Perception and Production of Cantonese Tones by Speakers with Different Linguistic Experiences

Mengyue Wu

A thesis submitted for the degree of

Doctor of Philosophy

Linguistics and Applied Linguistics

The University of Melbourne

November 2017
Declaration of Originality

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed: [Signature] On: 09/11/2017
Abstract

This thesis investigates the perception and production of Cantonese tones by speakers who differ systematically in their native prosodic systems and language learning experiences. These include native tone language speakers of Cantonese and Mandarin, English speakers with no experience of tone languages, and English speakers who have experience with tone languages through learning Mandarin. The core of the thesis consists of two perception studies (tone categorisation, presented in Chapter 5, and discrimination, presented in Chapter 6) as well as a production study (presented in Chapter 7). The categorisation study relies on a novel approach in which English-speaking participants categorise Cantonese tones in terms of their native intonation system, while Mandarin speakers categorise Cantonese tones in terms of their native tone system. The results are interpreted and discussed within the framework of the perceptual assimilation models (PAM, PAM-L2, PAM-S; see Best, 1995; Best & Tyler, 2007; So & Best, 2014).

The production study (Chapter 7) includes an imitation task and detailed analyses of F0 onset and offset ellipses plots for each speaker group. I focus on 1) the degree of overlap between the non-native speaker-produced tones and, 2) how much space each tone takes against the whole tonal space. I further analyse tone trajectories at 10% tone intervals, which is crucial to tones in languages like Cantonese, where contour tones take up a large proportion. Additional native perceptual judgement was provided by two native Hong Kong Cantonese speakers who have a linguistics major. The production results for each participant group are further compared to see whether L2 tone learning experience impacts non-native speech perception and production in the same way as the native prosodic system does. The results from the English monolingual participants, as well as the Mandarin speakers indicate that their
non-native tone ability is influenced by their native systems: English monolinguals pay more attention to pitch height while Mandarin speakers attend more to pitch contour information. The most striking result is the fact that the Mandarin learners outperform both the native Mandarin speakers and the English monolinguals in perceiving and producing the complex Cantonese tones, suggesting that learning Mandarin familiarises English speakers with the use of lexical pitch information and tunes their attention to pitch contour.

Finally, the perception and production studies allow a careful discussion of the link between perception and production, as well as differences in individual performance. Non-native perception and production abilities are positively linked for speakers with tone experience in either a first or a second language, while English monolinguals show no correlation between the perception and production of Cantonese tones.
Acknowledgements

This PhD was begun in 2013 and it owes much to my principal supervisor, Dr Brett Baker. Without his unreserved support and enthusiasm in this project, I would never have reached this point. Brett has given me the freedom to explore my own research interests as well as some key suggestions from which I benefitted enormously: I read broadly, obtained statistical training, delivered many public talks, and practiced academic writing in both English and Chinese. The training he provided concerns more than my PhD project—it has urged me to consider what properties make an academic outstanding.

I am truly grateful for Professor Janet Fletcher—her lectures in experimental phonetics were the most influential, practical, and challenging ones I took in the last five years. At the moment when I was at last able to work with EMU/R, I started to feel the true glamour of those spectrograms. In the numerous times when I was ashamed of my ignorance and entirely disappointed with myself, Janet told me that knowledge is not built in one day and learning is progressive. Her knowledge and wittiness always made me realise that I have a long way to go.

My deepest gratitude also goes to Dr Rikke Bundgaard-Nielsen, who mentored me throughout my candidature, especially at the beginning and towards the end. Her expertise in speech perception directed me through the foundation of the whole study; her confidence in me guided me through the darkest times. After my confirmation in 2013, she suggested that I visit the MARCS Institute for one year. This visit became the most unforgettable experience of my PhD. From a band of researchers who have been working in the same field as me, I learned how to design my experiments with E-Prime, how to perform behavioural experiments, how to juggle between recruiting a large number of participants, conducting experiments,
recording, and analysing data. I also expanded my horizons through the various research groups at MARCS and got to know how an institute functioned differently from departments in universities. Most importantly, I received guidance from Professor Catherine Best during this year, who helped me tremendously in fine-tuning the details of the experimental design and understanding how I could interpret her model, the perceptual assimilation model, and extend it to test prosodic features. Most of all, without Rikke’s suggestion and connections at MARCS, none of these outcomes would have been realised.

Thanks to all my friends in the Phonetics Lab: Rosey Billington, Eleanor Lewis, Katie Jepson and Josh Clothier for the ongoing support when I had ‘culture shock’, when I had questions with R, when I had trouble sleeping at night in the latter half of my candidature. We shared the stress of this journey and the joy of having exciting findings. Our cosy lab during those windy and freezing winter afternoons will be among my most cherished memories.

I also would like to thank all the friends I made outside linguistics in Melbourne—they gave me the release of being able to talk about everything other than research over those guilty brunch dates and late dinners. My final thanks should go to my mother and my partner, who understand and support me unconditionally.
# Contents

Abstract .......................................................................................................................... iii

Acknowledgements ......................................................................................................... v

Contents ............................................................................................................................ vi

List of Figures .................................................................................................................... x

List of Tables ...................................................................................................................... xi

List of Abbreviations ......................................................................................................... xiv

Chapter 1: Introduction ..................................................................................................... 1

1.1 Background ................................................................................................................ 1

1.2 Motivation and Aims ............................................................................................... 3

1.3 Thesis Structure ...................................................................................................... 6

Chapter 2: Tone and Intonation ....................................................................................... 8

2.1 Prosody ...................................................................................................................... 8

2.1.1 Prosodic typology ............................................................................................... 8

2.1.2 Autosegmental Metrical and Tones and Break Indices transcriptions ............. 11

2.1.3 Transfer of native prosodic systems ................................................................ 12

2.2 Tone Languages ....................................................................................................... 13

2.2.1 Cantonese tone system ...................................................................................... 16

2.2.2 Mandarin tone system ....................................................................................... 21

2.3 Intonation Languages ............................................................................................. 25

2.3.1 English intonation ............................................................................................. 25

2.3.2 Australian English ............................................................................................ 27

2.4 Comparison between Lexical Tones and Intonation ............................................. 31

2.5 Summary .................................................................................................................. 35

Chapter 3: Tone Perception and Production ................................................................. 36

3.1 Tone Perception ....................................................................................................... 37

3.1.1 Native tone perception ....................................................................................... 38

3.1.2 Non-native tone perception .............................................................................. 39

3.1.2.1 Non-native tone perception by speakers of other tone languages .............. 40

3.1.2.2 Non-native tone perception by speakers of non-tone languages ............... 45

3.2 Tone Production ...................................................................................................... 47

3.2.1 Native tone production ...................................................................................... 47

3.2.2 Non-native tone production .............................................................................. 49

3.2.2.1 Non-native tone production by speakers of other tone languages .......... 49

3.2.2.2 Non-native tone production by speakers of non-tone languages ............. 50

3.3 The Link between Tone Perception and Production .............................................. 51

3.4 Summary .................................................................................................................. 54

Chapter 4: Theoretical Models and Thesis Overview .................................................... 55

4.1 The Perceptual Assimilation Model ........................................................................ 57

4.1.1 Review of the perceptual assimilation model .................................................. 57

4.1.2 Extending the perceptual assimilation model and perceptual assimilation model-suprasegmental to tone perception and production .... 61

4.1.3 Perceptual assimilation model and the current thesis .................................... 65
Chapter 5: Categorisation of Cantonese Tones .......................................................... 77
  5.1 Background ........................................................................................................... 78
  5.2 Categorisation of Cantonese by Mandarin Speakers ........................................... 79
    5.2.1 Method ............................................................................................................. 79
      5.2.1.1 Participants ................................................................................................. 79
      5.2.1.2 Stimuli ....................................................................................................... 80
      5.2.1.3 Procedure .................................................................................................. 82
      5.2.1.4 Defining ‘Categorised’ ............................................................................... 82
    5.2.2 Results ............................................................................................................ 83
    5.2.3 Discussion ....................................................................................................... 87
  5.3 Categorisation by English Speakers without Tone Language Experience ............. 90
    5.3.1 Method ............................................................................................................. 90
      5.3.1.1 Participants ................................................................................................. 90
      5.3.1.2 Stimuli ....................................................................................................... 90
      5.3.1.3 Procedure .................................................................................................. 92
    5.3.2 Results ............................................................................................................ 92
    5.3.3 Discussion ....................................................................................................... 96
  5.4 Categorisation by English Speakers Who are Mandarin Learners ......................... 98
    5.4.1 Method ............................................................................................................. 98
      5.4.1.1 Participants ................................................................................................. 98
      5.4.1.2 Stimuli ....................................................................................................... 99
      5.4.1.3 Procedure .................................................................................................. 99
    5.4.2 Results ............................................................................................................ 99
    5.4.3 Discussion ....................................................................................................... 106
  5.5 General Comparison ............................................................................................. 107
  5.5.1 Categorisation by Mandarin speakers and Mandarin learners ............................ 109
  5.5.2 Categorisations by English speakers and Mandarin leaners ............................. 114
  5.6 Summary ............................................................................................................... 119

Chapter 6: Discrimination of Cantonese Tones ......................................................... 120
  6.1 Discrimination of Cantonese Tones by Tone Language Speakers ........................ 121
    6.1.1 Methods ........................................................................................................... 121
      6.1.1.1 Participants ................................................................................................. 121
      6.1.1.2 Stimuli ....................................................................................................... 121
      6.1.1.3 Procedure .................................................................................................. 122
      6.1.1.4 Analysis .................................................................................................... 123
    6.1.2 Results ............................................................................................................ 124
    6.1.3 Discussion ....................................................................................................... 127
  6.2 Discrimination of Cantonese Tones by Non-tone Language Speakers ................... 129
    6.2.1 Methods ........................................................................................................... 129
    6.2.2 Results ............................................................................................................ 130
List of Figures

Figure 4.1 Categorisation of L2 sounds by PAM .......................................................... 58
Figure 5.1. Pitch contour of the four Mandarin tones in /mɔː/ produced by the female speaker .......................................................................................................... 81
Figure 5.2. Pitch contours of the six Cantonese tones in /mɔː/ produced by the female speaker .......................................................................................................... 81
Figure 5.3. Mandarin listeners’ tonal categorisation percentage for each Cantonese tone and its goodness rating in brackets ................................. 85
Figure 5.4. Pitch contour of the five English tunes in /mɔː/ produced by the female speaker .......................................................................................................... 91
Figure 5.5. English listeners’ tonal categorisation percentage for each Cantonese tone and its goodness rating in brackets .......................... 94
Figure 5.6. The Mandarin learners’ tonal categorisation percentage into English tunes for each Cantonese tone and its goodness rating in brackets ...... 101
Figure 5.7. The Mandarin learners’ tonal categorisation percentage into Mandarin tone system for each Cantonese tone and its goodness rating in brackets .................................................................................................. 104
Figure 5.8. Mapping diversity for the six Cantonese tones perceived by Mandarin speakers and Mandarin Learners .......................................................... 112
Figure 5.9. Mapping diversity for the six Cantonese tones perceived by English speakers and Mandarin learners .......................... 117
Figure 6.1. The mean correct discrimination (in percentages) for each Cantonese tone pair by Mandarin listeners .......................................................... 125
Figure 6.2. The mean correct discrimination (in percentages) for each Cantonese tone pair by native Cantonese listeners .......................................................... 125
Figure 6.3. Mean discrimination of the category groups .......................................................... 126
Figure 6.4. The mean correct discrimination (in percentages) for each Cantonese tone pair by English listeners .......................................................... 131
Figure 6.5. The mean correct discrimination (in percentages) for each Cantonese tone pair by Mandarin learners .......................................................... 132
Figure 6.6. Mean discrimination of the category groups .......................................................... 133
Figure 7.1. Tone production by Cantonese speakers .......................................................... 147
Figure 7.2. Tone production by Mandarin speakers .......................................................... 148
Figure 7.3. Tone production by English speakers .......................................................... 148
Figure 7.4. Tone production by Mandarin learners .......................................................... 149
Figure 7.5. Results of Index 2 ......................................................................................... 151
Figure 7.7. Tonal contour by Mandarin speakers .......................................................... 152
Figure 7.8. Tonal contour by English speakers .......................................................... 153
Figure 7.9. Tonal contour by English speakers with Mandarin learning experience ......................................................................................... 154
Figure 7.10. Correlations between perception and production. .............................. 166
Figure 7.11. Correlations between perception and production by Mandarin speakers. .................................................................................................................. 167
Figure 7.12. Correlations between perception and production by English speakers. .................................................................................................................. 167
Figure 7.13. Correlations between perception and production by Mandarin learners. .................................................................................................................. 168
Figure 7.14. Mandarin speakers’ individual performances on perception and production......................................................................................................... 171
Figure 7.15. Mandarin learners’ individual performances on perception and production......................................................................................................... 171
Figure 7.16. English speakers’ individual performances on perception and production......................................................................................................... 172
# List of Tables

Table 2.1 Mandarin Tone Representations and Illustrations ........................................... 15
Table 2.2 Cantonese Tones and Break Indices Transcription of Lexical Tones........... 20
Table 2.3 Cantonese Tones and Break Indices Transcriptions of Boundary Tones .. 20
Table 2.4 Cantonese Tones and Break Indices Transcriptions of Break Indices....... 21
Table 2.5 Mandarin Tones and Break Indices ............................................................... 24
Table 2.6 Comparison between Cantonese and Mandarin Tones ............................. 25
Table 2.7 Australian English Tones and Break Indices .............................................. 28
Table 2.8 Comparison of the Tones and Break Indices Systems for English, Mandarin and Cantonese ................................................................. 33
Table 5.1 Summary of the t-tests of Each Choice—Mandarin Speakers............... 83
Table 5.2 Summary of the Categorisations of the Six Cantonese Tones— Mandarin Speakers ........................................................................................................ 86
Table 5.3 Summary of the Assimilation Patterns—Mandarin Speakers .............. 87
Table 5.4 English Stimuli and Tones and Break Indices Transcriptions ............ 92
Table 5.5 Summary of the t-tests of Each Choice—English Speakers .................. 93
Table 5.6 Summary of the Categorisations of the Six Cantonese Tones—English Speakers ........................................................................................................... 95
Table 5.7 Summary of the Assimilation Patterns—English Speakers ................. 96
Table 5.8 Summary of the t-tests of Each Choice—Mandarin Learners to English Intonation ................................................................. 100
Table 5.9 Summary of the t-tests of Each Choice—Mandarin Learners to Mandarin .............................................................................................................. 100
Table 5.10 Summary of the Categorisations of the Six Cantonese Tones— Mandarin Learners to English Intonation ............................................................. 102
Table 5.11 Summary of the Assimilation Patterns—Mandarin Learners to English Intonation ................................................................. 103
Table 5.12 Summary of the Categorisations of the Six Cantonese Tones— Mandarin Learners ........................................................................................................ 105
Table 5.13 Summary of the Assimilation Patterns—Mandarin Learners to Mandarin .............................................................................................................. 105
Table 5.14 Combination of English and Mandarin Categorisation—Mandarin Learners .............................................................................................................. 106
Table 5.15 Assimilation Fit of Cantonese Tones to Mandarin Tone Categories— Mandarin Listeners and Mandarin Learners ................................................. 111
Table 5.16 Assimilation Fit of Cantonese Tones to English Intonation Categories—English Listeners and Mandarin Learners ................................................. 116
Table 7.1 Token Numbers for Different Speaker Groups ....................................... 146
Table 7.2 Results of Index 1 ......................................................................................... 150
Table 7.3 Mean Duration of the Produced Tones by Different Speakers .......... 157
Table 7.4 Mean Duration for Each Tone Type and t-scores with Bonferroni Corrections between Multi-group Comparisons ................................ 158
Table 7.5 Summary of the Duration Rank ......................................................... 159
Table 7.6 Auditory Analysis of Non-native Productions .................................... 161
Table 7.8 Tone Confusion Patterns ................................................................. 162
Table 7.8 Tone Difficulty by Different Speaker Groups ..................................... 169
Table 7.9 Variance of Perception and Production Performance by Different Speakers ........................................................................................................ 170
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRT</td>
<td>Australian high-rising terminal</td>
</tr>
<tr>
<td>AP</td>
<td>Accentual phrase</td>
</tr>
<tr>
<td>AQI</td>
<td>Australian questioning intonation</td>
</tr>
<tr>
<td>CT</td>
<td>Cantonese tones</td>
</tr>
<tr>
<td>C_ToBI</td>
<td>Cantonese tones and break indices</td>
</tr>
<tr>
<td>CG</td>
<td>Category-goodness</td>
</tr>
<tr>
<td>ip</td>
<td>Intermediate phrase</td>
</tr>
<tr>
<td>IP</td>
<td>Intonational phrase</td>
</tr>
<tr>
<td>IPA</td>
<td>International phonetic alphabet</td>
</tr>
<tr>
<td>IPS</td>
<td>Intermediate intonational phrase</td>
</tr>
<tr>
<td>MT</td>
<td>Mandarin tones</td>
</tr>
<tr>
<td>NA</td>
<td>Non-assimilable</td>
</tr>
<tr>
<td>NLM</td>
<td>Native language magnet model</td>
</tr>
<tr>
<td>PAM</td>
<td>Perceptual assimilation model</td>
</tr>
<tr>
<td>PSOLA</td>
<td>Pitch-synchronous overlap-and-add</td>
</tr>
<tr>
<td>RQ</td>
<td>Research question</td>
</tr>
<tr>
<td>SBE</td>
<td>Southern British English</td>
</tr>
<tr>
<td>SC</td>
<td>Single-category</td>
</tr>
<tr>
<td>SLM</td>
<td>Speech learning model</td>
</tr>
<tr>
<td>TC</td>
<td>Two-category</td>
</tr>
<tr>
<td>ToBI</td>
<td>Tones and break indices</td>
</tr>
<tr>
<td>UC</td>
<td>Uncategorisable-categorisable</td>
</tr>
<tr>
<td>UU</td>
<td>Uncategorisable-uncategorisable</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Background

The rapid, complex and autonomous processes that underpin language use can be difficult to reconcile with the ease with which native speakers use their native language. For example, listening to speech involves hearing, understanding, and interpreting speech sounds (phonemes), words and parts of words (morphemes), and then making sense of the word order to form meaningful sentences. A crucial first step, however, is for speakers to perceive and identify the sounds—phonemes—that make up the words and sentences, as well as the prosodic information that adds to the meaning of the words and sentences. Non-native perception research has long focused on the acquisition and use of non-native phonemes; recently, the field has shifted in focus to prosodic features such as lexical tone, simultaneously adding to our knowledge of phones. However, the extent to which linguistic experience aids or hinders the perception and production of a new tone language, and the way in which perception and production are related, are still unclear.

Extensive research has supported the idea that linguistic experience influences the perception and production of a new or second language not only at the segmental level (Flege, McCutcheon & Smith, 1987; Lecumberri, Cooke & Cutler, 2010; Munro & Bohn, 2007; Polka, 1995), but also in terms of the prosodic features (So & Best, 2008, 2011). Prosodic features work along with segmental features as cues to differentiate words in speech. Of the great variety of important prosodic features, lexical tone has received particular attention (Francis et al., 2008; Gandour et al., 2003; Gottfried & Suiter, 1997; Hallé, Chang & Best, 2004; Xu, Gandour & Francis, 2006; So, 2006; So & Best, 2010; Hao, 2011), adding crucial information to the processing of tone languages. This particular interest in lexical tone is arguably
crucial, as tone languages constitute more than 70% of the world’s languages (Yip, 2002).

Native speakers of tone languages use their native tone system(s) as fluently and efficiently as they use their native phonemes. And being a native speaker of a tone language influences the perception (and production) of non-native tones, just as the native phoneme inventory of a speaker influences their perception and production of non-native phonemes (Burnham et al., 2014; Lee, Vakoch, & Wurm, 1996; Qin & Mok, 2011; So & Best, 2010, 2011, 2014; Wayland & Guion, 2004). Whether native tonal experience facilitates or interferes with non-native tonal perception is, however unclear, as the particular effect of the linguistic experience depends on discrepancies and similarities between the native and the non-native tone systems. For example, previous research has shown that Cantonese speakers make more errors identifying the Mandarin falling-rising tone than do Japanese and English speakers (So & Best, 2010), and English speakers outperform Cantonese speakers on both Mandarin tone identification and reading tasks (Hao, 2011). Additionally, the number of level tones in the listener’s native tone language may positively (Qin & Mok, 2011) or inversely (Chiao, Kabak & Braun, 2011) influence the perception of level tones. However, this assimilation pattern has rarely been examined with listeners from a simpler tonal background. In addition, it remains particularly unclear how L2 tonal experience influences listeners’ perception and production of non-native tones.

Most cross-language speech perception research is conducted through the theoretical frameworks of the perceptual assimilation model (PAM; Best, 1995) and the speech learning model (SLM; Flege, 1995), though both models are historically applied primarily to segmental studies. More recently, however, both PAM and SLM have been modified to also account for differences in the perception and production
of non-native prosodic features. PAM and SLM share a number of common assumptions but also have a significant number of differential predictions, especially regarding the relationship between perception and production. The models, and their extensions to the prosodic features, have been applied to non-native tone research with participants ranging from naïve listeners who speak a non-tone language (e.g., English) to learners of the target language (where the learners include tone and non-tone language speakers). How speakers from a tone background perceive and produce a new tone language has rarely been investigated at the same time; thus, investigating the link between perception and production is a crucial contribution of the present thesis.

