlhofer, B. (2004) Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics* 24, 209-239.

Seidlhofer, B. (2005) Language variation and change: the case of English as a lingua franca. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 59-75.

Seidlhofer, B. (2007) Comment on Rajadurai (2007). *World Englishes* 26/1, 99-100.

Seidlhofer, B. (2010) English as a Lingua Franca. University of Vienna, Vienna-Oxford International Corpus of English. Available online: http://www.univie.ac.at/voice/page/documents/oald_elf_page_r92.pdf Accessed 25 April, 2010.

- Seidlhofer, B. and Jenkins, J. (2003) English as a Lingua Franca and the politics of property. In Mair, C. (ed.) *The Politics of English as a World Language: New Horizons in Postcolonial Cultural Studies*. Amsterdam and New York: Rodopi, 139-156.
- Seidlhofer, B., Breiteneder, A. and Pitzl, M.-L. (2006) English as a Lingua Franca in Europe: challenges for applied linguistics. *Annual Review of Applied Linguistics* 26, 3-34.
- Setter, J. and Jenkins, J. (2005) State of the art review article: Pronunciation. *Language Teaching* 38, 1-17.
- Setter, J. (2008) Consonant clusters in Hong Kong English. *World Englishes* 27/3-4, 502-515.
- Sewell, A. (2009) World Englishes, English as a Lingua Franca and the case of Hong Kong English. *English Today* 25, 37-43.
- Sewell, A. (2010) Research methods and intelligibility studies. *World Englishes* 29/2, 257-269.
- Sewell, A. and Chan, J. (2010) Patterns of variation in the consonantal phonology of Hong Kong English. *English World-Wide* 31/2, 138-161.
- Shockey, L. (2003) *Sound Patterns of Spoken English*. Oxford: Blackwell.
- Smith, L.E. (1992) Spread of English and issues of intelligibility. In Kachru, B. *The Other Tongue: English Across Cultures*. Urbana, Illinois: University of Illinois Press.
- Smith, L. E. and Rafiqzad, K. (1979) English for cross-cultural communication: the question of intelligibility. *TESOL Quarterly* 13/3, 371-380.
- Smith, L.E. and Nelson, C.L. (1985) International intelligibility of English: directions and resources. *World Englishes* 3, 333-342.
- South China Morning Post (2007) Speak and span. *South China Morning Post*, 15 March.
- South China Morning Post (2008) Language policy change to 'alter public mindset'. *South China Morning Post*, 1 March.
- South China Morning Post (2009a) Marketing Guide 2009. Available at: http://advertising.scmp.com/ratecard/pdf/2009_saleskit_hongkong.pdf. Accessed 10 October 2009.
- South China Morning Post (2009b) BaptistU to double time spent on compulsory English learning. *South China Morning Post*, 28 February.
- South China Morning Post (2009c) Nearly half of high schools to teach in English. *South China Morning Post*, 15 December.

- Stibbard, R. (2004) The spoken English of Hong Kong: a study of co-occurring segmental errors. *Language, Culture and Curriculum* 17, 127-42.
- Stevens, P. (1992) English as an international language: directions in the 1990s. In Kachru, B. *The Other Tongue: English Across Cultures*. Urbana, Illinois: University of Illinois Press, 27-47.
- Swan, M. (2005) Legislation by hypothesis: the case of task-based instruction. *Applied Linguistics* 26/3, 376-401.
- Swan, M. and Smith, B. (2001) *Learner English: a Teacher's Guide to Interference and Other Problems*. Cambridge: Cambridge University Press.
- Szpyra-Kozłowska, J. (2005) Lingua Franca Core, phonetic universals and the Polish context. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 151-176.
- Tabachnick, B. and Fidell, L. (1996) *Using Multivariate Statistics*. Boston, MA: Allyn and Bacon.
- Tagliamonte, S. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tajima, K., Port, R. and Dalby, J. (1997) Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25/1, 1-24.
- Talbot, M., Atkinson, K. and Atkinson, D. (2003) *Language and Power in the Modern World*. Edinburgh: Edinburgh University Press.
- Tay, M. (1973) Problems in teaching pronunciation. *Regional English Language Centre Journal* 4/1, 1-8.
- Temperley, M. (1983) The articulatory target for final –s clusters. *TESOL Quarterly* 17, 421-436.
- Thomason, S. (2009) Why universals VERSUS contact-induced change? In Filppula, M., Klemola, J. and Paulasto, H. (eds) *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. New York and London: Routledge, 349-364.
- Thomason, S. and Kaufman, T. (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Tong, S. and James, G. (1994) *Colloquial Cantonese*. London: Routledge.
- Trudgill, P. (1974) *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, P. (1984) *Language in the British Isles*. Cambridge: Cambridge University Press.