An increasing number of studies support the idea that certain aspects of segmental assimilation are transferrable to the prosodic domain (Leung, 2008; Qin & Jongman, 2015; So, 2010; So & Best, 2010; So & Best, 2014). For instance, segmental assimilation usually consists of processing at both phonetic and phonological levels, and successful processing depends on phonemic equivalence and phonemic status, respectively. However, whether a similar approach is present in prosodic research is yet to be determined. Of all prosodic categories, tone is particularly interesting, as the phonemic function is embedded. As such, it is possible that phonetic and phonological assimilation are both present in non-native tone perception (Wu, Munro & Wang, 2014).

1.2 Motivation and Aims

This thesis examines the perception and production of Cantonese tones by speakers of Mandarin (a tone language), English (a non-tone language), and native English speakers with Mandarin L2-learning experience, as well as a native Cantonese-speaking group. The experimental design enables investigation of the way
in which speakers from a language with a smaller tone inventory categorise a more complex tone inventory and how this categorisation pattern influences their perceptual ability. The three languages involved (Cantonese, Mandarin and English) each have a unique prosodic system, and these different native prosodic systems influence the perception and production of tone and other prosodic features. Some interesting interactions can be expected between tone and intonation, as well as systematic problems acquiring a novel tonal system by speakers with and without tone experiences, and several studies have reported on the difficulty of non-native tone learning. This difficulty is experienced not only by learners with typologically different native languages (e.g., non-tone languages) but also by tone language speakers. However, achieving the correct tone is crucial to language understanding in tone languages—tonally minimal pairs exist and influence comprehension significantly in the same way as phonemically minimal pairs in English. Therefore, it is essential to understand precisely what happens during the tone perception and production processes; in particular, the potential benefits and limitations provided by different linguistic experiences.

‘Linguistic experience’ in this thesis refers to both the L1 and any previously acquired L2s (Rast, 2010; Sanz, Park & Lado, 2015). We know that different L1s influence the perception and production of tone languages (Leung, 2008; So, 2010; So & Best, 2014), but we know much less about the impact of a previously acquired language on a third (new) non-native language (L3), especially when the L2 and L3 are typologically similar. Indeed, Qin and Jongman (2015) have highlighted the issue that the transfer source is less obvious when listeners know more than one language. Several theoretical models have been proposed regarding L3 acquisition, but most of these focus on the perception and production of segments (Cabrelli Amaro, 2012;
Wrembel, 2012; Wrembel, Ulrike & Grit, 2010). Consequently, one motivation for the current study stems from the limited literature discussing L3 acquisition in relation to prosody and in particular, lexical tones.

The other motivation for this study arises from the fact that most previous tone studies of English-speaking participants involve American (or Canadian) English speakers (Hao, 2012; Leung, 2008). Few studies on the perception and production of tone languages have been conducted on Australian English speakers. Differences between varieties of the same L1 (i.e., American vs. Australian English) can influence cross-language perception (e.g., Chládková & Podlipský, 2011; Escudero, Simon & Mitterer, 2012; Escudero & Williams, 2012; Marinescu, 2012) and production (e.g., Lew, 2002; Marinescu, 2012; O’Brien & Smith, 2010; Simon et al., 2015). Further, Australian English has unique prosodic features; as such, it is necessary to investigate whether the existing research on American English can be replicated.

The current study departs from the existing literature by comparing participants who are native non-tone language speakers (English) with tone language learners (of Mandarin) where the target language is Cantonese, another tone language with a larger tone inventory than Mandarin. Moreover, it extends PAM to better account for tone perception and provide testable hypotheses also in this domain. SLM is also extended to investigate the relationship between tone perception and production, and provide testable hypotheses for that domain. This cross-language study examining the tone perception and production of Cantonese by speakers with varying levels of contact with tone languages will not only add empirical data to the increasingly important speech perception and production field.
More significantly, it will provide testable and comprehensive predictions for non-native tone research.

1.3 Thesis Structure

This thesis consists of three main sections: Chapters 1 to 4 introduce the research context and relevant theoretical models. Chapters 5 to 7 report on two perception experiments (tone categorisation in Chapter 5 and tone discrimination in Chapter 6, respectively) and one production experiment, with the methodology and results introduced separately. Chapter 8 summarises the findings and discusses the link between perception and production, concluding the thesis with a look at future research. An overview of each chapter follows below.

‘Chapter 2 Tone and Intonation’ introduces the linguistic uses of pitch in tone and intonation, and the Autosegmental Metrical (AM) approach, as well as the Tones and Break Indices (ToBI) transcribing traditions in Cantonese, Mandarin and English to illustrate the different prosodic systems these three languages possess.

‘Chapter 3 Literature Review of Tone Perception and Production’ includes a review of the key literature. Here, I introduce tone and then summarise the research conducted with different listeners. In the domains of non-native tone perception and production, research with tone and non-tone speakers is reviewed separately.

‘Chapter 4 Theoretical Models and Thesis Overview’ extends the two theoretical models to predict and account for tone perception and production. An overview of the research questions and experimental chapters are also provided in this chapter.

‘Chapter 5 Categorisation of Cantonese Tones’ and ‘Chapter 6 Discrimination of Cantonese Tones’ makes up this thesis’s perception study. Tone categorisation and discrimination results are discussed separately in Chapters 5 and
6, by groups: tone language groups (Cantonese and Mandarin speakers) and non-tone language groups (English monolinguals and English speakers who are Mandarin learners).

‘Chapter 7 Production of Cantonese Tones’ investigates native and non-native tone production results using a number of analytical methods: tone differentiation, tone movements, tone errors and duration differences. All of the four speaker groups are compared simultaneously to determine the differences in the way in which each produces tones. The link between perception and production is discussed here, as well as an investigation of individual differences.

‘Chapter 8 Discussion and Conclusion’ compares and discusses the perception and production results. It first summarises observations arising from the two experiments, before outlining answers to the research questions raised in Chapter 4. Conclusions are drawn, together with an outline of limitations and areas for future research.
Chapter 2: Tone and Intonation

The present chapter introduces the way in which prosody—and lexical tone in particular—operates in typologically different languages. The prosodic systems of the three languages relevant to the current study are discussed in detail. As is argued throughout this thesis, such discussion provides the foundation for any language specific predictions of cross-language perception and production, including cross-language perception of prosodic features such as tone.

2.1 Prosody

2.1.1 Prosodic typology

To understand how prosodic features work in one language and interact with other languages, it is important to comprehend what prosody is and how prosodic features categorise languages into different types. A prosodic structure is a hierarchical organisation of prosodic units from the smallest (mora or syllable) to the largest (intonation phrase or utterance). Prosody at the word and phrase levels forms the prosody of an utterance. A number of models have been proposed to explain the prosodic hierarchical structure (a review of these is given in Shattuck-Hufnagel and Turk [1996]). The study of prosody usually takes its surface manifestations, such as duration, intensity, and fundamental frequency (F0) to indicate the different levels of hierarchical structure (Beckman, 1996; Beckman & Edwards, 1990). These features can help divide sentences into different hierarchical structures: sentences into phrases, phrases into words and words into syllables. At the same time, these hierarchical patterns indicate the prosodic features.

A common model of prosodic typology proposes that prosody includes both prominence and phrasing (Jun, 2006). Prominence and phrasing exist at both the word and phrase level simultaneously. Word-level prominence is proposed to include
these types: tone, stress, and pitch accent. Tone languages have ‘prescribed pitches for syllables or sequences of pitches for morphemes or words’ (Cruttenden, 1994, pp. 8–9); that is, pitches have paradigmatic contrasts. Stress-accented languages maintain one syllable in the word as more prominent, as in English. The pitch information of these syllables does not carry lexical information but can be realised with a certain pitch pattern in the case of intonation. In lexical pitch accent languages, certain syllables are lexically specified with a pitch movement but no phonetic ‘stress’ in the sense of Beckman (1986), as in Japanese. However, recent research posits different definitions of stress and pitch accent typologies; Hyman (2006, 2009, 2010) has proposed a properties-driven typology approach.

Post-lexically, prominence is realised at the beginning of a prosodic unit (head) and/or the end of one (edge) (Beckman 1986; Beckman & Edwards 1990; Hyman 1978; Ladd 1996; Venditti, Jun & Beckman, 1996). Post-lexical prominence manifests through suprasegmental features such as pitch, duration and/or amplitude. If post-lexical prominence arises from lexical pitch accent—as in Japanese—duration or amplitude will not undergo change. By contrast, if it arises from stress accent—as in English—both duration and amplitude will be affected, relative to surrounding syllables. Prosody at the phrase level is an addition to all lexical prosodic typologies. All three categories mentioned above interact with post-lexical prosody, particularly intonation. A syllable with pitch information can carry sentence stress, a phrasal tone, or a boundary tone simultaneously.

Apart from prominence features, prosody also requires examination in terms of the phrasing pattern, which is categorised by the type of prosodic unit it is associated with. Like prominence, phrasing also includes lexical and post-lexical levels. Lexically, moras, syllables and feet can be identified with variations in
different languages. Difference at this level contributes to the impressionistic rhythm classes like mora-timed languages (e.g., Japanese), syllable-timed (Spanish) and stress-timed languages (English) (e.g., Abercrombie, 1967; Bloch, 1950; Lehiste, 1976; Pike, 1945). Post-lexically potential prosodic units include accentual phrase (AP), intermediate phrase (ip) and intonational phrase (IP).

Jun’s (2014) revised model of prosodic typology adds the parameter of macro-rhythm and updates the prosodic typology model to include a combination of prominence marking, macro-rhythm and word prosody. Prominence and macro-rhythm are at the phrase level, while word prosody is at the lexical level. Languages that maintain pitch accent at the lexical level (lexical pitch-accent languages) and at the post-lexical level (lexical stress-accent languages/post-lexical stress-accent languages) are all head-prominent languages. Tone languages also belong to this category, as the tonal specification of a syllable or mora marks the phrasal prominence in tone languages. When head-prominent languages also have prominence marking associated to the edge of a word boundary, they are head/edge languages. These languages either have lexical pitch accent and a word/AP boundary tone simultaneously, or have a post-lexical pitch accent and a simultaneous AP-like phrasal or boundary tone. Edge languages are those that only have AP-like phrasal/boundary tones, lacking lexical and post-lexical heads. French is an example of such a language.

According to Jun (2014), a macro-rhythm is defined as a ‘phrase-medial tonal rhythm whose unit is equal to or slightly larger than a word, and the tones forming a tonal unit can be pitch accents, lexical tones, or boundary tones’ (p. 526). Macro-rhythm degrees are generally categorised into three levels: strong, medium and weak. Four types of word prosodies are identified: stress, tone/lexical pitch accent, both of
stress and tone, none of stress and tone. According to Jun, combined with prominence and macro-rhythm features, languages are re-grouped into 15 types. Languages included in the current study are Australian English, Mandarin and Cantonese; these all belong to the group of head-prominent languages. Australian English has medium macro-rhythm and stress, while Mandarin and Cantonese share a similarly weak macro-rhythm. However, Mandarin has both tone and stress, while Cantonese has tone only. By investigating the prosodic typologies to which Mandarin, Cantonese and English belong, we will have a better idea of the similarities as well as the differences between them. On the basis of this understanding, a more precise prediction of how speakers perceive and produce non-native tones can be made.

2.1.2 Autosegmental Metrical and Tones and Break Indices transcriptions

Given the considerable language-to-language variation of prosodic systems, it is much easier to compare language prosody within a single framework. A number of models have been created to analyse and transcribe intonation systems; these exhibit great variation. In an Autosegmental Metrical (AM) model, an utterance can have tone targets of different pitch height (low and high) in a sequence, according to prosodic typologies. Currently, the tone and break indices transcription (ToBI)—an adapted version of the AM model—has been adopted in laboratory phonology research (e.g., Beckman, Hirschberg & Shattuck-Hufnagel, 2005; Fletcher & Harrington, 2001; Wong, Chan & Beckman, 2005), although it must be noted that ToBI does not provide a universal model and must be adapted to fit individual languages. The intonation framework inventory and its different ToBI conventions enable comparison across languages.
In a ToBI-style intonational analysis, the prosodic structures of an utterance can be represented by projecting separate prosodic information onto the four tiers: tone, orthographic, break index and miscellaneous. The tone tier is used to label tonal events (namely, the pitch accents) and/or phrase tones and boundary tones with edges marked. The break index tier uses numeric labels (0–4) at the end of each word, suggesting the hierarchical prosodic constituency and prosodic grouping. As the current study involves typologically different prosodic systems, a brief introduction to AM and ToBI will enable comparison of ToBI adaptations for these languages (see Sections 2.2.1 and 2.2.2).

2.1.3 Transfer of native prosodic systems

A number of studies have investigated the influence of a L1 prosodic system on L2 production. Aoyama and Guion (2007) compared the two English prosodic features (duration and F0) in productions by native Japanese and English children and adults. As discussed before, Japanese is a mora-timed language, while English is a stress-timed language, and these different prosodic systems explain differences in the absolute syllable and utterance durations produced by English speakers and Japanese speakers (English < Japanese), as well as differences in the F0 range between native and non-native English speakers (English < non-native English).

Native lexical stress systems also influence L2 production (see discussion in Nguyễn, Ingram & Pensalfini, 2008). For example, in English, stress can be correlated with differences in duration, intensity and vowel quality, whereas in Vietnamese, stress is only associated with differences in pitch and intensity. This difference might explain Vietnamese speakers’ difficulty in realising the duration contrast between accent-contrasted syllables in compound words and phrases or polysyllabic words and phrases. They were able to contrast F0 and intensity on
accent-bearing syllables while failing to deaccent those elements requiring narrow focus.

In terms of phonological quantity, the acquisition of the Swedish quantity contrast (i.e., short and long vowel duration contrasts) has been investigated by speakers with different language backgrounds: Estonian, English and Spanish (McAllister, Flege & Piske, 2002). In Swedish and Estonian, the differentiation of mid-vowels relies largely on a systematic difference in duration. However, duration is not the primary cue to differentiate English mid-vowels and it does not even exist in Spanish. Indeed, in a perception and a production task of the four Swedish vowel pairs that contrast in duration, Estonian participants (duration plays an important role in differentiating Estonian mid-vowels) outperformed English speakers (duration plays a less important role in differentiating English mid-vowels), while Spanish speakers (lacking duration differences in Spanish mid-vowels) performed the poorest. These results mirror the importance of durational contrasts across the three languages.

This section has briefly introduced the features of pitch and the possibility of L1 prosody transfer. The current study aims to investigate the link between the two uses of pitch: tone and intonation. These will be discussed in detail in the following sections.

2.2 Tone Languages

Tone is ‘the use of pitch in language to distinguish lexical or grammatical meaning’ (Yip, 2002, p. 1). The primary phonetic correlates of tone are F0 height, F0 movement and duration. Contrasting tones distinguish words in a manner quite similar to a phoneme change in a minimal pair; a tone change can result in a different word, just like for example, a voice onset time difference in the initial stop in the
English word ‘pat’ provides the primary method of differentiation from the English word ‘bat’. Tone is primarily a matter of pitch, but may also involve accompanying differences of segment duration and voice quality. For example, in Mandarin Chinese, syllables with T214, the dipping tone, are not only low in pitch but tend to have longer duration and a creaky/glottalised voice quality. Tone often functions similarly to segmental distinctions, involving a choice of categories from a paradigmatic set. It is meaningful to discuss contrasts between tones on a particular syllable without referring to the tones on another syllable. Accentual distinctions, by contrast, are syntagmatic: they involve contrast with adjacent syllables in a string. An example of Mandarin tone contrasts can be found in Table 2.1.

Tonal contours involve changes in pitch within one syllable, and while most tones exhibit some pitch change (Xu & Wang, 2001); tones produced with largely the same pitch throughout are considered to be ‘level’ tones. When a more significant pitch change occurs, a tone is likely to be classified as one of a range of ‘contour’ tone types. A rising tone is one where the pitch moves from a lower to a higher point in the speaker’s pitch range and a falling tone is one where the reverse pattern is evident. Sometimes a combination of rising and falling movement is carried by a single syllable. For example, Mandarin Chinese has four regular tones and a neutral tone. The neutral tone is mostly present in function words and possesses the same pitch value as the preceding tone. The numbers displayed in the pitch column in Table 2.1 represent the pitch of each tone at the beginning and end. For the Mandarin falling-rising tone, the pitch value in the middle represents the dipping movement of this particular tone. The numbers are given on a 1 to 5 scale, with 1 referring to the lowest pitch of the speaker and 5 to the highest pitch. The scale represents the linguistic tonal space. This 5-point scale was first introduced by
Chao (1930) and has since been adopted widely (e.g., Ladefoged & Johnson, 2001; Yip, 2002). Pitch height and tone movement can also be represented graphically, as shown in Table 2.1’s tone graphic column. In these graphics, the vertical line stands for the pitch range of a speaker’s voice and a line to its left indicates both pitch movement and relative pitch height. For example, \( \downarrow \) means that a tone falls from the top of the speaker’s pitch range to the bottom: this is known as the high-falling tone.

Table 2.1

**Mandarin Tone Representations and Illustrations**

<table>
<thead>
<tr>
<th>Tone number</th>
<th>Description</th>
<th>Tone graphic</th>
<th>Pitch</th>
<th>Example</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High-level</td>
<td>( \uparrow )</td>
<td>55</td>
<td>ma \textsuperscript{55}, mā</td>
<td>mother</td>
</tr>
<tr>
<td>2</td>
<td>High-rising</td>
<td>( \uparrow )</td>
<td>35</td>
<td>ma \textsuperscript{35}, má</td>
<td>hemp</td>
</tr>
<tr>
<td>3</td>
<td>Low-falling-rising</td>
<td>( \downarrow )</td>
<td>214</td>
<td>ma \textsuperscript{214}, mā</td>
<td>horse</td>
</tr>
<tr>
<td>4</td>
<td>High-falling</td>
<td>( \downarrow )</td>
<td>51</td>
<td>ma \textsuperscript{51}, mā</td>
<td>scold</td>
</tr>
</tbody>
</table>

Over 70% of the world’s languages have lexically contrastive tones; they are widespread in the Asia-Pacific region, Africa and America. African and American tonal languages have relatively simple tonal inventories: they tend to contrast relative tone heights, such as high and low or high, mid and low (Yip, 2002) to distinguish meaning. Tonal languages in Asia and the neighbouring Pacific regions include the Sino-Tibetan family (which includes the Chinese language family and the Tibeto-Burman language family), Austro-Tai (which includes Tai-Kadai, Miao-Yao and Austronesian), Vietnamese, and Papuan languages, as well as register-based
languages like most of Mon-Khmer. Asian tonal languages generally have richer tonal inventories, including a set of contour tones and contrasting level tones, meaning that they contrast on both pitch trajectory and height.

The following sections will focus on the two tone languages involved in this study: Cantonese and Mandarin, both of which include contour and level tones in their tone systems.

2.2.1 Cantonese tone system

Standard Cantonese, a Yue dialect of Chinese, is spoken in Hong Kong and Canton (Guangzhou) (for a comprehensive review of Cantonese, see Hashimoto [1972]). It is estimated that Cantonese is spoken by 66 million speakers in Hong Kong, Macao and Canton; it is ranked sixteenth among all of the world’s languages in terms of the total number of speakers (Grimes, 1996). Descriptions of Cantonese tones have varied throughout history. Most of the current research uses Chao tone letters. Jones and Woo (1912) use musical notation, while Mathews and Yip (1994) use prose. Some studies provide acoustic analysis of individual speakers’ tone production, for example, Hashimoto (1972) and Vance (1976).

Standard Cantonese refers to Cantonese spoken in Hong Kong and in Canton province; however, some differences have emerged over time (Bauer & Benedict, 1997). Variations are found only paradigmatically in the lexical tone and boundary tone inventory, not in the dense syntagmatic specification of tone. In Cantonese, every syllable bears a lexical tone, including all particles. Phrase-final syllables carry a specified boundary tone, which is a separate pragmatic morpheme specified for the phrase as a whole. The number of final particles range from 30 (Kwok, 1984) to 206 (Yau, 1980). For example, 呀 (aa3), 啊 (ge3), 啊 (laa1) are the three final particles used in neutral questions, assertions to emphasise, and in requests and imperatives
respectively. By contrast, standard Mandarin has only seven commonly used particles (Matthews & Yip, 1984). These final particles interact with tonal pragmatic morphemes (‘boundary tones’) to convey complicated pragmatic functions (Chan et al., 1998; Fung, 2000; Kowk, 1984). For instance, 了 (le), 呢 (ne) and 吧 (ba) do not bear any tone themselves but they will interact with boundary tones.

Currently, there is disagreement in the literature with respect to the number of tones that Cantonese maintains, although this is due largely to varying analysis methods. Four different inventories are proposed:

- a six-tone system (e.g., Matthews & Yip, 1994; Rose, 2000; Tong & James, 1994), consisting of three level and three contour tones
- a seven-tone system (e.g., Chik, 1980; Kuan et al., 1991), consisting of three level and three contour tones, as well as a high-falling tone. This last is no longer used contrastively in Hong Kong Cantonese, although it might be present in some speakers’ speech as a tone on the two sentence-final particles ‘sin’ and ‘tim’ (Matthews & Yip, 1994)
- a nine-tone system (e.g., Dodd & So, 1994; So & Dodd, 1994; Tse, 1978; Tse, 1993), again consisting of three level and three contour tones with the addition of t, which are three tones observed only in closed syllables ending in voiceless stops. These tones are referred to as ‘entering’ or ‘stopped’ tones. Matthews and Yip (1994) contend that entering tones are simply allotones of the basic tones
- a ten-tone system (e.g., Bauer & Benedict, 1997), consisting of the seven tones from Chik (1980) and the three stopped tones from Dodd and So (1994).
Clearly, while consensus has not yet been reached, the predominant position is that six basic lexically contrastive tones exist in Hong Kong Cantonese.

The unique syllable structure of Cantonese is one of the language’s most important characteristics, and it is easy to define at the phonological level. The link between word and syllable is quite strong, and is readily recognisable from the asymmetric distribution of onsets and codas. Cantonese syllables consist of an optional onset consonant and a rhyme, which has either a simple vowel, a simple vowel followed by an optional coda consonant, a vowel-glide diphthong, or a syllabic nasal.

The complex tone inventory is another signature characteristic of Cantonese (as mentioned at the beginning of Section 2.2.1). Compared to Mandarin, Cantonese has fewer disyllabic words; however, in Hong Kong Cantonese, some segmental effects fuse two syllables into a polysyllabic word in fast speech (Li, 1986; Wong, 1996; Wong, 2002). When the second of two syllables has undergone substantial weakening or an effective deletion of segmental information (in extreme cases, the simplification of contour tones and vowel), a merger may occur. However, fusion does not usually override the syllables’ lexical tones. In a very few of the most extreme cases, tone loss occurs (see Wong [2006] for a review of this phenomenon).

In addition to variation derived from fusion, a number of categorical segmental and tonal alternations are particularly interesting. This is especially so in Hong Kong Cantonese, due to its special geographical and historical context. Under certain conditions, Cantonese tones can change: this is known as ‘changed tone’ and two types have been identified. The first is tonal assimilation: this is phonetic in origin and occurs due to the influence of the tonal environment. In these instances, changes in tones do not affect word meanings. In certain bisyllabic words, if the first tone is
high-level, then the second syllable can be assimilated to high-level. The second type is morphological changed tones, which function as a morphological device for deriving new words. These tone changes affect the meanings of words. The original tone will change into high-rising or high-level to indicate that the word belongs to a colloquial register. This can alter the word morphologically, usually by giving a special meaning to concrete nouns, indicating that something is familiar or common.

The AM analyses for Cantonese and some proposed ToBI transcription conventions can be found in Wong et al. (2005). Cantonese ToBI (C_ToBI) is the ToBI convention used to annotate and transcribe Cantonese. Even though C_ToBI is based on Hong Kong Cantonese, it is suitable for transcribing other varieties. In general, this transcribing convention specifies six levels of transcription: 1) tones, 2) break indices, 3) any polysyllabic foot, 4) syllables, 5) words, and 6) miscellaneous.

Both lexical tone and boundary tone information is tagged onto the tone tier. Chao numbers are ordinarily used, with minor adjustments: the first number is doubled for non-checked contour tones; for checked syllables, the first letter is deleted to signify the shorter duration. For phrase-final boundary tones, six types have been identified in Hong Kong Cantonese. L% and H% indicate fall/rise from the final lexical tone. H:% shows a rise from the final lexical tone, but with a short plateau at the very end of the rise. HL% is used to label a final rise and then a fall from the final lexical tone. No extra tone at the end is indicted by %, while -% indicates a truncated rise of the final lexical tone. A frame-initial boundary used to mark the initial particle is represented by %-fi. These boundary tones can occur with or without a final particle/particle sequence. Tables 2.2 to 2.4 include the transcribing conventions and their descriptions for lexical tones, boundary tones with ToBI.
Table 2.2  

*Cantonese Tones and Break Indices Transcription of Lexical Tones*

<table>
<thead>
<tr>
<th>Tones</th>
<th>Non-checked syllable</th>
<th>Checked syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level tones</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-level</td>
<td>55</td>
<td>5</td>
</tr>
<tr>
<td>Mid-level</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Low-level</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td><strong>Rising tones</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-rising</td>
<td>335</td>
<td>35</td>
</tr>
<tr>
<td>Low-rising</td>
<td>223</td>
<td>--</td>
</tr>
<tr>
<td><strong>Falling tones</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-falling</td>
<td>221</td>
<td>21</td>
</tr>
<tr>
<td>High-falling*</td>
<td>553</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 2.3  

*Cantonese Tones and Break Indices Transcriptions of Boundary Tones*

<table>
<thead>
<tr>
<th>Type</th>
<th>Tier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L%</td>
<td>Tone</td>
<td>Fall from the final lexical tone</td>
</tr>
<tr>
<td>H%</td>
<td>Tone</td>
<td>Rise from the final lexical tone</td>
</tr>
<tr>
<td>H:%</td>
<td>Tone</td>
<td>Rise from the final lexical tone, with a short plateau at the very end of the rise; incredulity reading accompanied</td>
</tr>
<tr>
<td>HL%</td>
<td>Tone</td>
<td>Final rise and then fall from the final lexical tone</td>
</tr>
<tr>
<td>%</td>
<td>Tone</td>
<td>Phrase end with no extra tone</td>
</tr>
<tr>
<td>-%</td>
<td>Tone</td>
<td>Truncated rise of the final lexical tone</td>
</tr>
<tr>
<td>%fi</td>
<td>Tone</td>
<td>Frame-initial boundary used to mark the initial particle in phrase-framing particle pairs</td>
</tr>
</tbody>
</table>
Table 2.4

*Cantonese Tones and Break Indices Transcriptions of Break Indices*

<table>
<thead>
<tr>
<th>Types</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Foot internal syllable boundary</td>
</tr>
<tr>
<td>1</td>
<td>End of a syllable that is also end of foot</td>
</tr>
<tr>
<td>2</td>
<td>Intonation phrase end</td>
</tr>
<tr>
<td>1-</td>
<td>Uncertainty between 0 and 1</td>
</tr>
<tr>
<td>2-</td>
<td>Uncertainty between 1 and 2</td>
</tr>
<tr>
<td>c</td>
<td>An abrupt, disfluent cut-off of phonation</td>
</tr>
<tr>
<td>p</td>
<td>Prolongation at a disfluency (‘hesitation pause’)</td>
</tr>
</tbody>
</table>

2.2.2 Mandarin tone system

Mandarin has a number of different varieties, including three national standards (Guoyu in Taiwan, Putonghua in mainland China, and Huayu in Singapore) as well as many regional varieties. Yuan (1989) categorises Mandarin into four main varieties based on both geography and phonological differences: northern Mandarin, north-western Mandarin, south-western Mandarin and Jianghuai Mandarin. In addition, Mandarin is spoken as a native language in Taiwan, Singapore, Indonesia, Thailand and other parts of Southeast Asia, the United Kingdom, North America and South Africa.

Putonghua Mandarin is the official language of China, and is widely spoken in all provinces except for Canton, Hong Kong and Macao (where Cantonese is spoken). Mandarin is, however, a compulsory school subject for Cantonese speakers, and announcements at subways and railway stations are made in both Mandarin and Cantonese in Cantonese-speaking areas. Conversely, residents in non-Cantonese-speaking areas of China have little experience with Cantonese.
All Mandarin varieties have tone as a salient characteristic. Standard Mandarin varieties have four contrastive tones and thus constitute a smaller tonal inventory than does Cantonese (Duanmu, 2000). The first tone (T55 in Chao numbers) is a high-level tone. T35, a rising tone, is the third Mandarin tone. Tone 3 (T214) is a dipping tone in citation form, however when occurring in a T3T3 combination, the first dipping tone becomes Tone 2, a simple rising tone (T35). The fourth tone is a high falling tone (T51), characterising a sharp fall from high to low. In Mandarin, some morphemes, such as the agreement-soliciting particle –ba, the pragmatic particles –ma and –a, the verbal suffix –le, and the nominal suffix –zi, are inherently unspecified for tone. These morphemes carry a ‘neutral tone’, sometimes called ‘Tone 5’. This is a significant difference from the Cantonese tone system, where every syllable bears a lexical tone (as reviewed in Section 2.2.1). Unlike Cantonese, Mandarin has stress, inherent both in the lexical entry for some morphemes (neutral tone) and at the phrasal level. As the neutral-tone syllable cannot exist on its own, a bimoraic foot system has been proposed (Duanmu, 1990; Wright, 1983; Yip, 1980); this system is based on the idea that although a full-toned syllable can be a foot by itself, a neutral-tone syllable is necessarily footed together with the preceding full-toned syllable. In contrast, Shih (1986, 1997) suggests that even a full-toned monosyllabic word cannot form a foot by itself; this perspective emphasises Mandarin’s predominantly disyllabic rhythm. Additionally, Mandarin contains many more disyllabic words than does Cantonese.

Stress at the phrasal level is expressed in terms of an exaggerated pitch range on stressed components. Jin (1996) and Xu (1999) suggest that the expansion is most obvious on the focused word, with compression extending over the whole phrase after the stressed word. Manipulation can sometimes be realised more on lowering
the components following the focus word, resulting in a relative greater pitch excursion on the stressed word/syllable. In addition, Mandarin has its tone sandhi rules (for a review, see Peng et al. [2005]). Sandhi rules give rise to the superfoot concept in Mandarin. Boundary tones have been identified, along with global pitch range effects that can signal contrasting pragmatic meanings. These are comparable to Cantonese’s intonation phrase.

Combining the complex prosodic features and with the aim that it is applicable to as many varieties of Mandarin as possible, a ToBI system with eight tiers has been proposed for Mandarin—Pan M_ToBI (Peng et al., 2005). These eight tiers are word, romanisation, syllable, stress, sandhi, tone, break indices and codes. ‘Stress’ includes the relative degree of stress marked on each syllable, manifested by both segmental and prosodic features. ‘Tone’, which is of particular interest here, includes the marking of boundary tones and pitch range effects. Break indices indicate the hierarchy of disjunction to represent prosodic phrasing.

As the relationship between stress and tone sandhi is yet to be fully understood, both stress and sandhi annotations are included in the current ToBI system. On the stress tier, four levels are identified: S3 for a syllable with a fully realised lexical tone; S2 for a syllable with a substantial tone reduction; S1 for a syllable that has lost its lexical tonal specification and S0 for a syllable with a lexical neutral tone.

Unlike Cantonese, Mandarin has separate tiers for boundary tones and lexical tones. In this intonation tier, boundary tones and pitch range effects are present. For boundary tones, the traditional symbols L% and H% are applied. For global pitch range, all symbols are used to signal the beginning of a pitch range change: %reset for a new pitch downtrend or reset; %q-raise for a raised pitch range (e.g., in echo
questions); %e-prom for the local expansion of pitch range due to emphatic prominence and %compressed for a reduction in pitch range of syllables following %e-prom. Detailed descriptions are given in Table 2.5 (for a full discussion, see Peng et al. [2005]).

Table 2.5

*Mandarin Tones and Break Indices*

<table>
<thead>
<tr>
<th>Label</th>
<th>Tier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3</td>
<td>Stress</td>
<td>Syllable with fully realised lexical tone</td>
</tr>
<tr>
<td>S2</td>
<td>Stress</td>
<td>Syllable with substantial tone reduction</td>
</tr>
<tr>
<td>S1</td>
<td>Stress</td>
<td>Syllable that has lost its lexical tonal specification</td>
</tr>
<tr>
<td>S0</td>
<td>Stress</td>
<td>Syllable with lexical neutral tone</td>
</tr>
<tr>
<td>35</td>
<td>Sandhi</td>
<td>Tone 3 realised as a sandhi tone (rising tone)</td>
</tr>
<tr>
<td>214</td>
<td>Sandhi</td>
<td>Tone 3 realised as a low-dipping tone</td>
</tr>
<tr>
<td>H%</td>
<td>Tones</td>
<td>High boundary tone (at the end of an utterance)</td>
</tr>
<tr>
<td>L%</td>
<td>Tones</td>
<td>Low boundary tone (at the end of an utterance)</td>
</tr>
<tr>
<td>%reset</td>
<td>Tones</td>
<td>Beginning of a new pitch downtrend or pitch reset</td>
</tr>
<tr>
<td>%q-raise</td>
<td>Tones</td>
<td>Beginning of a raised pitch range</td>
</tr>
<tr>
<td>%e-prom</td>
<td>Tones</td>
<td>Beginning of local expansion of pitch range due to emphatic prominence</td>
</tr>
<tr>
<td>%compressed</td>
<td>Tones</td>
<td>Beginning of reduction of pitch range of syllables following the expansion of pitch range under %e-prom</td>
</tr>
</tbody>
</table>

The differences between Cantonese and Mandarin are quite apparent. For example, Mandarin maintains stress in addition to tone, while stress does not exist in Cantonese (Beckman & Venditti, 2010). Additionally, Cantonese and Mandarin are not mutually intelligible, and have different phonological properties and lexicons and—to some extent—different syntax. Table 2.6 compares the general tone systems of Cantonese and Mandarin, where we can see that Cantonese has two additional
lexical tones and some tones that share the same pitch contour but have different pitch height. Both languages have high-level and high-rising tones, but Cantonese has additional level tones and an additional rising tone. In contrast, Mandarin has a high-falling tone with no precise Cantonese counterpart. Informally, in Mandarin, pitch contour is reported as the most important cue for native speakers to discriminate among tones, while both pitch height and contour are considered the most salient features of Cantonese tones (Yip, 2002).

Table 2.6

Comparison between Cantonese and Mandarin Tones

<table>
<thead>
<tr>
<th>Tonal Contour</th>
<th>Tones</th>
<th>Cantonese</th>
<th>Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>High</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Rising</td>
<td>High</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Falling</td>
<td>High</td>
<td>x</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Dipping</td>
<td>Low-falling</td>
<td>x</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>=6</td>
<td>=4</td>
</tr>
</tbody>
</table>

2.3 Intonation Languages

2.3.1 English intonation

Intonation languages use pitch variation to signal focus, juncture, pragmatic inference and discourse function (Beckman, 1986). According to Ladd (2008), intonation is ‘the use of suprasegmental phonetic features to convey “post-lexical” or sentence-level pragmatic meanings in a linguistically structured way’ (p. 4). For example, ‘Mary can drive.’ has a significantly different meaning from ‘Mary can drive?’. Intonation refers to the phrase/sentence-level uses of pitch that convey
distinctions related to sentence modality and speaker attitude, phrasing, discourse grouping and information structure (Himmelmann & Ladd, 2008). Intonational features at major prominent syllables and boundaries are the differences between pitch accents and boundary tones.

Similar to lexical tone, intonation involves different patterns of fundamental frequency; however, while tone operates at the word or lexical level, intonation operates at the phrase or sentence (i.e., post-lexical) level. In English, it is possible for a sentence to consist of only one word (a monosyllabic sentence), but more often sentences consist of more than one word. When a sentence is made up of several words, it may contain more than one intonation pattern—the sentence will be separated into IPs that are dominated by the accented word. This accented word contains the metrically strongest syllable, which is often referred to as the nuclear or tonic syllable.

A number of models have been proposed to explain the prosodic hierarchical structure. That used here is Beckman and Pierrehumbert’s (1986) model. Two levels of prominence are identified: at the syllable and phrase level. Usually a word contains more than one syllable, and some syllables are more prominent than others. English exhibits a difference between strong (stressed) and weak (unstressed) syllables. A stressed syllable usually involves a long or short vowel of full vowel quality, while an unstressed syllable usually has a schwa or weak lax vowel as its nucleus. These syllables are then grouped into left-headed feet. Prosodic words consist of feet, each of which contains one stressed syllable. Within one prosodic word, only one foot can have the most prominent syllable, which carries the word’s main stress. Prominence at the phrase level happens over the intermediate IP, which can consist of one or more prosodic words that optionally bear pitch accents.
associated with their main stress. The tonic or nuclear stressed syllable is the last pitch-accented syllable in the ip. IPs can consist of one or more ip. The prosodic patterns over IPs are formed by the pitch accents that signal prominence, along with phrase and boundary tones that demarcate the edges of these post-lexical prosodic constituents.

2.3.2 Australian English

Australian English has long been regarded as a variation of Southern British English, but it differs significantly in the phonetic characteristics of vowels, as well as some allophonic and reduction processes (Cox, 2008; Cox & Palethorpe, 2007). Prosodic features and voice quality differences exist between Australian and other English varieties. This thesis focuses on the prosodic features of Australian English. A general transcription of Australian English tunes within a ToBI AM framework is given in Table 2.7, where tonal categories and their general pitch description and major break indices are listed.
Table 2.7
Australian English Tones and Break Indices

<table>
<thead>
<tr>
<th>Intonation events</th>
<th>Pitch description</th>
<th>Australian English ToBI label</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pitch accents</strong></td>
<td>Simple high</td>
<td>H*</td>
</tr>
<tr>
<td></td>
<td>Simple low</td>
<td>L*</td>
</tr>
<tr>
<td></td>
<td>Rising</td>
<td>L+H*</td>
</tr>
<tr>
<td></td>
<td>‘scooped’</td>
<td>L*+H</td>
</tr>
<tr>
<td></td>
<td>Downstepped high</td>
<td>!H*</td>
</tr>
<tr>
<td></td>
<td>Downstepped rising</td>
<td>L+!H*</td>
</tr>
<tr>
<td></td>
<td>Downstepped ‘scooped’</td>
<td>L*+!H</td>
</tr>
<tr>
<td></td>
<td>Downstepped high from preceding H tone</td>
<td>H+!H*</td>
</tr>
<tr>
<td><strong>Phrase accents</strong></td>
<td>High</td>
<td>H-</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>L-</td>
</tr>
<tr>
<td></td>
<td>Downstepped high mid</td>
<td>!H-</td>
</tr>
<tr>
<td><strong>Boundary tones</strong></td>
<td>High</td>
<td>H%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>L%</td>
</tr>
<tr>
<td><strong>Additional pitch labels</strong></td>
<td>Highest pitch value for intermediate phrase (excluding phrase accents or boundary tones)</td>
<td>HiF0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Break indices</th>
<th><strong>Prosodic structure</strong></th>
<th><strong>BI</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>InP</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Source: Adopted from Fletcher et al. (2005)

‘Uptalk’ is commonly used in Australian English; this is the use of a high-rising terminal contour on statements. The Australian National Database of Spoken Language corpus (Fletcher, Grabe & Warren, 2005) identifies five rising types in Australian speakers: simple low-rises, L* L-H%; simple low-onset high-rises, L* H-H%; simple high-rises, (L+) H* H-H%; fall-rises H* L-H%; and expanded-range fall-rises; H*+L H-H%. Different rises have different functions. The two simple high-rises (L* H-H% and H* H-H%) have distinct functions within a dialogue act.
framework: H* high-rises are used for information requests (yes/no questions), while L* high-rises are used for explanations, opinions and instructions (Fletcher, Stirling, Mushin & Wales, 2002). In this ToBI-labelled map-task, 97% of the H* H-H% rises were information requests, while L* L-H% were used more for statement conditions, including acknowledgement/answer, acceptance dialogue acts and back channels. Simple low-rises (i.e. L* L-H%) are usually associated with backward-looking communicative functions, differing from a low-onset high-rise (L* H-H%). Fifty-six per cent of simple high-rises (L* H-H%) are floor-holding. The proportion increases to 68% when including expanded-range complex rises. This is likely a result of speakers wanting to confirm frequently with the other participant. In expanded-range fall-rises (H*+L H-H%), the turning point of the rising portion can occur very late in a nuclear accented word that is also intonational phrase-final, resulting in a very rapid final rise, although this has yet to be verified experimentally (Fletcher et al., 2005).

The distinction between statement rises and question rises has inspired numerous investigations and discussions. Ritchart and Arvaniti (2014) found that the size of a rise is related to an utterance’s function: floor-holding statements have twice the pitch range of non-floor-holding ones. Apart from pitch ranges, rise alignments differed for question and statement rises: question rises begin earlier, relative to the accented syllable, than statement rises. The F0 endpoints for question and statement high-rises were not differentiated consistently, but the F0 start points were often distinct (Fletcher & Harrington, 2001). The ToBI transcriptions of the two utterance types can be quite different: the question rise is often labelled H* H-H%, while the statement rising terminal is usually labelled L* H-H%, which has a much lower start. Technically, therefore, these two kinds of rising intonation are not phonetically
identical. In recent comparisons of New Zealand and Australian English, statement rises in Australian English were realised using a wider pitch range than question rises, which is a unique characteristic that differentiates Australian English from the rising tones in other English varieties (Warren & Fletcher, 2016). Some further variation with the statement high-rise has been identified since Fletcher and Harrington (2001). Fletcher and Loakes (2006) found that many of the L* H-H% statement rises are in fact part of compound fall-rise tunes, which may account for the high incidence of low-onset statement rises in the 2001 study. It is later suggested that L* H-H% can also occur in some questions (McGregor & Palthorpe, 2008). Interestingly, none of these studies (based on map-task dialogues) found evidence that female speakers use more high-rises than males, contradicting Warren and Britain’s (2010) findings for New Zealand English.

Australian English speakers differentiate statement and question rises by the use of higher pitch accents; that is, higher starting points for the rise on questions than on statements (Fletcher & Harrington, 2001). A more recent perception study (Fletcher & Loakes, 2010) revealed that Australian speakers categorise more high-rise (L* H-H%, H* H-H%) intonation patterns as questions, while more statement responses are associated with L* L-H%. However, within the two high-rises, participants are more confident identifying short sentences, with H* H-H% interpreted as question intonation. Even in longer sentences, L* H-H% receives more statement than question responses. These results indicate that the distinction between L* H-H% and H* H-H% is quite salient. H* H-H% is most commonly associated with questions, while the other two rising tones are more commonly associated with statements.
2.4 Comparison between Lexical Tones and Intonation

As stated in Beckman and Venditti (2010), both tone and intonation ‘[refer] to patterned variation in voiced source pitch that serves to contrast and to organise words and larger utterances’ (p. 1). The functions of lexical tones and intonation are different: for tones, the contrastive function of F0 works at the lexical level, while intonation works at the post-lexical level. The change of lexical tone instigates the change of word meaning. Although intonation does not change word meaning, it constitutes part of the meaning of the whole utterance. English intonation interferes with both the perception and production of Mandarin tones (Chen, 1997). English speakers from this study replaced many tones with mid-level tones. This is explained as the transfer of English intonation patterns and the result of a smaller pitch range in English. It is proposed that for English speakers, both level tones and falling tones are easier to produce than rising tones. Within level tones, mid-level tones are easier to produce compared to high- and low-level tones. This hierarchical difficulty aligns with Li and Thompson (1977). This is also confirmed by comparing English intonation patterns with Mandarin tones and analysing common tonal errors (Gui, 2003). To produce a rising tone, a greater physiological effort is required (Ohala & Ewan, 1973); fewer occurrences of low-high sequences exist in languages compared to high-low sequences (Hyman, 1978; Hyman & Schuh, 1974).