- Trudgill, P. (1986) *Dialects in Contact*. Oxford: Blackwell.
- Trudgill, P. (1999) Standard English: what it isn't. In Bex, T. and Watts, R.J. (eds) (1999) *Standard English: the Widening Debate*. London: Routledge, 117-128.
- Trudgill, P. (2005) Finding the speaker-listener equilibrium: segmental phonological models in EFL. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 213-228.
- Trudgill, P. and Hannah, J. (1982) *International English: a Guide to Varieties of Standard English*. London: Edward Arnold.
- Tsui, A.M and Bunton, D. (2002) The discourse and attitudes of English language teachers in Hong Kong. In Bolton, K. (ed.) *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 57-78.
- Ur, P. (2006) A different ball game. Paper presented at the 40th Annual IATEFL Conference, Harrogate, England.
- Valdman, A. (1989) Classroom foreign language learning and language variation: the notion of pedagogical norms. In Eisenstein, M. (ed.) *The Dynamic Interlanguage: Empirical Studies in Second Language Variation*. New York: Plenum, 261-277.
- van den Doel, R. (2006) *How Friendly are the Natives? An Evaluation of Native-Speaker Judgements of Foreign-accented British and American English*. PhD dissertation. Utrecht: LOT.
- van den Doel, R. (2007) International Intelligibility in EIL. *Asian EFL Journal* 9/4, 28-38. Available: http://www.asian-efljournal.com/December_2007_EBook.pdf#page=28 Accessed 9 May, 2008.
- Van der Walt, C. (2000) The international comprehensibility of varieties of South African English. *World Englishes* 19, 139-53.
- Wardhaugh, R. (2009) *An Introduction to Sociolinguistics*. Oxford: Blackwell.
- Waters, A. (2007) ELT and 'the spirit of the times'. *ELT Journal* 61/4, 353-359.
- Weber, A. and Cutler, A. (2004) Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50, 1-15.
- Wee, L. (2002) *Lor* in colloquial Singapore English. *Journal of Pragmatics* 34/6, 711-725.
- Weiner, D., Freedheim, J. Schinka, W. and Velicer, W. (2003) *Handbook of Psychology: Research methods in Psychology*. Hoboken, NJ: Wiley.
- Wells, J.C. (1982) *The Accents of English*. Cambridge: Cambridge University Press.
- Widdowson, H. G. (1994) The ownership of English. *TESOL Quarterly* 28, 377-389.

- Widdowson, H.G. (2004) A perspective on recent trends. In Howatt, A.P.R. and Widdowson, H.G. *A History of English Language Teaching*. Oxford: Oxford University Press, 353-372.
- Williams, J. (1987) Non-native varieties of English: A special case of language acquisition. *English World-Wide* 8, 161-199.
- Wilson, R. and Keil, F. (2001) *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.
- Wolfe-Quintero, K., Inagaki, S. and Kim, H.-Y. (1998) *Second Language Development in Writing: Measures of Accuracy, Complexity and Fluency*. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
- Wolff, H. (1959) Intelligibility and inter-ethnic attitudes. *Anthropological Linguistics* 1/3, 34-41.
- Wrembel, M. (2005) An overview of English pronunciation teaching materials. Patterns of change: model accents, goals and priorities. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (eds) *English Pronunciation Models: A Changing Scene*. Bern, Switzerland: Peter Lang, 421-438.
- Xu, Z. (2002) From TEFL to TEIL: changes in perceptions and practices: teaching English as an International Language (EIL) in Chinese universities in P.R. China. In Kirkpatrick, A. (ed) *Englishes in Asia: Communication, Power, Identity and Education*. Perth: Curtin University of Technology, 225-244.
- Young, C. (2003) 'You dig tree tree to NUS': understanding Singapore English from the perspectives of international students. In Lee, Gek Ling, Ho, L., Gek, Ling Lee, Meyer, J.E.L., Varaprasad, C. and Young, C. (eds) *Teaching English to Students from China*. Singapore: Singapore University Press, 94-106.