Another difference between English and Mandarin is the realisation of stress. In English, the realisation of stress relies largely on the acoustic correlates of average F0, intensity, syllable duration and vowel quality (Zhang, Nissen & Francis, 2008). In Mandarin, the realisation of stress relies largely on expanded pitch range, lengthened duration and greater intensity (Shen, 1990). Every heavy syllable has a lexical tone (or pitch accent) in Chinese but not in English (Duanmu, 2013).
Interestingly, it has been found that words are often lengthened for emphasis in the production of English by Mandarin speakers (Schack, 2000). This lengthening could imply that English speakers possess similar transfers when speaking Mandarin. Additionally, White (1981) suggests that English speakers hear the Mandarin high-level tone as stressed and the falling-rising one as unstressed or very weakly stressed. Apart from these differences, intonation contours are similar to the pitch patterns of lexical tones, with pitch movements from high to low and vice versa. All tone contours can thus possibly be traced in English intonation, but they are typically spread over more than one syllable. Lexical tone density is much higher in tone languages than average pitch accent density in English.

If we compare the language-specific ToBI systems for English with the systems available for Mandarin, and Cantonese (see Table 2.8), the Mandarin-specific ToBI has five additional tiers and Cantonese has two. In the break index column of Table 2.8, the numbers with brackets stand for the similar and different uses of numbers to represent prosodic structure: 0 indicates a weakened word boundary, 1 a phrase-medial word boundary, 3 a minor phrase boundary such as ‘ip’, and 4 a major phrase boundary such as ‘IP’ and 2 for mismatch.

On the tone tier, tones are either lexical; the head of a prosodic unit (such as pitch accent [marked by *]); or the boundary tone marking the edge of a prosodic unit, such as an AP, an ‘ip’, or an ‘IP’. For Mandarin and Cantonese, tones for marking pitch range information (e.g., %reset, %q-raise) are grouped together with the boundary tone of the highest prosodic group.

With respect to prosodic units, English maintains two distinct units: ‘ip’ and ‘IP’; only Mandarin has a breath group, while Cantonese possesses word and IP distinction. Unlike Cantonese, which uses numbers to indicate lexical tones,
Mandarin uses romansi (romanisation), a separate tier. Neither Mandarin nor Cantonese has pitch accent (a * tone), but both have smaller ‘intonational’ tone inventories, as the boundary tone occurs at the edge of the largest prosodic unit.

Table 2.8

Comparison of the Tones and Break Indices Systems for English, Mandarin and Cantonese

<table>
<thead>
<tr>
<th>Language</th>
<th>Types of tiers</th>
<th>Types of break indices (BI)</th>
<th>Types of tones on the tone tier</th>
<th>Prosodic units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2,3 (ip)</td>
<td>L-.H-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 (IP)</td>
<td>L%, H%</td>
<td></td>
<td>IP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>! (for H pitch accent), &lt;, &gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandarin</td>
<td>Romansi 0,1</td>
<td>L%, H%, %reset, %q-raise</td>
<td></td>
<td>Breath group</td>
</tr>
<tr>
<td>Syll</td>
<td>2 (minor group)</td>
<td>%e-prom, %compress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td>3 (major group)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandhi</td>
<td>4 (breath group)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>5 (prosodic group)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td>Syllable 0,1</td>
<td>Lexical tones (55,33,22,335,</td>
<td></td>
<td>Wd</td>
</tr>
<tr>
<td>Foot</td>
<td>2 (‘IP’)*</td>
<td>223,221,553)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>L%, H%, H:%, HL%, %, fi</td>
<td></td>
<td>IP</td>
</tr>
</tbody>
</table>

Source: Jun (2006)

As mentioned earlier (Sections 2.2.1, 2.2.2 and 2.3.1), with respect to prominence, both English and Mandarin have lexical stress but Cantonese does not; however, both Cantonese and Mandarin have lexical tone distinctions while English
does not. English has post-lexical pitch accent, while Mandarin exhibits a strong phonological association between stressed syllables and lexical tones. Mandarin has many more polysyllabic words; in Cantonese, most syllables are ‘potentially free-standing morphemes’ (Wong et al., 2005). There is no contrast between stressed syllables and reduced syllables. Unlike Mandarin, even Cantonese particles have lexical tones. Apart from lexical tones, Cantonese has a rich inventory of boundary tones, which can be added to the final lexical tone to indicate intonational boundaries.

The traditional declarative tone in English is the rising-falling (H* L-L%) intonation, where the fall starts after the tonic word. It is claimed that the falling Mandarin tone (T51) is phonetically similar to the sentence-final intonation in English (Hayes, 2011). Native English speakers also have the impression that the falling tone in Mandarin is the only ‘normal’ tone (Broselow, Hurtig & Ringen, 1987) and this tone is better imitated by non-native speakers, and particularly well-imitated by musicians (Gottfried, 2007). Chiang’s (1989) finding regarding the misuse of T51 on sentence-level words by English speakers supports the above proposal, which is a transfer source of the English intonation pattern. Broselow, Hurtig and Ringen (1987) suggest that the falling tone was perceived in a different way from the other three tones by native English speakers. The advantage of T51 was seen as a transfer of English intonation as was also the case when T51 was misperceived as T55. When English listeners hear the T51, they might take the latter falling part as the sentence-final intonation and the former part (which has the exact same F0 onset as T55) as T55 itself. This confusion has also been reported by Gottfried and Suiter (1997). As argued in a number of studies (Pierrehumbert, 1980; Pike, 1945; Trager & Smith, 2009), English intonation has its underlying form as H
and L tone targets. The contours in English intonation are interpolations between these tone targets. Gandour (1983) has further supported this hypothesis with a dissimilarity-rating task showing that English speakers rely more on pitch height than pitch contour.

The differences and similarities between intonation and lexical tones are obvious and have a profound influence on both the perception and production of each system. The intonation contours are typically realised over a wider domain than is the case for lexical tones. Intonation is derived from the pragmatic system of English, while lexical tones are projected from the lexicon in Mandarin/Cantonese. The English prosodic system might still be of use to native speakers acquiring a tone language, although some unconscious transfer may have a negative influence. As White (1981) suggests, the tone mistakes that English speakers make during production do not randomly replace one tone for another, but rather occur during the L1 transfer of their intonation system. English speakers might use their intonation system as a ‘filter’ when perceiving and producing lexical tones.

2.5 Summary

This chapter has briefly reviewed a few key aspects regarding prosody: prosodic typology, transcription of prosody, and pitch perception and production. It has also focused on the two different linguistic uses of pitch: tone and intonation, also examining tone and intonation in the three relevant languages (Cantonese, Mandarin and Australian English). The chapter compared the three prosodic systems and discussed the possibility of prosodic transfer between these three languages. Chapter 3 examines the previous literature on tone perception and production by speakers from tone and non-tone language backgrounds.
Chapter 3: Tone Perception and Production

When investigating native and non-native speech perception and production, most research has focused primarily on the segmental speech sounds of languages, the vowels and the consonants. How and when listeners develop perceptual sensitivity to prosodic features such as stress, rhythm, tone and/or intonation when acquiring a new language is less well known, and as discussed in Chapter 2, languages differ greatly from each other prosodically. This is likely to result in substantial cross-language mismatches and learning challenges. How listeners perceive and produce non-native prosodic cues, especially across this typology, has garnered much attention in the recent literature, being the focal point of several interesting studies. Recent studies have even shown that prosodic errors are more prominent than individual segment errors in effective L2 communication (Anderson-Hsieh et al., 1992; Munro & Derwing, 1995; Trofimovich & Baker, 2006). Similarly, L2 prosodic acquisition can be shaped by the L1 system—native prosodic experience can both facilitate and hinder L2 learning, as reviewed in Section 2.1.3. For example, Dutch listeners are better at detecting English stressed syllables than native English speakers (Cutler et al., 2007), while Vietnamese speakers transfer their L1 tonal and syllable-timing features in perceiving and producing English stress and rhythm, which influences their acquisition of English negatively (Nguyêñ et al., 2008). Of all the prosodic features, tone is of particular interest. This is because—in the majority of languages—tones function similarly to phonemes in that they can change word meaning (as reviewed in Section 2.2). With this similarity, whether tone perception and production will be similar to the perception and production of segments is yet to be determined. The following sections will review tone perception and production.
separately, with the further distinction made between participants from tonal versus non-tonal backgrounds, speakers with versus without L2 tone learning experience.

3.1 Tone Perception

A significant amount of research has highlighted L1 and L2 speech perception and production on the segmental level, (e.g., consonant and vowel perception and production: Best, 1995; Best & Tyler, 2007; Flege, 1995; Polka, 1991; Strange, 1995; Strange, 2007). Much cross-language speech perception and production research has attempted to explain the influence that one’s L1 can have on L2 acquisition. This is further supported by extensive research indicating that linguistic experience has an influence on perception (Best, 1995; Best & Tyler, 2007; Lecumberri et al., 2010; Polka, 1991, 1995) and production (Flege, 1995; Flege et al., 1997; Flege et al., 1987; Munro & Bohn, 2007) of a new or second language at the segmental level.

In recent years, research on speech perception has also focused on prosodic features such as stress, prosody, intonation and tone. Lexical tone has received particular attention (Francis et al., 2008; Gandour et al., 2003; Hallé et al., 2004; Hao, 2011; So, 2006; So & Best, 2010; Xu, Gandour & Francis, 2006). This is of substantial practical value; tones have been reported by language learners to be quite difficult to perceive and produce in a second language (Francis et al., 2008; Qin & Mok, 2011; So, 2006). Incorrectly perceived tones can affect the understanding of speech significantly (Gandour et al., 2003; Hallé et al., 2004).

Cross-language tone perception research often relies on tone categorisation, identification and discrimination tasks to investigate how L2 listeners perceive linguistic tones; depending on the research question of each particular study, participants range from naïve listeners (listeners with no experience of a given L2) to
beginning learners of the L2 (see Hao, 2011; So & Best, 2010). Other researchers have employed language learning training sessions, conducting pre- and post-tests to assess tone learnability (e.g., Francis et al., 2008; So, 2006; Wayland & Guion, 2004).

3.1.1 Native tone perception

Tone perception by native speakers is typically investigated to provide a benchmark for perception by L2 speakers. In studies comparing perception by L1 and L2 speakers, discrimination accuracy results are commonly provided to support differences. For example, Hallé et al. (2004) found that the L1 tone-discrimination accuracy rate ranged from 84.4% to 94%, with a mean value of 88%. Importantly, this study provides evidence that tones are perceived categorically by L1 speakers, much in the same way as contrastive vowels and consonants by L1 speakers. L1 speakers also show increased sensitivity towards category boundaries, while L2 speakers (from a non-tone language background) fail to show a similar pattern of increased perceptual sensitivity at boundaries.

Research from a very different angle (i.e., neuropsychology) has also contributed to our understanding of L1 tone perception. A left-hemisphere advantage of processing linguistic information has been supported by a number of studies. Van Lancker and Fromkin (1973, 1978) determined that tone language speakers have a right-ear (left-hemisphere) advantage when distinguishing tones, while English speakers (non-tone) do not show this advantage. It is thus proposed that tone languages have a closer link between segmental structure and tone information. This explains the phenomenon that tone functions as near-phonemic information for L1 tone language speakers. Repp & Lin (1989) provide empirical data to investigate whether tone and non-tone language speakers show different integration of
segmental and tonal dimensions. Both groups show a processing asymmetry between consonants and tones while only Mandarin speakers show the asymmetry between vowels and tones. This might be related to the fact that vowels are the segments that carry tones, explaining why Mandarin speakers maintain this advantage over English speakers.

As discussed in Chapter 2, tone (along with most other prosodic features) is realised mainly on the nucleus of a syllable, and tone perception is thus intimately related to vowel perception, although vowel perception may be somewhat easier than tone perception. Indeed, L1 judgements are quicker and more accurate when words differ in vowels than when they differ in tones (Keung & Hoosain, 1979; Taft & Chen, 1992). Further, while tone information is quite important to language users, it takes longer for L1 speakers to process tonal information than segments. We know that F0 is the most salient cue to L1 tone perception compared to duration or relative amplitude (Abramson, 1962; Lin & Repp, 1992). Despite the importance of F0, however, it is also the case that tones are not well perceived when amplitude information is removed (Abramson, 1972). Indeed, added amplitude information can enhance perception accuracy. It is worthwhile noting that tones cannot be identified through differences in amplitude alone (Abramson, 1972; Whalen & Xu, 1992).

3.1.2 Non-native tone perception

The above section examined how L1 speakers perceive tones. The current section will discuss the way in which tones are perceived by L2 speakers, with and without L1 tone language backgrounds (L2 speakers with and without tone backgrounds will be introduced separately). The results from studies of L2 tone perception are complex and can seem contradictory. It is also clear that prior tone experience may either facilitate or interfere with L2 perception depending on the
specific discrepancies and similarities between the speaker/listener’s L1 and L2 tone systems. Under some conditions, non-tone language speakers can even outperform L1 tone language speakers on certain L2 tone contrasts. Indeed, tonal speakers do not perceive Mandarin T214, a falling-rising tone, better than do non-tonal speakers (Hao, 2011; So & Best, 2010). Additionally, the number of level tones in a listener’s L1 tone language may directly (Qin & Mok, 2011) or inversely (Chiao et al., 2011) influence the perception of level tones in an L2. However, perceptual assimilation has rarely been examined in listeners from a simpler tonal background, and it remains unclear how the L1 system will influence listeners’ perceptions of more complex L2 tones.

Some studies suggest that L2 tone acquisition is influenced by the phonological and phonetic constraints of the L2 tone system itself, regardless of L1 background (Wang, Behne, Jongman & Sereno, 2004). Such studies often claim that acoustically contrastive tones are acquired first, while tones with similar acoustic features are processed later or with greater difficulty. Other research claims that previous linguistic experience (L1 background) affects tone perception in a second language significantly, just as it does in phoneme perception (Hao, 2011; So, 2006; So & Best, 2010). Burnham et al. (2014) conclude that universal and language-specific factors work in tandem during L2 tone perception processes. These contrasting results highlight the need for further research in this field: a number of issues are still contentious and questions remain to be explored, in particular the influence of the L1 on L2 tonal learning.

### 3.1.2.1 Non-native tone perception by speakers of other tone languages

As discussed in Section 3.1.2, tonal experience can assist L2 tone perception, but the evidence is somewhat inconsistent. This inconsistency is highlighted by Lee
et al. (1996) who conclude that Cantonese speakers perform better than do English speakers in discriminating Mandarin tones. Interestingly, Mandarin speakers do not discriminate Cantonese tones better than do English speakers. The authors’ explanation for this is that Cantonese has more contrastive tones than does Mandarin and is thus more difficult to perceive for Mandarin speakers. This effect might be language-specific instead of universal. Indeed, So and Best (2010) argue that this evidence is not conclusive, as the Cantonese participants from Lee et al.’s (1996) study originated from Hong Kong, where they had extensive exposure to Mandarin that might account for the performance difference between Cantonese and English speakers. Leung (2008) however, presents similar results: their L1 Cantonese listeners also outperform English listeners in terms of Mandarin tone perception; but, again, those Cantonese participants had previous experience with the target Mandarin tones.

Another study by Wayland and Guion (2004) suggests that a tone language background has a positive influence on second language tone perception. Here, the authors found that tone language speakers (Mandarin) improve significantly in both discriminating and categorising Thai tones after L2 training, whereas non-tone speakers (English) show no significant improvement following training. This study is particularly interestingly, as the Mandarin- and English-speaking participants reached similar levels of discrimination and categorisation accuracy in the pre-test before training.

Despite the results reviewed above, other research suggests that having a tone language background does not always help perception in another tone language (Francis, Ciocca, Ma & Fenn, 2008; Hao, 2011; So, 2006; So & Best, 2010; Wang, 2006). For example, Wang (2006) found that tonal language speakers (Hmong)
performed less accurately than do pitch-accent language speakers do (Japanese) when perceiving Mandarin tones. Hmong has seven contrastive lexical tones while Mandarin has four. Due to the mismatch between the L1 and target tone inventories, the Hmong tonal inventory may have negatively affected Hmong listeners’ perceptual ability to discriminate Mandarin tones. These findings accord with those of So (2006), who investigated Mandarin tone identification by Cantonese and Japanese speakers. In that experiment, participants were tested three times: immediately after a brief familiarisation session for Mandarin tones, one or two days after training with auditory sessions, and one month after training. At first, the two groups of listeners were comparable to each other, and both groups showed significant progress after training. In addition, the A prime score (seen as a measure of perceptual sensitivity) indicated that the Japanese participants were more sensitive to Mandarin tones than were the Cantonese participants. In particular, this study examined error patterns and found that tones that were more similar to participants’ L1 prosodic inventory were more difficult to discriminate.

The results outlined above are also consistent with those of Hao (2011), who found that Cantonese listeners identify the T35-T214 pair poorly as they perceptually map the pair to one single Cantonese tone. In general, this study supported the idea that an L1 tone system interfered with L2 tone perception: Cantonese speakers identified and produce fewer Mandarin tones accurately compared to English speakers. However, according to the mapping task in which Cantonese listeners participated, not all error patterns could be explained by L1 linguistic experience. Other factors might have been working in tandem, making T35-T214 the most difficult pair for Cantonese speakers to perceive, even if they were mapped into two different native categories where good discrimination was expected.
An L2 tone disadvantage for L1 tone language speakers has also been found between Mandarin and Cantonese when Mandarin speakers perceive Cantonese. English listeners performed better than Mandarin speakers in both pre- and post-tests in a training study to identify Cantonese tones (Francis et al., 2008). A greater between-group difference was found after participants had undergone training, such that non-tone language speakers improved more than tone language speakers did. This finding contrasts with Wayland and Guion’s (2004) findings that English speakers did not improve significantly after training. The study also highlights significant group differences in the most difficult tone pairs: native Mandarin listeners primarily rely on F0 contours to perceive their native tones and pay more attention to direction rather than height, while English speakers have more difficulty with pitch contour. When two tones had the same contour pattern, they were quite difficult for Mandarin speakers to discriminate. Similar findings arose from Qin and Mok’s (2011) study, where even though Mandarin speakers were better at discriminating Cantonese tones in general than English and French speakers, they performed worse on discriminating the three Cantonese level tones.

Further, the number of level tones in the native tone system may exert an influence on L2 tone perception. Indeed, it is likely that having a more complex L1 tone system can enhance a listeners’ sensitivity towards phonetic distinctions (Bohn & Best, 2012; Zheng, Munhall & Johnsrude, 2010), despite Wang’s (2006) findings reported above. For example, Cantonese speakers, whose L1 has three level tones, have a greater sensitivity to both phonetic and phonological differences than Mandarin speakers, whose L1 has only one level tone (Zheng et al., 2010). In contrast, Chiao et al. (2011) found that the more level tones there are in one’s L1 tone system, the poorer is one’s ability to perceive L2 level tones. Vietnamese (a
tone system that involves one level tone) listeners and English (a non-tone system) listeners outperformed Taiwanese (a tone system with two level tones) listeners in discriminating the four level tones in Toura. This is an African Niger-Congo language, and it is argued that the Taiwanese listeners confused level Toura tones with those from their L1 Taiwanese system, leading to poor discrimination ability. The available evidence also suggests that Mandarin speakers outperform Cantonese speakers in discriminating Thai level tones (Burnham et al., 2014).

Interestingly, other studies have shown that having a tonal L1 can be both detrimental and beneficial for L2 tone perception: facilitation and interference from L1 tone experience can occur simultaneously (Burnham et al., 2014; So & Best, 2010). A three-group perception study (So & Best, 2010) concluded that L1 prosodic structure does not always facilitate the categorisation of L2 tones and that the categorisation pattern may be language-specific rather than universal. In some cases (for Mandarin T35 and T51), Cantonese listeners perform best, whereas on T55, Japanese listeners outperform Cantonese listeners; on T214, Cantonese listeners have the poorest levels of performance accuracy. Further, Cantonese listeners display a similar sensitivity to pitch height and pitch movement, as Cantonese tones are (also) distinctive in these features, while Japanese listeners focus on pitch variations to differentiate lexical meaning, as Japanese is a pitch-accented language. This suggests that, just as the L1 prosodic systems are language-specific, the observed error patterns are also language-specific. They are dependent on the mismatch between L2 and L1 systems and the differences in the phonetic realisations within these systems, although the difficulty of some tone pairs with similar phonetic features is independent of language. Similarly, tonal experience does help Mandarin and Cantonese listeners outperform English listeners in perceiving Thai tones, but these
two groups have no advantage when compared to Swedish (a pitch-accented language) listeners under auditory-only (only heard the tones) and auditory-visual conditions (saw the speaker and heard the tones at the same time). Conversely, English speakers discriminate Thai tones more effectively than both tone and pitch-accent language speakers when provided with visual-only information (could only see the speaker, not hear the tones).

3.1.2.2 Non-native tone perception by speakers of non-tone languages

As indicated above in Section 3.1.2.1, L1 speakers of non-tonal languages may perform differently to L1 speakers of tonal languages, who may be able to recruit their L1 tone inventory for L2 tone perception when acquiring an L2 tonal language. As there is no L1 lexical tone system for listeners to map onto, the ways in which these participants perceive L2 tones present an interesting puzzle. As reviewed in Chapter 2, both tone and intonation are cued by F0: tone is lexical and intonation is post-lexical, and if non-tone speakers can perceive intonation according to their L1 prosodic system, it is likely that non-tone language speakers without L1 tone experience might recruit aspects of their L1 prosodic system in order to perceive tones. The influence of L1 prosodic systems on the perception of intonation contours is supported by several studies (Grabe, Lang & Zhao, 2003; He et al., 2012; Huang et al., 2007; Ulbritch, 2008). The inference is that non-tone language speakers may perceive tones in the same way as they perceive intonation (see Section 2.4). However, it is also likely that L2 listeners cannot perceive tones categorically (as L1 tone language speakers do), but rather in a psychoacoustical way (Hallé et al., 2004).

Different perceptual cues are used by tone and non-tone language speakers. Gandour (1983, 1984) shows that English speakers attend to pitch height information when perceiving L2 tones; Cantonese listeners attend to pitch height as well as pitch
contour information. This difference in strategy may result in English speakers’ difficulties in perceiving tones with similar pitch height but different contours. A number of studies support the position that both L1 tonal language speakers and L1 non-tonal language speakers share the same level of confusion with L2 tones, and suggest that L2 tone perception is difficult regardless of listeners’ L1 background (tonal or non-tonal) (Hao, 2011; Qin & Mok, 2011; So & Best, 2010). So and Best (2010) suggest that some Mandarin tone pairs (T55–T35, T55–T51 and T35–T214) are difficult for Cantonese, Japanese and English listeners, given that these tone pairs have similar phonetic features. Similar results are documented for speakers of English (Hao, 2011) and German (Ding, Hoffmann & Jokisch, 201). Studies examining the perception of Cantonese tones (Qin & Mok, 2011) also reveal similar patterns with the two rising tones T23 and T25, such that Mandarin, English and French listeners have trouble discriminating between these tones due to their high degree of phonetic similarity. Interestingly, this pair is the last of the Cantonese tones to be acquired by L1 children, and even Cantonese-speaking adults report difficulties in discriminating between these tones (Mok, Zuo & Wong, 2013; To, Cheung & McLeod, 2013), suggesting that the high degree of phonetic overlap may be problematic even for native listeners.

At this stage, very little is known about whether English speakers rely on their L1 intonation system to help with the discrimination or perception of tones as non-speech patterns. A recent study by So and Best (2011) supports this notion (that a L2 prosodic system will be assimilated to the L1 one in L2 perception). In this study, English and French speakers categorised Mandarin tones into written ‘flat pitch, exclamation, question and statement’ intonation types1. The study shows that

---
1 ‘Flat pitch, exclamation, question and statement’ were the descriptive tags presented in their study
both English and French speakers categorically perceive Mandarin tones using their own L1 intonation system. However, it is very difficult to assess what the four provided intonation tags represent—they are quite ambiguous and may have influenced the participants’ performance and the associations between intonation and tone. However, Japanese speakers (a pitch-accented language) can assimilate Mandarin tones to their pitch-accent inventory using the same written descriptions (So, 2010).

3.2 Tone Production

L2 tone production has received an increasing amount of attention in the past few years, as the focus of cross-language research has gradually extended to the prosodic domain. However, ample evidence is still lacking regarding whether production is moulded by previous linguistic experiences to the same extent as has been argued for the domain of speech perception.

3.2.1 Native tone production

In general, L1 speakers of a given tonal language, who do not suffer from any hearing loss or impairment, achieve near-ceiling accuracy in the production of their L1 tones. L1 tone language speakers acquire their native tones at an early age and make very few tonal errors. Indeed, infants from tone language backgrounds use pitch to indicate different meanings as early as eight months of age, and prior to producing their first lexical word at around 10 to 12 months of age (for typically developing children [Clumeck, 1980]). Tone perception and production has even been argued to start before segmental acquisition (Burnham & Francis, 1997). The time-course of the emergence of different tones may vary between the many tone languages. For example, Li and Thompson (1977) argue that Mandarin infants produce falling tones earlier than rising tones, as falling tones require less
physiological effort. In contrast, Thai infants are reported to produce rising tones earlier than falling ones (Tuaycharoem, 1979). This suggests that tone production varies significantly across languages. Other research into adult native tone production by Mandarin speakers has found that the four Mandarin tones are produced in different tonal spaces but with some degree of overlap (Yang, 2014). T214 and T51 have the greatest degree of overlap. A notable overlap between the neutral tone and the other four tones has also been identified.

As pitch contours and F0 are more salient cues than duration for tone perception, variations in tone duration have often been ignored or under-investigated, despite the fact that systematic relationships are often found between tone height and duration. For example, Abramson (1962) examined native Thai tone production and found that the mid and low tones were longer than the high tones. Similar systematic relationships between tone contour and tone duration have also been observed, such as in the work of Earle (1975). Earle shows that in Vietnamese, hto (the mid-falling-rising) is the longest, followed sequentially by ngang (mid-level), huyel (low-falling), sal (mid-rising), ngâ (glottalised mid-rising) and nisi (mid-falling). Analyses of Mandarin tones by Dreher and Lee (1968), Chuang and Hiki (1972) and Howie (1974) further suggest that the falling-rising tone (dipping) is the longest. However, these authors disagree about which is the shortest tone—either T51 (Dreher & Lee, 1968) or T55 (Howie, 1974). Similarly, no agreement exists regarding the longest. Fok (1974) and Kong (1987) compared the duration of Cantonese tones with variations of the shortest and longest tones. Fok (1974) found that the low-rising tone (T23) has the longest duration and the high-level tone (T55) has the shortest. In contrast, Kong (1987) found that the high-rising tone was the longest, with the high-
and low-level tones the shortest. However, some agreement exists supporting the mid-level tone as the longest of the level tones.

To summarise, a common trait shared by Mandarin and Cantonese is that rising tones have the longest duration and falling tones the shortest. If there are two rising tones, one will be longer. A positive relationship between an upward F0 and longer duration has also been found (Ohala & Ewan, 1973). This was later argued as universal by Gandour (1977): rising tones have a longer duration than falling ones, and level tones with higher frequency have a shorter duration. However, the latter conclusion contradicts Kong’s (1987) findings with Cantonese tones: that the mid-level tone has a longer duration than the low-level one. Thus, the relationship between duration and F0 might not be simply linear. All the studies mentioned above have examined the duration in tone production by L1 speakers.

3.2.2 Non-native tone production

The following section will present evidence on production by non-native speakers. Unsurprisingly, this evidence is likely to reveal that L1 speakers outperform L2 speakers in tone production. However, little research has been conducted that compares L2 tone production by tonal and non-tonal language speakers, with few clues to determine how L1 tone experience influences L2 tone production. In the following sections, I will outline L2 tone production by speakers from tone and non-tone languages respectively.

3.2.2.1 Non-native tone production by speakers of other tone languages

Relatively little research has been conducted on L2 tone production by L1 speakers of other tone languages. As discussed in Section 3.1, it is reasonably predictable that L2 tone production will be moulded by the L1 tone system, similar to the findings of segmental research. Most studies favour a positive influence from the
speaker’s native language: having a tone language background will be advantageous when producing a new tone language. Leung (2008) found that tone language speakers produce L2 tones significantly better than do non-tone L2 language speakers; this was determined by investigating Mandarin tone production by Cantonese learners of Mandarin versus English speakers. However, it should be noted that the tone language speakers in this study had all learned Mandarin, while the English speakers had no prior experience with lexical tones. Further, as with perception studies, the difference between L1 and L2 systems determined the difficulty of L2 tone production. Although the position that tone language speakers produce L2 tones better is not conclusive, this study provides a clear and precise example of how studies of tone production can be compared with perception results. However, this study only investigated how Cantonese speakers assimilate Mandarin tones: how tones are mapped from a smaller tone inventory (Mandarin) to a larger one (Cantonese). More research is required to examine the perceptual assimilation of native tones in the reverse direction to understand tone production more effectively.

Negative influence from the L1 is present in L2 production as well. For example, Hao (2011) suggests that Cantonese speakers make more errors than do English speakers in both mimicry and reading T35 and T51. This negative influence aligns with a perception study conducted with the same participants. Nevertheless, the most difficult tone pair found was T35–T214.

3.2.2.2 Non-native tone production by speakers of non-tone languages

Few studies have examined non-native tone production by naïve listeners from a non-tone language background, as most production studies examining non-tone language speakers involve learners of the target tone language. However, it is clear that even a small amount of experience with the target language is likely to
influence results, as researchers have found that tone production improves with experience (Flege, Takagi & Mann, 1997; He et al., 2008). For example, tone error patterns differ between early and advanced Mandarin learners (with L1 English), and the errors that early learners make are more clearly related to their L1 (Shen, 1989) than those made by advanced learners, whose errors fall evenly into two categories: tonal register errors (too high or too low) and tonal contour errors. Interestingly, the errors are distributed evenly among all four Mandarin tones (Miracle, 1989). Another study with intermediate learners further suggests that T55 and T51 are generally easier to produce than T35 and T214 for L2 speakers (Yang, 2014). Tones with a lower register are more difficult to produce than are those with a higher register. The errors mostly stem from a register at the start or endpoint, except for T35, which has more contour errors. Yang (2014) proposes that falling intonation is the stress marker in English; thus, English speakers tend to replace the rising tone with falling tones. The production maps show that unlike native Mandarin speakers who produce tones in three main categories, English speakers can only differentiate either one or two categories. These non-native speakers could not produce the pitch differences required in Mandarin tones.

The above review clearly indicates that more research is required in investigating how speakers with different L1 prosodic backgrounds produce non-native tones and whether there is a tone language advantage in production.

3.3 The Link between Tone Perception and Production

The link between tone perception and production has long been a conundrum, with researchers tackling the issue from different angles. It has been of great interest to researchers in the field of children’s development, with most studies determining that children first establish a perceptual category and then attempt to match their
output to this category. Research from this domain supports the assumption that phonemic perception precedes production (Edwards, 1974; Menyuk & Anderson, 1969). The supporting evidence for this link is multifaceted: children who are deafened pre-lingually suffer from severe speech loss if they are not implanted with a hearing device promptly upon diagnosis (Geers, Nicholas & Sedey, 2003; Schauwers, Gillis, Daemers, De Beukelaer & Govaerts, 2004); adults undergoing hearing loss will lose control of F0 and intensity (Cowie, Douglas-Cowie & Kerr, 1982). The importance of feedback (auditory perception) in ensuring production accuracy is supported by ample clinical research (Fukawa, Yoshioka, Ozawa & Yoshida, 1988; MacKay, 1968; Siegel, Schork, Pick & Garber, 1982). Speakers accommodate their speech style rapidly so it is similar acoustically to their auditory feedback. A study monitoring brain activity with functional magnetic resonance imaging (fMRI) during both production and perception found similar functional activity, indicating the existence of a self-monitoring system and providing neuropsychological evidence for the link between perception and production (Zheng et al., 2010).

Studies that have investigated the link between perception and production vary considerably in their methodology. Methods include different tasks among different types of populations, including naïve listeners, learners, bilinguals and listeners with cochlear implants. Empirical studies usually take four positions: 1) how shifts in perception lead to shifts in production; 2) how perceptual training improves perception and production; 3) how adding a production component instigates a change of perception recalibration; and 4) how perception performance is related to production performance.
This question has also been long investigated by empirical research into L2 perception and production. Some of these studies have also explored this topic in the prosodic domain, which is relevant to the present thesis. In general, the results from previous studies have not given a clear picture; several show no correlation between participants’ ability to produce and perceive a given contrast, segment or consonant sequence (e.g., Darcy, Park & Yang, 2011; de Jong, Hao & Park, 2009; Golestani & Pallier, 2007; Kabak & Idsardi, 2007; Sheldon & Strange, 1982; Shin & Iverson, 2011). Conversely, other studies have reported correlations between production and perception accuracy (e.g., Flege, 1993, 1995; Flege et al., 1997; Rochet, 1995).

Bent (2005) investigated the perception and production of Mandarin tones by naïve English speakers and found no direct link between the two sets of abilities. However, some evidence of perception leading production was found:

1. Perception scores were generally quite high, while some difficulty was present in production tasks. This suggests that even with perceptually sensitive speakers, production ability sometimes lags. This is consistent with research results from the segmental domain.

2. The most difficult monosyllabic pair in production was still perceived well, indicating that perception precedes production.

3. The most difficult trisyllabic pair in production was the same pair that participants had most trouble with in perception, showing that without correctly perceiving a contrast, accuracy in production is very unlikely.

The finding that no link exists between perception and production with naïve non-tone speakers is not unique: several studies have found similar results (de Jong et al., 2009; DeKeyser & Sokalski, 1996; Yang, 2014) or partial correlation between non-native tone perception and production (Hattori & Iverson, 2010). However, this
is contradicted by Xu et al.’s (2011) research, which shows that tone perception and production performance is highly correlated—perhaps because a tone must be perceived accurately before it can be produced accurately. Wang, Jongman and Sereno (2003) indicate that a correlation is present after a short period of training.

3.4 Summary

This chapter has reviewed previous studies that examined the perception and production of lexical tones. The literature has been discussed separately according to categories and groups: by L1 and L2 speakers (further grouped into tone and non-tone language speakers). The last section reviewed the link between perception and production. A considerable number of studies have explored the influence of a speaker’s native language system, but no clear conclusion regarding this has been reached at this point. Controversial results have been obtained when discussing the relationship between perception and production within the prosodic domain. This chapter has provided a background to, basis for and explanation of the necessity for this current investigation. Chapter 4 will introduce the two relevant speech modals that explain L2 tone perception and production, and the potential link between the two modalities. A thesis overview will also be provided in the latter part of Chapter 4.
Chapter 4: Theoretical Models and Thesis Overview

This section will first briefly introduce the development of different second language perception theories and then review the two most relevant theoretical models—PAM (Best, 1995) SLM (Flege, 1995). Following the presentation of each of the two models, I will propose expansions to each of them in the following ways: PAM is extended in order for the model can provide separate predictions for L2 tone perception by non-native speakers from tone and non-tone backgrounds. The proposed extension also endeavours to draw predictions for L2 tone production. In turn, SLM is extended particularly to account for the relationship between tone perception and production.

This thesis builds on decades of research that has clearly demonstrated the importance of cross-language research. Only by adopting a cross-language perspective can we ascertain the language-dependent and universal traits behind speech perception and production. More recent theories and models attempting to explain and unveil the ‘magic’ interactions between language and the human mind have embraced this knowledge. These have been developed as models that compare the perception and production of speech sounds from second languages. These models include PAM (Best, 1994, 1995), SLM (Flege, 1986, 1990, 1995; cf. Guion et al., 2000) and Kuhl’s native language magnet model (NLM) (Grieser & Kuhl, 1989; Iverson & Kuhl, 1996; Kuhl, 1991, 1992; Kuhl, Williams, Lacorda, Stevens & Lindblom, 1992). All authors have noted that the frequently observed patterns of difficulty with foreign language or L2 phoneme perception are related to the listener’s L1 speech system.

The difference between these frameworks is how they see the relationship between the new speech and L1 systems. NLM focuses primarily on first language
development, and studies within the NLM framework typically observe the
behaviour and development of young children; they do not study cross-language
perception and production studies with adult populations, as in the present thesis.
PAM and SLM share a focus on the differences in speech perception by both naïve
and experienced L2 listeners respectively; as such, these models make direct
predictions for performance across the lifespan. Both PAM and SLM are extremely
relevant to the current study, as the participants involved in this research will include
naïve and experienced adult L2 listeners.

Speech perception research and models, such as PAM and SLM, continue to
excite ongoing and often intense theoretical debate. For instance, while most
researchers posit that speech perception and non-speech perception is handled by the
same auditory processes (e.g., cf. theoretical overview in Best [1995]), others suggest
that speech perception involves a specialised system not employed in the perception
of non-speech sounds (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967;
Liberman & Mattingly, 1989). In addition to the debate about whether speech
perception relies on general or specialised perceptual systems, the question of
whether perceptual mechanisms operate on acoustic or articulatory information is
also controversial. However, new areas of focus, such as tone perception and
production, will undoubtedly incite further debate. Although some studies have
investigated tone perception and production and have tried to extend these models to
account for prosodic features, the models still need further development and rigorous
testing to account fully and satisfyingly for this aspect of language use.
4.1 The Perceptual Assimilation Model

4.1.1 Review of the perceptual assimilation model

PAM’s key claim, as formulated by Best (1994; 1995), is that perceptual limitations determine the difficulty that L2 learners have in learning an L2. PAM proposes that—depending on the degree of similarity and discrepancy between L1 and L2 phonemic systems—L2 learners classify L2 phones into existing L1 categories. PAM is the only model that provides specific predictions about listeners’ L2 discrimination and assimilation. It does so through formulating hypotheses about how L2 phones match to L1 phonemic categories, which makes it easier to predict a clear discrimination pattern. PAM proposes that both the L1 abstract phonological and the language-specific phonetic realisations of the phonemes determine listeners’ assimilation of L2 systems. These assimilation patterns, detailed below, form the foundation of a further set of different types of L2 contrasts, displayed in Figure 4.1.

As is clear from the figure, PAM proposes three possible ways in which L2 phones can be categorised. First, an L2 sound can be either a speech phone or a non-speech sound. If it is categorised as a speech sound, it shares some commonalities with the L1 sound system, which can further be classified into categorised or uncategorised. A categorised consonant or vowel of the L2 phoneme will be assimilated into an L1 category, and will have assimilation goodness from poor to excellent. An uncategorised exemplar has some similarity with more than one phoneme but does not resemble any single phoneme. If an L2 sound is quite different from the L1 phonemes, it will be classified as a non-speech sound. The L1 phoneme inventory thus affects the way that L2 phones are perceived, depending on the assimilation pattern of a given L2 phone to the native phoneme inventory. A L2 phone can be perceived in a near-native fashion, in a moderately ‘accented’ fashion, or in a highly
‘accented’ fashion (Best, 1994a, 1995). Indeed, according to Best (1995), when two L2 phones are separated by an L1 phone boundary, L1 phonology can help L2 discrimination. When both phones are similar to the same L1 phone, L1 phonology should hinder discrimination. However, when L2 sounds are perceived as non-speech ones, they are neither aided nor hindered by L1 phonology. For example, Best, McRoberts and Sithole (1988) tested English speakers’ perception of Zulu clicks. Instead of mapping the clicks to an L1 category, English participants perceived them as non-speech. Consistent with PAM’s predictions, these non-assimilable contrasts (Zulu clicks) were discriminated very well—with goodness from good to very good—compared with the other speaker group from a click language background whose click inventory differed from Zulu (Best et al., 1988).

Source: Best (1995)

*Figure 4.1* Categorisation of L2 sounds by PAM.
The essence of PAM therefore is phone pairs—this not only provides a clear definition of possible discrimination contrasts, but also specific predictions about the discrimination difficulty for each pair. The L2 contrasts postulated by PAM are summarised as follows:

1. **Two-Category (TC):** members of the L2 contrast assimilate to two different native categories. If a sound contrast is categorised as TC, the contrasts should be phonemic in both L1 and L2. Hence, this contrast will be easy to discriminate.

2. **Category-Goodness (CG):** each member of the L2 contrast assimilates to the same L1 category with one of the members being more deviant from the L1 sound than the other. The extent to which an L2 learner can discriminate sound contrasts from a CG group depends on the distance of the two members from the L1 category. If these two sounds differ greatly from each other as well as the L1 sound, it will still be possible to discriminate them. However, if they are both close to the L1 category, discrimination will be more difficult.

3. **Single-Category (SC):** both L2 phones assimilate to one phoneme in the L1 category, and both are equally deviant from the L1 sound. Considering sounds from the SC group, a discrimination task will be quite difficult, as the two sounds are equally close to the same L1 category.

4. **Uncategorisable-Categorisable (UC):** one of the contrast members is uncategorisable, while the other is categorisable. As the two phonemes are quite different, the discrimination should be quite good.

5. **Uncategorisable-Uncategorisable (UU):** both members are uncategorised as defined above. The discrimination of this contrast should
have little influence from the native system and the discrimination accuracy should be fair to good, depending on the distance between the L2 phonemes and the closest L1 ones.

6. **Non-assimilable (NA)**: both members have great discrepancy with the L1 phone inventory and are categorised into a non-speech category. Thus, NA (both non-speech sounds) contrasts should have an accuracy of good to excellent, depending on the perceived difference between these two sounds (Best, 1995; So & Best, 2014).

Indeed, UU and UC can be further categorised into different subtypes with clear predictions for discriminability (So & Best, 2014): when both phonemes are perceived as similar to (or categorised into) the same set of L1 categories, the contrast is labelled ‘same set’(s). When a partial overlap exists in the perceived similarity to L1 categories, the contrast is labelled ‘partial overlap’ (o). When there is no perceived overlap, the contrast is labelled ‘no overlap’ (no). Contrasts with no overlap are predicted as easy to discriminate, while partial overlap is more difficult and the same set discrimination is the most difficult. To conclude, PAM predicts that the discrimination of a given contrast will be poor when two L2 phones are categorised and assimilated into one L1 phonemic category. In contrast, the outcome will be excellent when these two phonemes are assimilated into two different L1 categories.