Appendix 1: details of speakers and speech samples in the Hong Kong English mini-corpus.

Speaker No.	Gender M/F	Occupation (if known)	Programme and date of broadcast	Setting	Total length of extracts (seconds)
1	M	Politician	The Pearl Report, March 2006	Location interview	36
2	M	Politician	"	Location interview	40
3	M	Industry spokesperson	"	Location interview	21
4	M	Politician	"	Location interview	29
5	M	Politician	"	Studio interview	24
6	F	NGO spokesperson	"	Location interview	73
7	M	Civil servant	"	Location interview	24
8	M	Politician	"	Location interview	51
9	M	Company spokesperson	The Pulse, April 2007	Location interview	55
10	M	Unknown	"	Location interview	30
11	M	Unknown	"	Location interview	32
12	M	Politician	"	Studio interview	131
13	M	Educational administrator	The Pulse, May 2007	Studio discussion	130
14	M	Politician	"	Studio discussion	85
15	M	Politician	"	Studio discussion	171
16	M	Politician	The Pulse, June 2007	Public address	41
17	F	NGO spokesperson	"	Location interview	30
18	F	Civil servant	"	Location interview	16
19	M	Unknown	"	Studio discussion	116
20	F	Politician	The Pulse, December 2007	Studio discussion	110
21	M	Journalist	"	Studio discussion	63
22	M	Journalist	"	Studio discussion	229
23	M	Politician	The Pulse, February 2008	Public address	23
24	M	Politician	"	Studio discussion	174
25	M	NGO spokesperson	"	Studio discussion	122

Appendix 2a: Part 1 of the survey form used in the study (NB: Part 1 was completed first for all speakers, followed by Part 2).

Part 1: listen to each speaker and put a tick in one box for each of the questions a-f.

Speaker 1

		Agree strongly						Disagree strongly
a	The speaker sounds like a Hong Kong person.	☐	☐	☐	☐	☐	☐	
b	This speaker has a lot of pronunciation errors.	☐	☐	☐	☐	☐	☐	
c	This speaker is easy to understand.	☐	☐	☐	☐	☐	☐	
d	I like the way this speaker sounds.	☐	☐	☐	☐	☐	☐	
e	This speaker's accent is acceptable as a model for pronunciation teaching purposes in HK.	☐	☐	☐	☐	☐	☐	
f	This speaker has a high level of education and / or a high status job.	☐	☐	☐	☐	☐	☐	

Appendix 2b: Part 2 of the survey form used in the study.

Part 2: listen again and then try to decide which words, sounds or other features were most important in helping you make the decisions you made in Part 1. You can refer to any of these areas:

Vowel sounds (V) Consonant sounds (C) Consonant clusters (CC)
Word stress (WS) Connected speech: sentence stress, rhythm, linking etc. (CS) Intonation (I)

For ‘negative’ features, please mark the transcript by underlining the relevant parts and using the above codes. For example, if you think there is a consonant problem in the word ‘supermarket’ you can mark it like this:

supermarket
C

NB: Please do not mark more than THREE features per speaker. Decide which features were most important in forming your impression.

If you do not think there are any errors, you do not need to mark anything. You may note ‘positive’ features and/or further explain your Part 1 answers in the space provided.