In terms of the relationship between perception and production, PAM assumes that perception and production relies on the same mechanism but working from opposite ends, as PAM has its basis in articulatory phonology. It is suggested that perception and production share the same gestural representations in nature; thus, a direct link can be expected. Best and Tyler (2007) extended PAM to L2
acquisition (PAM-L2) and made a series of predictions about the particular aspects that changed L2 perception and production. PAM-L2 predicts a lag between perception and production, with perception coming earlier, as speakers will have had to perceive a sound from themselves or others before producing it. Thus, a change of perception will prompt a change of production. Such a pattern is commensurate with findings that indicate perceptual learning helps improve production (Akahane-Yamada, Strange, Downs-Pruitt & Masuda, 1998; Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Wang et al., 2003a). It can also be inferred that production errors have their roots in perception: without perceiving a sound accurately, one has little chance of producing it correctly.

### 4.1.2 Extending the perceptual assimilation model and perceptual assimilation model-suprasegmental to tone perception and production

The extension of PAM to prosodic systems (PAM-suprasegmental [PAM-S]) (So & Best, 2014) proposes similar assimilation patterns to those at the segmental level (L2 phones can be categorised/uncategorised, and depending on the discrepancy between L1 and L2 categories, categorised L2 phone pairs can further be grouped into SC, TC or CG), based on previous findings from prosodic studies. According to PAM-S, a given L2 prosodic realisation can be either categorised or uncategorised. Similar to the segmental perception listed in Section 4.1.1, when both prosodic realisations in a contrast are categorised, the contrasts could be SC (when two fall into the same L1 category), TC (when two fall into two different categories), or CG (when two fall into one category but one fits better). This is based on the discrepancy between the L2 and L1 prosodic sounds. The discrimination of a given contrast will be poor when two L2 sounds are categorised and assimilated into one L1 prosodic category. In contrast, the outcome is excellent when these two phonemes
are assimilated into two different L1 categories. With this PAM-S model, for the first time PAM provides criteria for deciding an L2 phone to be uncategorised and detailed predictions for contrasts involving an uncategorised phone. As they suggest, to be counted as categorised, an L2 phone must satisfy two criteria: the chosen category should have significantly more choices than both chance level and other categories. When an L2 phone is not assimilated into a certain L1 category it is seen as uncategorised. This suggests that an uncategorised L2 phone can be assimilated into no L1 category, or two competing L1 categories. Following this argument, UC or even UC pairs can sometimes have similar assimilation patterns. Depending on the chosen L1 categories, UU and UC pairs can be further categorised into same set (when two L2 phones have common chosen L1 categories), no overlap (when two L2 phones have no common chosen categories) and partial overlap (when two L2 phones have partially common chosen categories). Similarly, pairs with no overlap would be the easiest contrast, while partial overlap will be more difficult and the same set will be the most difficult. This is a very meaningful add-on: it complements previous PAM predictions in that it further defines the differences between categorised/uncategorised and further categorising UU and UC pairs.

PAM provides a clear framework from within which it is possible to make predictions about the relationship between perception and production, and also relate perceptual and production difficulties. As PAM predicts that perception leads production, and that they are intimately connected, listeners’ perception and production should be related—if a learner perceives L2 tones well, he or she should also be able to produce them reasonably accurately. Moreover, the errors one makes in production should be directly relatable to perception errors. For example, if a learner misperceives a particular tone, he or she should also have problems when
producing it. This extension of PAM to tone production is proposed according to the aspect of PAM-L2 that involves discussion of another model: SLM. SLM, which will be introduced in detail below, provides excellent predictions for speech and tone production.

Traditionally, a distinction is made between phonetic and phonological assimilation within segmental perception studies. Here, phonetic assimilation occurs when one L2 sound is perceived as a phonetic equivalent in the L1, and phonological assimilation occurs when the same phonemic status is shared by the L1 and L2 categories. However, few studies have examined this issue when extending these models to the prosodic domain. One of the major differences between PAM and SLM lies here as well: SLM proposes that category assimilation occurs only at the phonetic level. By contrast, PAM posits that both phonetic and phonological levels are possible. However, according to the SLM, assimilation can occur between dissimilar L1 and L2 phones as well as between similar categories.

Within the PAM framework, phonetic assimilation occurs when listeners rely on acoustic similarities to assimilate an L2 sound to their L1 system. L2 phonetic categories are perceived as similar with L1 phonetic categories, based on acoustic or gestural properties. Evidence from Cantonese speakers’ perception of Mandarin tones supports the effects of acoustic similarities on tone perception (Leung 2008; So, 2012; So & Best, 2010) (e.g., high-level tone and high-rising tone). Wu et al. (2014) have confirmed with Thai and Mandarin speakers that listeners assimilate L2 tones to their L1 tone category according to the most similar acoustic properties, such as F0 height or F0 contours, and sometimes even partial phonetic features. This is explained as listeners being forced to make a choice even when they can find no better match.
To date, only a few studies have tested the predictions of PAM extensions to tone perception (Chiao et al., 2011; Hu, 2011; Leung, 2008; So & Best, 2008; So, 2010; So & Best, 2011). Results from these studies confirm that L2 tones are mapped onto tone language speakers’ L1 tone categories, which works in a similar way to segmental features. Some studies support PAM’s predictions regarding extending findings to the tone domain. So and Best (2010) found that the discrimination patterns of Cantonese listeners categorising Mandarin tones supports PAM’s predictions that TC discrimination is better than CG discrimination, and further that the accuracy of UC discrimination is likely to exhibit significant within-group differences. To Cantonese listeners, three out of four Mandarin tones are ‘categorised’. The Mandarin level tone (T55) was assimilated to Cantonese high-level (T55) (i.e., they are a CG pair). Mandarin falling (T51) was assimilated to Cantonese high-level (T55). Mandarin rising (T35) was assimilated to Cantonese high-rising (T25), while Mandarin falling-rising (T214) did not fall into any certain Mandarin category and was thus seen as uncategorised by Cantonese listeners. This is quite relevant to the current study, where UC can be further grouped and three different patterns are revealed (see details in Chapter 5). The predictions from PAM extensions are also supported by experienced L2 speakers. Even with L2 experience, the influence of L1 properties is still difficult to eliminate. Cantonese speakers who were Mandarin learners discriminated Mandarin tones as well as did L1 Mandarin speakers. However, the two speaker groups showed different error patterns: Cantonese speakers with Mandarin experience perceived T51 as the most difficult, while Mandarin speakers found T35 more difficult (Leung, 2008).

However, conflicting results have arisen in different studies, even sometimes occurring with similar participant groups. These results do not always support the
predictions of PAM extensions. In So and Best (2011), Cantonese speakers were found to have more problems with CG pair (T35–T214) than SC (T55–T51), which contradicted PAM’s prediction that SC should have the poorest discrimination. However, Hao (2011) found that T35 and T214 in Mandarin were assimilated into different categories by Cantonese listeners, which should have led to a TC pair and excellent discrimination; instead, the results revealed that this pair was the most difficult for Cantonese speakers. In a reversed situation where Mandarin speakers discriminated Cantonese, the results supported PAM’s prediction that TC pairs always have excellent discrimination. Conversely, two CG pairs showed poor discrimination, contradiction PAM’s predictions. Similarly, in Reid et al. (2014), Mandarin speakers’ discrimination of Thai tones was generally in line with PAM’s predictions that TC had higher levels of discrimination than SC and CG. Additionally, Cantonese speakers discriminated SC and TC Thai tones equally well, in a manner inconsistent with PAM’s predictions. The authors explained that this might be due to the greater complexity of the Cantonese tone system. This might help Cantonese speakers’ ability to be more sensitive to tones in a way that Mandarin and Thai listeners are not. Another reason for Cantonese speakers’ better performance was that Cantonese speakers applied a greater level of phonological processing when perceiving both speech and non-speech, more than with Mandarin speakers showing increased sensitivity to the acoustic differences (Zheng et al., 2010).

4.1.3 Perceptual assimilation model and the current thesis

As discussed in Chapter 3, although tone perception research has constantly been broadened and deepened, a wide range of crucial questions remain unanswered. Indeed, we still do not know whether the discrimination accuracy of different tone
contrasts is consistent with PAM’s predictions. Moreover, within the range of tone perception by speakers of other tone languages, most studies have examined the tone perception of a language with fewer tones (e.g., Mandarin) by speakers of languages with a larger tone inventory (e.g., Cantonese) (Leung, 2008; So & Best, 2008). Only limited data are available for the reverse situation. Further, although tone language speakers mapping L2 tones to their L1 category has been confirmed, little understanding exists regarding how L2 tones are mapped to the L1 category. Finally, little is known about how non-tone language speakers who are learners of a tone language assimilate a new tone system. In terms of tone production, very little work has been undertaken within a PAM framework; this is likely due to PAM/PAM-L2’s focus on speech perception. One of the key goals of this thesis is therefore to test PAM-S in the domain of tone perception and provide an extension of PAM into tone production. A set of hypotheses based on PAM-S are described in the final three paragraphs of this section.

For non-tone language speakers, most tones are likely perceived as speech, although not categorisable according to a native phonological entity (e.g., the post-lexical intonation system), as both tone and intonation involves different F0 patterns. However, they have different applications: tone is lexical while intonation is post-lexical. Depending on the L1 prosodic system, a tone might be so similar to a L1 intonational structure that it will be possible for L2 listeners to categorise it using intonational categories. For example, tone might be associated with a monolexemic sentence. Thus, the L2 tones will be either uncategorisable or categorisable, with contrasts formulated as UC and UU. For a UU contrast, the L1 system should exert little influence on discrimination and the goodness should be fair to good, depending on the distance between the L2 and the closest L1 phonemes. However, a UC
contrast should have excellent discrimination results, as the two tones differ a great deal from each other.

For tone language speakers, L2 tones will most likely be perceived as categorisable with respect to a speaker’s L1 tone inventory. It is likely that some tone pairs will be TC, while others will be CG—and in rare cases perhaps even SC—as has been demonstrated in the studies discussed above (Hao, 2011; Leung, 2008) in a manner similar to that proposed by So and Best (2014). Two tones perceived as belonging to two different L1 (and perhaps L2) categories will form a TC categorisation pattern and will be easy to discriminate. If two L2 tones are perceived as instances of the same L1 (and perhaps L2) tonal category, they will be classified as a CG pair; the level of discrimination difficulty is predicted by the articulatory, acoustic and perceptual distance between the two members from the L1 category. If these two tones differ greatly from each other, as well as the L1 tone, they will still be easy to discriminate. However, if they are both close to the L1 category, discrimination will be more difficult. When two tones form a SC pair, it will be extremely difficult to discriminate them, as they are assimilated to the same L1 category with the same distance to the L1 tone. Assimilation results are given in Sections 5.1.2, 5.2.2 and 5.3.2 and specific predictions for the discrimination studies are given in Section 6.1.

4.2 The Speech Learning Model

4.2.1 Review of the speech learning model

The other theoretical model, SLM, has been the predominant framework for L2 production work. SLM was developed by Flege (1995) and his colleagues to explain the mechanisms underlying second language speech perception and production (mainly production). As SLM focuses primarily on the ultimate
attainment of an L2 phonological system, studies within an SLM framework are typically conducted with L2 speakers who have spoken the language for a several years. The model claims that most production errors are rooted in perception errors: without L1-like perception, L1-like production of speech is impossible.

SLM’s core theoretical contributions consist of four postulates and seven hypotheses derived from those postulates. Some SLM hypotheses are concerned with the relationship and development of a person’s L1 and L2 phonological systems in general. Here, SLM proposes that ‘the mechanisms and processes used in learning the L1 sound system remain intact over the life span’ (Flege, 1995, p. 239). In other words, there is no biologically determined ‘critical period’ within which language learning must happen, as has been previously posited by the critical period hypothesis (Lenneberg, 1967; Penfield & Roberts, 1959). PAM agrees on this with SLM.

The observation that most L2 learners find it difficult to discriminate some L2 sound contrasts (as they perceive them as instances of the same phonological category) is labelled the ‘similarity effect’ in the SLM framework (Flege 1987, 1988, 1995). This is quite similar to PAM’s prediction about SC: when two phonemes are perceived as instances of the same category in the L1 system, they will be very difficult to discriminate. In contrast, when a greater difference between L1 and L2 phones exists, it is assumed L2 learners find it easier to interpret different L2 phones as instances of different phonological categories. Indeed, in this case, if an L2 phone is perceived as highly different from sounds in the L1 inventory, a new category will be established. The properties of the new category will match those of the L2 phones closely. SLM thus predicts that L2 speech sounds that are absent in the L1 phonology system will be easier to acquire than those that overlap or are perceived as similar to
the existing L1 phonemes. These will be much more difficult to acquire, and are likely to be produced with an L2 accent. It is posited that L2 production will reflect L2 perception, as perception and production are linked to the same mental representation. According to SLM, how accurately L2 sounds are perceived predicts the accuracy of their production.

Like PAM, SLM predicts that listeners will learn a novel language through the filter of their first language. Specifically, SLM predicts that similar phonemes will be assimilated into a composite category. A process of assimilation and dissimilation over the course of learning results in the learning of L2 categories. SLM also makes very strong claims about the relationship of perception and production during learning. Specifically, the model claims that perception leads production (always occurring first in terms of learning), and that perception and production become closer to one another over the course of learning. SLM argues that problematic perception will lead to imperfect production, but it does not predict that all production errors are perceptually based: perception and production are linked indirectly and they may not share representations. SLM, from a psychoacoustic perspective, proposes that some representations are different, as perception has its roots in psychoacoustic elements while production is articulatory. Conversely, while PAM itself does not make strong claims regarding the production of novel contrasts, it does posit that speech perception and production share representations. Because of this general claim, we can infer that learning in one modality should be correlated strongly to learning in the other modality. As PAM posits a direct relationship between the two modalities, it must be the case that learning in each modality will be correlated under this. More studies favour SLM’s
indirect link between perception and production: that they possess separate representations, with complex links mapping one onto the other.

4.2.2 Extending the speech learning model to tone perception and production

As SLM does not provide specific predictions based on the difference between L1 and L2 systems, few studies have applied SLM as a model in the prosodic domain. The model I am proposing here extends SLM in the following ways:

1. For tone language speakers, L2 speakers will map L2 tones to the L1 categories, according to a similarity effect, as with vowels and consonants. An L2 tone from a completely different category than the L1 tone might be easier to perceive and produce than one perceived as being in the same category.

2. In the case of non-tone language speakers, it is likely they will use their L1 prosodic patterns to perceive L2 tones. Tones similar to existing prosodic patterns might be more difficult to perceive and produce for such learners, while tones with no overlap might be easier.

3. L2 tone perception and production are not directly linked. Perception precedes production and a problematic perception will lead to imperfect production.

Chapter 7 will present and discuss the production results, with the link between perception and production examined in the latter part of this chapter as well (Section 7.7). The evidence indicates that linguistic experiences shape the production of a new tone language in a similar way as they do in perception. SLM’s position
regarding the link between perception and production is supported by the current study.

4.3 Thesis Overview

The extensions of PAM and SLM presented above invite a number of research questions (RQ) pertaining to non-native tone perception and production. I outline four such research questions below, and then present a series of experiments (see Chapters 5, 6 and 7) that address these questions.

RQ 1: how are tones from a large tone inventory mapped to tones in a small inventory? Does this experience hinder or help? (This is addressed in the categorisation study in Chapter 5).

RQ 2: how do non-tone language speakers assimilate tones to their L1 prosodic system? (This is addressed in the categorisation study in Chapter 5).

RQ 3: does L1 and L2 tonal experience help in perceiving and producing another tonal language? (This is addressed in the discrimination study in Chapter 6 and the production study in Chapter 7).

RQ 4: what is the relationship between tone perception and production? (This is addressed in the discrimination study in Chapter 6 and the production study in Chapter 7).

To answer these four questions, it is necessary to conduct a series of perception and production experiments: a categorisation study, a discrimination study, and a production study. Detailed descriptions of the participants, procedures and results of these studies will be presented in Chapters 5 to 7 respectively. However, a brief introduction of the study’s aims and findings will be provided here to indicate how they are designed to answer the questions. The participant groups
and target languages are well thought through to ensure that we maximise the opportunity to understand the influences of previous linguistic experiences on perceiving and producing novel tones. None of the recruited participants had received consecutive years of musical training as several studies have demonstrated differences between musicians and non-musicians on successful discrimination of unfamiliar tones (Delogu, Lampis & Belardinelli; Gottfried, 2007; Marie et al., 2011).

4.3.1 Categorisation study (Chapter 5)

This categorisation study investigates how non-native tones (Cantonese) are perceived by speakers whose own native language has fewer tones (Mandarin speakers), whose native language does not have lexical tones (English speakers), and whose native language does not have lexical tones but where the second language has fewer tones (English speakers who are intermediate Mandarin learners). The analysis is presented within the PAM-S framework. The results by tone language speakers indicate both phonetic and phonological assimilation of Cantonese tones by Mandarin speakers. The results also suggest that non-tone language speakers can assimilate Cantonese tones to their native prosodic system. Native non-tone language speakers with L2 tone experience can take advantage of both their L1 and L2 experiences to assimilate non-native tones. The assimilation results determined the grouping patterns of a tone pair (TC, CG, SC or UU, UC), providing predictions for the other part of the perception experiment: the discrimination study in Chapter 6. For the first time, UU and UC pairs were further grouped into lower classifications, which enabled the test of predictions on these pairs formulated by PAM-S.
4.3.2 Discrimination study (Chapter 6)

This study investigates how native prosodic systems and L2 learning experience shape non-native tone discrimination. The same speaker groups from the categorisation study, along with a controlled group of native Cantonese speakers participated in this study. Native Cantonese speakers discriminated tones the best, followed by English speakers with Mandarin experience, Mandarin speakers and English speakers. The discrimination results were compared with predictions from PAM-S. The results from Mandarin speakers are most consistent with predictions from PAM-S: that TC > CG, UC-no overlap > UC-overlap > UC-same set. For English speakers, TC > CG, UC-no overlap > UC-overlap, and UU-overlap were the most easily discriminated pairs. However, even the mean accuracy of TC was higher than CG with English speakers; a few TC pairs showed lower accuracy than CG ones. For English Mandarin learners into English, the accuracy ranking of the tone groups is: TC ≥ CG > SC, UC-no overlap > UC-same set; for English Mandarin learners into Mandarin, TC ≥ CG, UC-no overlap > UC-overlap. Not all TC pairs were better discriminated than the CG pairs. Additionally, for all speaker groups, UC did not always have moderate to excellent discrimination, contradicting what PAM-S/PAM-L2 has proposed. The results from this study will be compared with the results from the production experiment (Chapter 7) to examine the relationship between perception and production.

4.3.3 Production study (Chapter 7)

This study investigates how native prosodic systems and L2 learning experience shape non-native tone production. The same speaker groups—speakers from tone language backgrounds (native Cantonese speakers and Mandarin speakers), and non-tone language backgrounds (English monolinguals, and English
speakers with Mandarin learning experience)—produced the six Cantonese tones in an imitation task. The results reinforce the influence of native prosodic systems on L2 tone production, regardless of tone or non-tone backgrounds. Mandarin speakers have more problems with pitch height, and English speakers tend to produce every tone in a level shape, which echoes the findings from previous perception studies. Further, Mandarin speakers’ ability to integrate their native sensitivity to pitch height along with their Mandarin training in pitch contour contributes to their exceptional performance in producing the new tone language. Further, the production results were compared with perception results to examine the relationship between the two modalities. The results show that speakers with either L1 or L2 tonal experiences display positive correlations between their perception and production, while speakers with no tonal experience indicate no correlation between the two abilities.

4.3.4 Justifications for languages and participants chosen

Chapter 2 detailed the importance and difficulty of perceiving and producing speech sounds within the same, and across two, prosodic typologies. The perception and production of different tone languages and between tone and intonation languages are the current study’s focus. As Chapter 3 reviewed, how tones are categorised and perceived, especially when the L1 has a smaller tone inventory compared with the new tone language, is not very clear. The perception of L2 tones by speakers coming from non-tone language backgrounds has been examined; however, no agreement or conclusion has been reached and the research has been undertaken without a unified methodology, as the comparison between the two prosodic systems is complex. Production by either tone language or non-tone language speakers requires more research, especially with the same participants as in the perception studies. Most importantly, what kind of influence L2 tone experience
may exert on the perception and production of a new tone language has been investigated rarely. The link between perception and production will be worthy of investigation, as previous research has found contradictory results. Chapter 4 provided frameworks and tools with which to design this experiment. PAM was used here to provide predictions based on categorisation results, while SLM helped to understand how production was related to perception, even though PAM initiates different opinions regarding the link between these two modalities.

From the above description of the three experiments we can see that the three languages involved in the whole design are Cantonese, Mandarin and English. As Chapter 2 introduced with great detail, Cantonese and Mandarin are two lexical tone languages that differ from each other not only in the number of tones (Cantonese has six contrastive tones while Mandarin has four), but also in the tones’ traits (all Mandarin tones have different contours while Cantonese tones are differentiated by both F0 register and contour). English, on the other hand, uses F0 information to convey meaning post-lexically. Australian English, as a dialect of English, has unique prosodic patterns and its L1 speakers can use both F0 register and contour information to differentiate different intonation patterns.

The current study takes the Cantonese tone system as the target tone system for participants to perceive and produce. Participants from a tone language background are Cantonese L1 speakers, Mandarin L1 speakers, while the non-tone language speakers are Australian English speakers. Another group of participants are L1 Australian English speakers who have been learning Mandarin as a second language. In this way, we have participants who come from a larger tone language, a smaller tone language, a non-tone language and L1 non-tone but L2 tone background. This selection of languages and participants maximises the contrast in
prosodic systems; as such, we can examine the influence of L1 and L2 prosodic systems and their interaction on non-native tone perception and production. This has a significant potential for such research.

4.4 Summary

The increased attention paid to speech perception and production in the prosodic domain highlights the serious need for theoretical models providing comprehensive and testable predictions concerning this level. While existing versions of PAM/PAM-L2/PAM-S and SLM have been hugely influential and successful in phoneme (vowel and consonant) perception and production, little work has hitherto been done to extend PAM to tone perception and production. The current model combines PAM and PAM-L2, along with corresponding traits from SLM, in an attempt to fill the gap of tone production, and the relationship between tone perception and production. First, these extensions will enable the formulation of testable hypotheses for the perception and production of tones by speakers with different linguistic experiences, by using PAM. Second, it will allow a greater focus on the relationship between perception and production with the combination of SLM and PAM-L2.

The following three chapters (5 to 7) will introduce the three studies in detail: categorisation, discrimination and production, including the participant recruitment, experimental materials and procedures, results and a discussion of the results.
Chapter 5: Categorisation of Cantonese Tones

This chapter contains the introduction, method, results and discussion of the categorisation study, which is the first part of the study’s perception facet. Different speaker groups who differ in their lexical tone experiences assimilated Cantonese tones to their L1/L2 prosodic systems. The aim is to determine how speakers from different language backgrounds categorise complex Cantonese tones. The categorisation mappings will form our predictions for their discrimination performance, based on PAM/PAM-S. The categorisation patterns by the three participant groups—L1 Mandarin speakers, English monolinguals and L1 English speakers with Mandarin experience—will be introduced separately. As demonstrated by the research on tone perception reviewed in Section 3.1, previous linguistic experiences influence non-native tone perception. Unsurprisingly, research has shown that some speakers of L1 tone languages may successfully use their native prosodic system in perceiving a new tone system (e.g., Hao, 2011; Leung, 2008; So & Best, 2011). What is less clear, likely due to minimal research on this topic, is how L2 tone systems are perceived by speakers with a smaller tone inventory, and by tone-naïve non-tone language speakers. Similarly, it is unclear if L2 learners of tone languages with non-tone L1s can use knowledge from the L2 tone system to aid perception of an L3 tone system. Indeed, only one paper to date has examined speakers coming from a non-tone background but who have learned a tone language as second language (Qin & Jongman, 2015).

The following sections present the categorisation of Cantonese tones by three speaker groups: native Mandarin speakers, English monolinguals, and native English speakers with Mandarin learning experience. In doing so, the chapter addresses RQs 1 and 2 (see Chapter 4). The results show that both tone and non-tone speakers can
assimilate non-native lexical tones to their native prosodic system. Moreover, not only L1, but also L2 learning experience influences the assimilation pattern.

5.1 Background

As discussed in Chapter 3, it is a well-established fact that the perception of L2 tones is influenced by an individual’s L1 tone language experience (Burnham et al., 2014; Lee et al., 1996; So, 2008; So & Best, 2010; So & Best, 2014; Wayland & Guion, 2004). However, whether this L1 experience facilitates or interferes with L2 perception remains unclear. Existing research suggests that this depends on both the discrepancies and the similarities between the specific L1 and L2 tone systems in question. A particularly pertinent question is how L1 experience with a comparatively simple tone system might influence listeners’ perceptions of more complex L2 tones: this is also unclear (Qin & Mok, 2011).

PAM (Best, 1995, see Chapter 4 for a review) has increasingly been extended to account for cross- and second language speech perception of prosodic features, most notably in the form of PAM-S (So & Best, 2014). The predictions of PAM/PAM-S have been tested in a number of tone perception studies (cf. Chiao et al., 2011; Hao, 2011; Leung, 2008; Reid et al., 2014; So & Best, 2008; So & Best, 2011). PAM-S makes clear predictions about the discriminability of L2 tones based on their categorisation (or lack thereof) into the available L1 tone categories. These predictions are consistent with the results from studies concluding that L2 tones are mapped onto tone language speakers’ L1 tone categories, and that this L1 influence is difficult to overcome, even with training (Leung, 2008). Research also suggests that some difficulties in discrimination are universal, regardless of listeners’ language backgrounds. These difficulties might be due to the phonetic similarities of the particular pair (Burnham et al., 2014; So & Best, 2010). Support for PAM
predictions has also been found in the discriminability of tone pairings classified as PAM TC and CG contrasts respectively, such that L2 tone TC contrasts are easier to discriminate than L2 CG contrasts (Qin & Mok, 2011; Reid et al., 2014; So & Best, 2011). However, different categorisation methods greatly influence individual study results.

Typical *segmental* perception studies differentiate between phonetic and phonological assimilation: Phonetic assimilation occurs when one L2 phone is perceived as the phonetic equivalent of a tone in the L1 category. In contrast, phonological assimilation occurs when the same phonological behaviour (the application of L1 phonological knowledge) is evident in both the L1 and L2 categories. Few studies have examined this issue in terms of prosodic features (*suprasegmental* properties). A recent study (Wu et al., 2014) suggests that phonological assimilation only occurs in experienced listeners, while other findings indicate that phonological assimilation may also occur in inexperienced listeners (So, 2012; So & Best, 2010b).

### 5.2 Categorisation of Cantonese by Mandarin Speakers

#### 5.2.1 Method

**5.2.1.1 Participants**

Twenty L1 Beijing-accented Mandarin speakers (mean age 23.8 years, standard deviation ($SD$) = 2.85) participated in this experiment. All participants had been born and raised in Beijing, and had arrived in Australia after they had turned 18. They had little exposure to Cantonese and claimed that Cantonese was a foreign language to them. The language background questionnaire for participant recruitment can be found in Appendix A.
5.2.1.2 Stimuli

The stimuli for the present study were selected to test the categorisation of Cantonese tones into the Mandarin tone system and the English intonation system. Thus, a syllable existing in all three languages is preferable. The string /mɔː/ was chosen as it exists in Cantonese (‘mo’ 摸), English (‘more’), as well as in Mandarin. In fact, ‘Mo’ carrying all four Mandarin tones correspond to four actual Mandarin words: ‘摸 touch’, ‘磨 scrub’, ‘抺 swipe’ and ‘末 powder’. These words are in daily use in Mandarin and before the task began, I confirmed that all Mandarin participants could recognise them. This design enables investigation into whether Cantonese tones can be assimilated into the Mandarin tone system by native Mandarin speakers and Mandarin learners.

The 18 Cantonese tokens (6 tones× 3 repetitions) were recorded by a female L1 Cantonese speaker (25.6 years old); the 12 Mandarin tokens (4 tones× 3 repetitions) were recorded by a female L1 Mandarin speaker (23.9 years old). The most clearly pronounced tone production from the three repetitions was chosen as the final stimuli by a native speaker of Mandarin.

Stimulus recording was conducted at MARCS Auditory recording booth at Western Sydney University, with a Technica Audio AT892CT4 head-mounted microphone positioned directly in front of the speaker in a sound-attenuated booth. The microphone was connected to a digital recording device, a Dell Dimension E521 computer with a Sigma C-Major Audio sound card, located in an adjacent sound-attenuated booth. The recording software Cool Edit was used, with a sampling rate of 44010Hz, and a resolution of 16 bits.
The pitch contours extracted from the stimuli are illustrated in Figures 5.1 and 5.2. For the Mandarin tones, T1 and T2 have similar pitch offsets, while T2 and T3 share similar onsets.

*Figure 5.1. Pitch contour of the four Mandarin tones in /mɔː/ produced by the female speaker*

*Figure 5.2. Pitch contours of the six Cantonese tones in /mɔː/ produced by the female speaker*

From Figure 5.2, we can see that Cantonese has a more complex tone system and a more crowded tonal space: four tones (T2, T4, T5 and T6) have quite similar pitch onsets. Among the three level tones (T1, T3 and T6), the difference between the high- and mid-level tone (T1 and T3) is about twice that between the mid- and low-level tones (T3 and T6): 60Hz to 30Hz. Low-falling (T4) starts at the same pitch as the low-level, but then drops. The two rising tones, T2 and T5, both start at around 140Hz, but rise to 220Hz and 170Hz, respectively.
5.2.1.3 Procedure

Participants were asked to categorise the randomised individual presentations of 120 trials of the target word (/mɔː/ tones) (6 tones × 20 repetitions) as one of the four Mandarin tones: level, rising, dipping and falling. In addition, an ‘unknown’ choice was provided. The 120 tokens were randomised in E-Prime 2.0. During the experiment, the stimuli tokens were presented individually from a laptop (Sony SVT131A11W), on the screen of which several choices were provided, corresponding to the Mandarin tone categories (written in pinyin form) with the addition of an unknown choice. Each response ‘button’ was hyperlinked to a pre-recorded example of the corresponding Mandarin tones. The ‘unknown’ button was not hyperlinked to an example.

Listeners were instructed to click on the button and compare the target Cantonese syllable and the four Mandarin syllables and then choose the most similar one and type a goodness rating (1 to 5) for that syllable, with 1 being least alike and 5 being very alike. They were instructed to choose ‘unknown’ when they could not identify a target word’s tone with any in the L1 tone category. They could listen to the stimuli as many times as they wished. The maximal comparisons for each token were 6 times and the minimal was 1 time. Participants became faster as the task proceeded. It took approximately 10 minutes for each participant to finish the task. A screenshot of the experiment screen is provided in Appendix B, Figure B.1.

5.2.1.4 Defining ‘Categorised’

The current study applies the definition of ‘categorised’ presented in So and Best (2014). Here, and thus in the present study, a tone is considered categorised only if it satisfies two criteria: the number of choices for the chosen category should be significantly higher than 1) chance level, and 2) other presented options. If a given
L2 tone fails to satisfy both of these criteria, it will be considered uncategorised. In the current study, the participants were presented with five competing choices (four Mandarin tones plus one ‘unknown’ response option), for each L2 Cantonese tone. The response patterns for each Cantonese tone were subjected to \( t \)-tests against chance level (20% in this case) and other competing choices.

**5.2.2 Results**

The total number of responses for each tone category was 400 (20 participants \( \times \) 20 repetitions). To test whether the participants’ patterns of categorisation differed from chance performance, I conducted a series of \( t \)-tests against chance performance (chance level for each tone is 20%, with the provided number of response options). The results of the \( t \)-tests are provided in Table 5.1.

<table>
<thead>
<tr>
<th>Cantonese tone</th>
<th>Chosen Mandarin tone</th>
<th>Percentage</th>
<th>Df</th>
<th>( t )-test</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1 (T55)</td>
<td>Tone 1 (T55)</td>
<td>92</td>
<td>19</td>
<td>53.817</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 2 (T25)</td>
<td>Tone 2 (T35)</td>
<td>54</td>
<td>19</td>
<td>12.764</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 2 (T25)</td>
<td>Tone 3 (T214)</td>
<td>34</td>
<td>19</td>
<td>5.270</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 3 (T33)</td>
<td>Tone 1 (T55)</td>
<td>70</td>
<td>19</td>
<td>16.327</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 4 (T21)</td>
<td>Tone 3 (T214)</td>
<td>68</td>
<td>19</td>
<td>15.363</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 5 (T23)</td>
<td>Tone 2 (T35)</td>
<td>40</td>
<td>19</td>
<td>7.774</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 5 (T23)</td>
<td>Tone 3 (T214)</td>
<td>44</td>
<td>19</td>
<td>8.295</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Tone 6 (T22)</td>
<td>Tone 1 (T55)</td>
<td>79</td>
<td>18</td>
<td>18.616</td>
<td>( p &lt; 0.001 )</td>
</tr>
</tbody>
</table>

The categorisation results are as summarised in Figure 5.3. All three Cantonese level tones were categorised as instances of the only Mandarin level tone
(MT155). Indeed, CT1 (T55) was categorised as the high-level tone in Mandarin 92% of the time, with a goodness rating of 3.9. For CT3 (T33) and CT6 (T22), the Mandarin level tone was chosen 70% and 79% of the time respectively, with a goodness rating of 3.3 and 3.0. The two Cantonese rising tones CT2 (T25) and CT5 (T23) were categorised into MT2 (T35) and sometimes MT3 (T214). For these two rising tones, thus, two categories in Mandarin were chosen (above 20% chance level)—MT2 (T35) and MT3 (T214). However, upon closer examination, we can see that for CT2 (T25), MT2 (T35) is the primary choice (54%), which is significantly higher than the other choice of MT3 (T214) (34%). By contrast, MT2 (T35) was selected 40% of the time, and MT3 (T214) 44% of the time for CT5. CT4 (T21) was categorised into MT3 (T214) in 68% of cases with a goodness rating of 3.3, while interestingly, for 30% of the time, MT4 (T51) was chosen, with a higher rating of 3.5.

A chi-square test revealed a significant association between Cantonese tones and the chosen Mandarin categories $\chi^2 (20) = 2425.146, p < .001$. This was further examined in a two-way repeated-measures ANOVA (CT × MT), which revealed a significant main effect of CT, $F(5, 14) = 45.178, p < .001$, as well as a significant effect of MT, $F(3, 285) = 106.065, p < .001$, on listeners’ mean assimilations. The CT × MT interaction was also significant, $F(6, 285) = 246.359, p < .001$. 

84
The total number of responses for each tone category was 400 (20 participants × 20 repetitions).

The symbols * (p < .001) show that the mean is significantly above the chance level (20%).

CT = Cantonese tones, MT = Mandarin tones

Figure 5.3. Mandarin listeners’ tonal categorisation percentage for each Cantonese tone and its goodness rating in brackets

Individual one-way ANOVAs on the percentage of Mandarin tone choices for each Cantonese tone target were also conducted to investigate the interaction between Mandarin tone choices and Cantonese tone categories. The Mandarin tone effect was significant for each Cantonese tone: CT1, $F(1, 36) = 2611.713, p < .001$; CT2, $F(2, 54) = 102.377, p < .001$; CT3, $F(2, 54) = 229.241, p < .001$; and CT4, $F(1, 36) = 89.605, p < .001$; CT5, $F(2,54)=45.673, p <.001$; CT6, $F(1,36)=214.438, p <.001$.

Within the tone groups with more than one category selected above the chance level, the percentages of CT2 being categorised as MT2 and MT3 are significantly different (p <.001), while the differences between CT5 being categorised as MT2 and MT3 are not significant (p = .259).
According to the two criteria established previously, the categorised type for each Cantonese tone can be decided: five of the six tones are categorised, while CT5 (T23) is uncategorised. Regarding CT1 (T55), CT3 (T33) and CT6 (T22), which are all mapped onto the same category, a t-test for the goodness rating was performed. With goodness ratings differing significantly from each other, the tone pairs formed by CT1, CT3 and CT6 are considered CG instead of SC. Tone pairs which include CT5 constitute UC pairs—to be specific—depending on whether the pair shares overlaps. UC pairs are further grouped into UC-no (no overlap), UC-o (partly overlap) and UC-s (same-set).

Table 5.2

Summary of the Categorisations of the Six Cantonese Tones—Mandarin Speakers

<table>
<thead>
<tr>
<th>Cantonese tones</th>
<th>Mandarin tones</th>
<th>Status</th>
<th>Percentage; rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT 1 (T55)</td>
<td>MT 1 (T55)</td>
<td>C</td>
<td>(92%; 3.9)</td>
</tr>
<tr>
<td>CT 2 (T25)</td>
<td>MT 2 (T35)</td>
<td>C</td>
<td>(54%; 3.3)</td>
</tr>
<tr>
<td>CT 3 (T33)</td>
<td>MT 1 (T55)</td>
<td>C</td>
<td>(70%; 3.3)</td>
</tr>
<tr>
<td>CT 4 (T21)</td>
<td>MT 4 (T51)</td>
<td>C</td>
<td>(68%; 3.3)</td>
</tr>
<tr>
<td>CT 5 (T23)</td>
<td>MT 2 (T35)</td>
<td>U</td>
<td>(40%; 3.4)</td>
</tr>
<tr>
<td></td>
<td>MT 3 (T214)</td>
<td></td>
<td>(44%; 3.3)</td>
</tr>
<tr>
<td>CT 6 (T22)</td>
<td>MT 1 (T55)</td>
<td>C</td>
<td>(79%; 3.0)</td>
</tr>
</tbody>
</table>
A summary of the assimilation patterns of tone contrasts is presented in Table 5.3, where Cantonese pairs T1-T2, T1-T4, T2-T3, T2-T4, T2-T6, T3-T4 and T4-T6 are TC groups, T1-T3, T1-T6 and T3-T6 are CG pairs, T1-T5 and T3-T5 are UC-no overlap, T4-T5 and T5-T6 are UC-overlap, and T2-T5 is in the UC-same set pattern. According to the predictions made by PAM, PAM-L2 and PAM-S, the discrimination of TC should be good, CG should be moderate, UC-no overlap should be good, UC-overlap moderate and UC-same set should be poor. These contrasts will form the core of the second part of the perception study, presented in Chapter 6.

### 5.2.3 Discussion

This study examines how L1 Mandarin speakers categorise Cantonese tones into their own tone system, which has a smaller tone inventory. The results indicate that in most cases, L2 Cantonese tones are categorised as the most acoustically similar L1 Mandarin tone counterparts. The fact that all three level tones are categorised as the only available Mandarin level tone clearly demonstrates that even partial similarity can stimulate phonetic assimilation. However, the differences in goodness ratings suggest that Mandarin speakers are indeed able to differentiate F0 height: Mandarin listeners found CT1 the best fit for MT1, although they also chose
MT1 for CT3 and CT6. Interestingly, for the two rising tones CT2 (T25) and CT5 (T23), listeners were debating between MT2, which is a rising tone (T35) and MT3 (T214), which has an allophonic tone as a rising tone (T35). When the Cantonese rising tone is categorised as the rising tone in Mandarin, this means that the assimilation happens at the phonetic level. However, when the rising tone is assimilated to the allotone of the Mandarin dipping tone (T214), this means that phonological assimilation has also applied for those speakers.

As the categorisation method of the current results differs slightly from some previous research, direct comparison is somewhat difficult. For example, earlier studies of tone categorisation from Cantonese to Mandarin (Qin & Mok, 2014) did not employ participants to categorise tones; rather, the researchers mapped the relationship between the two tone languages by comparing acoustic similarities and differences. CT2 (T25) and CT5 (T23) were both categorised to MT2 (T35), with CT2 being a better exemplar than CT5. Based on this, CT2 and CT5 fell into the CG contrast. The reason for this result was that the researchers focused only on phonetic assimilation rather than listener choices. However, some results reviewed in Chapter 3 show that phonological assimilation might also occur in this situation (Huang, 2001; Leung, 2008; So & Best 2010). Thus, relying solely on phonetic similarities could result in the loss of these phonological assimilation phenomena.

With respect to phonological tone assimilation, Best and Tyler (2007) proposed that this level can only be accessed by experienced listeners. So (2012) and So and Best (2010) later found that phonological assimilation is also possible for inexperienced listeners, where Cantonese listeners categorise both the Mandarin high-level tone and high-falling tone into the Cantonese high-level tone. Indeed, as discussed in Section 2.2.1, the two tones (high-level and high-falling) are free
variants in Cantonese. In the current study, the low-falling tone and rising tone are perceived by Mandarin speakers as allophonic variants of the Mandarin falling-rising tone (MT3). Thus, when the rising tone (T23) and the low-falling tone (T21) are both categorised into the falling-rising tone, we could say that the phenomenon of phonological assimilation is present. For Mandarin speakers, phonological assimilation is likely to occur due to allophonic tone patterns in the native language, such as when a falling-rising tone will be assimilated to its allotonic variants, a rising tone or a low-falling tone, and vice versa. Thus, to Mandarin speakers, these three tones are assimilated as phonologically similar tone categories, even though they have different F0 height and contours. According to the above data, when low-falling (CT4 [T21]) or rising tones (CT2 [T25] and CT3 [T23]) are assimilated as the Mandarin falling-rising tone (MT3), then phonological assimilation is present. If we establish the criteria as being the modal response, then from the fit index we can determine that MT3 is the modal response for CT4 and CT5. The fact that CT4 is categorised as MT3 aligns with predictions made by Qin and Mok (2014); thus, another explanation for this could be that listeners pay attention selectively to the former part of the falling-rising tone.

Wu et al. (2014) argue that sometimes a choice is made due to the participants being obliged to choose one tone from their L1 category; sometimes they choose one with only partially similar features, as they cannot find a better match. In the current study, even though the listeners were given an ‘unknown’ button, listeners chose ‘unknown’ only in a few cases. Even where there was no perfect fit, they still tried to find a tone that shared even some limited similarities with the L2 tone category.
5.3 Categorisation by English Speakers without Tone Language Experience

This sub-section of Study 1 investigates how speakers from a non-tone language background categorise the six Cantonese tones. As reviewed in Chapter 2, English is typologically different from Cantonese and Mandarin, as it uses pitch only at the post-lexical level. However, non-tone language speakers can still make use of their own prosodic system to perceive lexical tones (see Chapter 3.2.2). We thus predict that English speakers will categorise Cantonese tones into those (Australian) English intonation patterns that share similar F0 shapes. As previous evidence shows, Australian English speakers can discriminate rising intonation contours by both the height and range of rise (Fletcher & Harrington, 2001). Thus, we predict that our L1 Australian English-speaking participants will be able to categorise the two rising Cantonese tones (T23 and T25) into rising intonation contours with different rising ranges.

5.3.1 Method

5.3.1.1 Participants

Twenty L1 Australian English monolinguals (Mage = 22.7, SD = 3.25) participated in this study. All speakers were undergraduate students at the University of Western Sydney. No participants had experience with Cantonese nor had they received extensive musical training. All passed a pure tone hearing screening (250–8000Hz at 25dB HL) experiment first, to ensure that all listeners could discriminate tones at a basic level.