Speaker 1

They don’t see an advantage in doing anything risky, and they don’t have to because they think that they have all the cards now

Any other comments about this speaker:

Correlations

		ERR PW	ERR PSYLL	NO ERR	WPM	SYLL PM	SYLL PW	LEX COMP	WTU	CTU	MIN FREQ	MAX FREQ	SPAN	LENGTH
ERRPW	Pearson	1	.984(**)	.885(**)	-.357	-.539	.014	-.341	-.359	.077	-.291	.158	.284	-.606
	Sig. (2-tailed)	.	.000	.004	.385	.168	.974	.408	.382	.857	.484	.710	.495	.111
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
ERRPSYLL	Pearson	.984(**)	1	.888(**)	-.209	-.455	-.153	-.411	-.306	.178	-.345	.120	.278	-.692
	Sig. (2-tailed)	.000	.	.003	.619	.257	.718	.311	.460	.673	.403	.777	.505	.057
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
NOERR	Pearson	.885(**)	.888(**)	1	-.386	-.608	.009	-.164	-.272	.219	-.221	.459	.518	-.382
	Sig. (2-tailed)	.004	.003	.	.345	.110	.984	.699	.515	.602	.599	.252	.189	.351
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
WPM	Pearson	-.357	-.209	-.386	1	.863(**)	-.830(*)	-.346	.374	.332	-.453	-.549	-.263	-.446
	Sig. (2-tailed)	.385	.619	.345	.	.006	.011	.401	.362	.422	.260	.159	.529	.268
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
SYLLPM	Pearson	-.539	-.455	-.608	.863(**)	1	-.441	-.060	.256	-.057	-.404	-.673	-.398	-.181
	Sig. (2-tailed)	.168	.257	.110	.006	.	.274	.887	.541	.894	.321	.067	.329	.667
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
SYLLPW	Pearson	.014	-.153	.009	-.830(*)	-.441	1	.566	-.422	-.690	.421	.229	-.006	.621
	Sig. (2-tailed)	.974	.718	.984	.011	.274	.	.144	.298	.058	.298	.586	.989	.100
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
LEXCOMP	Pearson	-.341	-.411	-.164	-.346	-.060	.566	1	.082	-.237	.636	-.003	-.318	.672
	Sig. (2-tailed)	.408	.311	.699	.401	.887	.144	.	.847	.573	.090	.995	.443	.068
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
WTU	Pearson	-.359	-.306	-.272	.374	.256	-.422	.082	1	.765(*)	-.109	-.042	.017	-.101
	Sig. (2-tailed)	.382	.460	.515	.362	.541	.298	.847	.	.027	.796	.921	.968	.812
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
CTU	Pearson	.077	.178	.219	.332	-.057	-.690	-.237	.765(*)	1	-.218	.264	.342	-.342
	Sig. (2-tailed)	.857	.673	.602	.422	.894	.058	.573	.027	.	.605	.528	.406	.408
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
MINFREQ	Pearson	-.291	-.345	-.221	-.453	-.404	.421	.636	-.109	-.218	1	.040	-.461	.607
	Sig. (2-tailed)	.484	.403	.599	.260	.321	.298	.090	.796	.605	.	.925	.250	.111
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
MAXFREQ	Pearson	.158	.120	.459	-.549	-.673	.229	-.003	-.042	.264	.040	1	.868(**)	.417
	Sig. (2-tailed)	.710	.777	.252	.159	.067	.586	.995	.921	.528	.925	.	.005	.304
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
SPAN	Pearson	.284	.278	.518	-.263	-.398	-.006	-.318	.017	.342	-.461	.868(**)	1	.069
	Sig. (2-tailed)	.495	.505	.189	.529	.329	.989	.443	.968	.406	.250	.005	.	.870
	N	8	8	8	8	8	8	8	8	8	8	8	8	8
LENGTH	Pearson	-.606	-.692	-.382	-.446	-.181	.621	.672	-.101	-.342	.607	.417	.069	1
	Sig. (2-tailed)	.111	.057	.351	.268	.667	.100	.068	.812	.408	.111	.304	.870	.
	N	8	8	8	8	8	8	8	8	8	8	8	8	8

** is significant at the 0.01 level (2-tailed).

* is significant at the 0.05 level (2-tailed).

Appendix 3: correlation matrix for the thirteen speaker variables.