5.3.1.2 Stimuli

The string /mɔː/ was used for the stimuli, as it resembles ‘mo’ in Cantonese and ‘more’ in English. The Cantonese stimuli were the same as in the categorisation
by Mandarin speakers. English ‘More’, carrying five different intonation patterns was chosen as the corresponding L1 match: ‘More?’, ‘More!’, ‘More.’, ‘More…’, and ‘More?!’. The English stimuli were recorded by a female Australian speaker (age 28.5), born and raised in western Sydney, under similar recording conditions. Intonation contours of the English stimuli are shown in Figure 5.4.

![Figure 5.4. Pitch contour of the five English tunes in /mɔː/ produced by the female speaker.](image)

The five intonation patterns for English are as follows: ‘More?’ and ‘More?!’ are rising, with ‘More?’ having a sharper trajectory and higher range; ‘More!’ and ‘More.’ both have a falling contour, but the falling trajectory in ‘More!’ starts earlier in the token and has a greater excursion than ‘More.’; while ‘More…’ is a level pattern. The ToBI transcriptions of the five intonations are given in Table 5.4. This experimental procedure was inspired by So and Best (2010); however, these authors did not provide model, naturally occurring intonation patterns for participants. Rather than relying on participants’ imagined intonation patterns, this study asked the participants to match Cantonese tones with recordings of these English intonation tunes.
Table 5.4

*English Stimuli and Tones and Break Indices Transcriptions*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ToBI Transcription</strong></td>
<td>L* H-H%</td>
<td>L+H* L-L%</td>
<td>H* L-L%</td>
<td>H*H-L%</td>
<td>H* H-H%</td>
</tr>
<tr>
<td><strong>Tune</strong></td>
<td>High-rise—rise from low pitch</td>
<td>Rise-fall</td>
<td>Fall</td>
<td>Level</td>
<td>High-rise</td>
</tr>
</tbody>
</table>

5.3.1.3 Procedure

In a manner similar to that employed for the L1 Mandarin participants, L1 English participants were asked to categorise the randomised individual presentations of 120 trials of the target word (/mɔː/ + tones) (6 tones × 20 repetitions) into the five English intonation categories—‘More?’, ‘More.’, ‘More!’, ‘More…’, ‘More?!’. Similarly, an ‘unknown’ button was provided. All other procedures replicated those undertaken with Mandarin speakers and reported in Section 5.1.3. An experiment screenshot can be found in Appendix B, Figure B.2.

5.3.2 Results

The current study provided six Australian English intonation choices for each of the six Cantonese tones (including an ‘unknown’ category. As a result, the chance level for each category is 17% (100/6). Every choice over 17% has been examined with t-tests, with the results provided in Table 5.5.
Table 5.5

*Summary of the t-tests of Each Choice—English Speakers*

<table>
<thead>
<tr>
<th>Cantonese tone</th>
<th>Chosen English</th>
<th>Percentage</th>
<th>Df</th>
<th>$t$-test</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1 (T55)</td>
<td>More…</td>
<td>81</td>
<td>19</td>
<td>30.52</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 2 (T25)</td>
<td>More…</td>
<td>31</td>
<td>19</td>
<td>4.77</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More?!</td>
<td>31</td>
<td>19</td>
<td>6.13</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 3 (T33)</td>
<td>More…</td>
<td>63</td>
<td>19</td>
<td>13.72</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 4 (T21)</td>
<td>More.</td>
<td>94</td>
<td>19</td>
<td>68.51</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 5 (T23)</td>
<td>More?</td>
<td>31</td>
<td>19</td>
<td>5.21</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More.</td>
<td>56</td>
<td>19</td>
<td>12.03</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 6 (T22)</td>
<td>More.</td>
<td>44</td>
<td>19</td>
<td>9.84</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More…</td>
<td>38</td>
<td>19</td>
<td>16.08</td>
<td>$p &lt; 0.001$</td>
</tr>
</tbody>
</table>

The categorisation results are summarised in Figure 5.5. For 81% of the time, the Cantonese high-level tone (T55) was categorised as the intonation tune ‘More…’ in English, with a goodness rating of 3.7. For the other two level tones, CT3 (T33) and CT6 (T22), ‘More…’ was chosen for 63% and 38% of the time respectively, with a goodness rating of 2.8 and 3.0. In particular, the low-level tone attracted a greater number of ‘More.’ choices, with a percentage of 44% and a goodness rating as high as 3.5. The high-rising tone (T25) had dual categories: ‘More…’ and ‘More?!’, with equal likelihood of selection (31%), while the former had a higher goodness rating (2.8) than the latter (2.3). The low-rising tone was mainly categorised into ‘More.’ (56%), but the goodness rating was relatively low (2.1). English listeners in this study reached the highest agreement on the categorisation of the low-falling tone (T21), with ‘More.’ selected 94% of the time: the goodness rating is the highest (4.1) as well.
Note: The total number of responses for each tone category was 400 (20 participants × 20 repetitions).

The symbols * ($p < .001$) show that the mean is significantly above the chance level (17%).

Figure 5.5. English listeners’ tonal categorisation percentage for each Cantonese tone and its goodness rating in brackets.

Using the two criteria established previously, the categorised type for each Cantonese tone can be determined: four tones (CT1 [T55], CT3 [T33], CT4 [T21] and CT5 [T23]) are categorised and CT2 [T25], along with CT6 [T22], are uncategorised. Regarding the two pairs CT1-CT3, and CT4-CT5, which are each mapped onto the same category, a $t$-test for the goodness rating was performed. With goodness ratings significantly different from each other, tone pairs involving any of these four tones are considered to be CG instead of SC. Tone pairs involving CT2 or CT6 constitute UC pairs. Specifically, depending on whether the pair shares any overlap, UC pairs were further grouped into UC-no overlap, UC-partial overlap and UC-same set. Further, the pair formed by CT2-CT6 is a UU pair, and in the current case a UU pair with overlap, as they share the ‘More…’ category.
Table 5.6

**Summary of the Categorisations of the Six Cantonese Tones—English Speakers**

<table>
<thead>
<tr>
<th>Cantonese tones</th>
<th>English intonation</th>
<th>Status</th>
<th>Percentage; rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT 1 (T55)</td>
<td>More…</td>
<td>C</td>
<td>(81%; 3.7)</td>
</tr>
<tr>
<td>CT 2 (T25)</td>
<td>More…</td>
<td>U</td>
<td>(31%; 2.8)</td>
</tr>
<tr>
<td></td>
<td>More?!</td>
<td></td>
<td>(31%; 2.3)</td>
</tr>
<tr>
<td>CT 3 (T33)</td>
<td>More…</td>
<td>C</td>
<td>(63%; 2.8)</td>
</tr>
<tr>
<td>CT 4 (T21)</td>
<td>More.</td>
<td>C</td>
<td>(94%; 4.1)</td>
</tr>
<tr>
<td>CT 5 (T23)</td>
<td>More.</td>
<td>C</td>
<td>(56%; 2.1)</td>
</tr>
<tr>
<td>CT 6 (T22)</td>
<td>More.</td>
<td>U</td>
<td>(44%; 3.5)</td>
</tr>
<tr>
<td></td>
<td>More…</td>
<td></td>
<td>(38%; 3.0)</td>
</tr>
</tbody>
</table>

The summary of the assimilation tone contrasts is listed in Table 5.7, where T1-T4, T1-T5, T3-T4 and T3-T5 are TC groups; T1-T3 and T4-T5 are CG pairs, T2-T4 and T2-T5 are UC-no overlap, T1-T2, T1-T6, T2-T3, T3-T6, T4-T6 and T5-T6 are in the UC-overlap group, and T2-T5 is in the UC-same set group. According to the predictions made by PAM, PAM-L2 and PAM-S, the discrimination of TC should be good, CG should be moderate, UC-no overlap should be good, UC-partial overlap should be moderate and UC-same set should be poor. Similarly, these contrasts will be tested in the second part of the perception study—the discrimination experiment, which will be presented in Chapter 6.
Table 5.7

Summary of the Assimilation Patterns—English Speakers

<table>
<thead>
<tr>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tone 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tone 2</strong></td>
<td>UC-o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tone 3</strong></td>
<td></td>
<td>CG</td>
<td>UC-o</td>
<td></td>
</tr>
<tr>
<td><strong>Tone 4</strong></td>
<td></td>
<td>TC</td>
<td>UC-no</td>
<td>TC</td>
</tr>
<tr>
<td><strong>Tone 5</strong></td>
<td></td>
<td>TC</td>
<td>UC-no</td>
<td>TC</td>
</tr>
<tr>
<td><strong>Tone 6</strong></td>
<td>UC-o</td>
<td>UU-o</td>
<td>UC-o</td>
<td>UC-o</td>
</tr>
</tbody>
</table>

5.3.3 Discussion

English speakers mostly chose ‘More.’ (the falling tune) or ‘More…’ (the level tune) when asked to assimilate the six Cantonese tones to the five different English intonation patterns and one unknown category. These two English tunes worked as their ‘default’ choices for the Cantonese tones. Participants showed most agreement on the low-falling tone (T21)—94% of the choices were made for ‘More.’, with a goodness rating score as high as 4.1. ‘More.’, which has a falling tune, is the most familiar intonation to English speakers. Two of the three level tones (T55 and T33) were mainly categorised into ‘More…’, which has a level contour. However, the low-level tone (T22) was uncategorised for English speakers. They found two matching intonations, ‘More…’ and ‘More.’, for this target level tone, with ‘More.’ being the category chosen most. Although “More…” is more of a level pitch, English speakers chose the falling pitch contour (‘More.’) to go with the low-level Cantonese tone. The other uncategorised tone was the high-rising tone (T25), where participants chose ‘More…’ and ‘More?!’ equally. ‘More…’, with the level pitch contour, was also chosen as the counterpart of one of the rising tones. For the low-rising tone (T23), the main category was ‘More.’, which has the contradictory
pitch contour (falling vs. rising). However, the fall in our stimuli was quite shallow and Australian English is a ‘rising’ variety that uses rising intonations very frequently. The secondary choice was ‘More?’, the question intonation. As discussed previously, the main difference between the two high-rising intonations of ‘More?’ and ‘More?!’ is that ‘More?!’ has a higher start when compared to ‘More?’.

Even though it was not indicated by the main choice of the two Cantonese rising tones, a greater number of participants chose ‘More?!’ for the high-rising tone (T25) and ‘More?’ was the secondary choice after the falling tune (‘More.’) chosen for the low-rising tone (T23).

The current results indicate that most agreement was reached on CT1, CT3 and CT4, which are either level or a low-falling tone. Similar preferences were found regarding the Mandarin falling tone by English speakers in other studies. L1 English speakers have the impression that the Mandarin falling tone is the only ‘normal’ tone and the falling tone is perceived differently from the other three tones by L1 English speakers (Broselow, Hurtig & Ringen, 1993). The perceptual advantage of MT4 (T51) is seen as a transfer of English intonation, as is the case when MT4 (T51) is misperceived as MT1 (T55). When English listeners hear the falling Mandarin tone, they might take the latter’s falling part as the sentence-final intonation and the former part, which has the same F0 onset as MT1. As argued in a number of studies (Pike, 1945; Trager & Smith, 2009; Liberman, 1978; Pierrehumbert, 1980), English intonation has its underlying form based on level pitch targets (as reviewed in Chapter 2). The contours in intonation are interpolations between high- and low-level tone targets.

In general, most of the chosen intonation patterns did not share similar pitch contour patterns with the target Cantonese tones. This study shows clearly that
English speakers are very sensitive to pitch register differences, as the two rising tones in Cantonese are categorised quite differently. Australian English speakers differentiate statement and question rises by using higher pitch accents; that is, higher starting points for the rise on questions than on statements (Fletcher & Harrington, 2001). Findings from the present study that these Australian English speakers could differentiate ‘More?’ from ‘More?!’ (both are high rises but “More?!” has a much lower onset) suggest their ability to detect pitch range difference.

5.4 Categorisation by English Speakers Who are Mandarin Learners

This sub-section of Study 1 investigates how speakers from a non-tone language background, but who have L2 tone experience (here, Mandarin), categorise novel L3 tones (here, the six Cantonese tones). The aim is to determine how different the influences provided by the two prosodic systems (L1 and L2) are, and whether L2 tone experience can transfer to the tone system of an unfamiliar L3, something which has rarely been examined previously.

5.4.1 Method

5.4.1.1 Participants

Eighteen L1 Australian English speakers with intermediate Mandarin learning experience ($M$ age = 24.3, $SD$ = 3.72) participated in this study. The Mandarin learners were mostly undergraduate students studying Chinese at the University of Melbourne, and the rest were from language institutes in Sydney. These participants have all learned over 250 Chinese characters when they were tested. No participants had experience with Cantonese, nor had they received extensive musical training.
5.4.1.2 Stimuli

The stimuli used in this task combined all the Cantonese, Mandarin and English stimuli used in previous tasks with Mandarin and English speakers.

5.4.1.3 Procedure

Participants were asked to categorise the randomised individual presentations of 120 trials of the target word (/moː/ tones) (6 tones × 20 repetitions), first into the five English intonation categories—‘More?’, ‘More.’, ‘More!’, ‘More…’ and ‘More?!’—and then into the four Mandarin tone categories—level tone, rising tone, dipping tone and falling tone. In addition, an ‘unknown’ button was provided for both tasks. Procedures were the same as in the previous two experiments.

5.4.2 Results

The categorisation results are illustrated in Figure 5.6. The chance level for categorising into English intonation is 17% (100/6 categories), while for categorising into the Mandarin tone, the chance level for each tone is 20%. Participants’ choices over the two chance levels in English and Mandarin categories were examined with t-tests and are summarised in Tables 5.8 and 5.9 respectively. As shown in Figure 5.6, for the three level tones, the high-level (T55) and the mid-level (T33) tones are categorised into the same English tune ‘More…’ 95% and 78% of the time, respectively. The biggest category for the low-level tone (T22) is the intonation ‘More.’, but the secondary category is ‘More…’. For the high-rising tone (T25), ‘More?’ and ‘More?!’ were both chosen, with ‘More?!’ having a slightly higher proportion and goodness rating. The low-rising tone was mainly categorised into ‘More?!’, for 63% of the time, while 29% of the time, ‘More?’ was chosen. The low-falling tone (T21) was categorised as ‘More.’ 65% of the time, and categorised as
‘More!’ 31% of the time, the latter with a goodness rating as high as 4.2, higher than that for the main choice, which was 3.9.

Table 5.8

Summary of the t-tests of Each Choice—Mandarin Learners to English Intonation

<table>
<thead>
<tr>
<th>Cantonese tone</th>
<th>Chosen English</th>
<th>Percentage</th>
<th>Df</th>
<th>t-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1 (T55)</td>
<td>More…</td>
<td>95</td>
<td>17</td>
<td>56.08</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 2 (T25)</td>
<td>More?</td>
<td>43</td>
<td>17</td>
<td>19.49</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More?!</td>
<td>48</td>
<td>17</td>
<td>8.58</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 3 (T33)</td>
<td>More…</td>
<td>78</td>
<td>17</td>
<td>36.33</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 4 (T21)</td>
<td>More.</td>
<td>65</td>
<td>17</td>
<td>28.71</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More!</td>
<td>31</td>
<td>17</td>
<td>7.62</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 5 (T23)</td>
<td>More?</td>
<td>29</td>
<td>17</td>
<td>6.69</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More?!</td>
<td>63</td>
<td>17</td>
<td>23.40</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Tone 6 (T22)</td>
<td>More.</td>
<td>59</td>
<td>17</td>
<td>18.55</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>More…</td>
<td>35</td>
<td>17</td>
<td>8.13</td>
<td>$p &lt; 0.001$</td>
</tr>
</tbody>
</table>

Table 5.9

Summary of the t-tests of Each Choice—Mandarin Learners to Mandarin

<table>
<thead>
<tr>
<th>Cantonese tone</th>
<th>Chosen Mandarin tone</th>
<th>Percentage</th>
<th>Df</th>
<th>t-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT 1 (T55)</td>
<td>MT 4 (T51)</td>
<td>66</td>
<td>17</td>
<td>28.537</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>MT 1 (T55)</td>
<td>31</td>
<td>17</td>
<td>19.382</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>CT 2 (T25)</td>
<td>MT 3 (T214)</td>
<td>49</td>
<td>17</td>
<td>11.762</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>MT 2 (T35)</td>
<td>44</td>
<td>17</td>
<td>8.543</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>CT 3 (T33)</td>
<td>MT 1 (T55)</td>
<td>93</td>
<td>17</td>
<td>39.281</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>CT 4 (T21)</td>
<td>MT 4 (T51)</td>
<td>78</td>
<td>17</td>
<td>26.836</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>CT 5 (T23)</td>
<td>MT 2 (T35)</td>
<td>84</td>
<td>17</td>
<td>18.249</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>CT 6 (T22)</td>
<td>MT 1 (T55)</td>
<td>61</td>
<td>17</td>
<td>9.074</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>MT 3 (T214)</td>
<td>33</td>
<td>17</td>
<td>6.341</td>
<td>$p &lt; 0.001$</td>
</tr>
</tbody>
</table>
Note: The total number of responses for each tone category was 360 (18 participants × 20 repetitions).

The symbols * (p < .001) show that the mean is significantly above the chance level (17%).

Figure 5.6. The Mandarin learners’ tonal categorisation percentage into English tunes for each Cantonese tone and its goodness rating in brackets.

In the tone groups with more than one category over the chance level (CT2, CT4, CT5 and CT6), three tunes were chosen a significantly higher number of times as the main category over the secondary choice (p < .001), while the difference between CT5 being categorised as ‘More?’ and ‘More?!’ was not significant (p = .178). Again, using the two criteria established previously in Section 5.2.1.4, the categorisation type for each Cantonese tone can be decided: five of the six tones count as ‘categorised’, while CT2 (T25) is ‘uncategorised’, as shown in Table 5.10. Regarding pairs CT1-CT3, CT4-CT6, which are each mapped onto the same category, a t-test for the goodness rating was performed. With goodness ratings significantly different from each other, the tone pairs formed by CT4 and CT6 are CG instead of SC. CT1 and CT3 count as SC, as their goodness ratings are not significantly different from each other. Tone pairs which include CT2 will UC pairs;
to be specific, depending on whether the pair shares overlaps, UC pairs are further grouped into UC-no overlap, UC-partial overlap and UC-same set.

Table 5.10

Summary of the Categorisations of the Six Cantonese Tones—Mandarin Learners to English Intonation

<table>
<thead>
<tr>
<th>Cantonese tones</th>
<th>English intonation</th>
<th>Status</th>
<th>Percentage; rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>More…</td>
<td>C</td>
<td>(95%; 3.8)</td>
</tr>
<tr>
<td>Tone 2</td>
<td>More?</td>
<td>U</td>
<td>(43%; 2.5)</td>
</tr>
<tr>
<td></td>
<td>More?!</td>
<td></td>
<td>(48%; 3.0)</td>
</tr>
<tr>
<td>Tone 3</td>
<td>More…</td>
<td>C</td>
<td>(78%; 3.5)</td>
</tr>
<tr>
<td>Tone 4</td>
<td>More.</td>
<td>C</td>
<td>(65%; 3.9)</td>
</tr>
<tr>
<td>Tone 5</td>
<td>More?!</td>
<td>C</td>
<td>(63%; 2.2)</td>
</tr>
<tr>
<td>Tone 6</td>
<td>More.</td>
<td>C</td>
<td>(59%; 3.1)</td>
</tr>
</tbody>
</table>

A summary of the tone contrast pair categories is given in Table 5.11, where T1-T4, T1-T5, T1-T6, T3-T4, T3-T5, T3-T6, T4-T5 and T5-T6 are TC pairs, T4-T6 is a CG pair, T1-T2, T2-T3, T2-T4 and T2-T6 are UC-no overlap, and T2-T5 is UC-same set. According to the predictions made by PAM, PAM-L2 and PAM-S, the discrimination of TC should be good, CG should be moderate, UC-no overlap should be good and UC-S should be poor.
Table 5.11

Summary of the Assimilation Patterns—Mandarin Learners to English Intonation

<table>
<thead>
<tr>
<th></th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 2</td>
<td>UC-no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 3</td>
<td>SC</td>
<td>UC-no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 4</td>
<td>TC</td>
<td>UC-no</td>
<td>TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 5</td>
<td>TC</td>
<td>UC-s</td>
<td>TC</td>
<td>TC</td>
<td></td>
</tr>
<tr>
<td>Tone 6</td>
<td>TC</td>
<td>UC-no</td>
<td>TC</td>
<td>CG</td>
<td>TC</td>
</tr>
</tbody>
</table>

When categorising the Cantonese tones into their L2 tone system (Mandarin), English speakers with Mandarin experience showed different patterns to that exhibited by Mandarin L1 speakers, as shown in Figure 5.7. The high-level tone (T55) was mainly categorised into the falling Mandarin tone and then into the high-level tone in Mandarin. The mid-level tone was matched onto the Mandarin high-level tone with high agreement (93%). The low-level tone was categorised mainly onto the falling-rising tone (T214), while 33% chose the high-level tone. The low-falling tone had the Mandarin falling tone as the dominant category, with an agreement of 78% and a goodness rating as high as 3.7. The high-rising tone was partially categorised into the Mandarin high-rising tone (44%) and the remainder chose the Mandarin falling-rising tone (49%). The low-rising tone had the Mandarin high-rising tone (84%) as the main category, but the goodness rating was relatively low (2.3).
Note: The total number of responses for each tone category was 360 (18 participants × 20 repetitions).

The symbols * \((p < .001)\) show that the mean is significantly above the chance level (20%).

**Figure 5.7.** The Mandarin learners’ tonal categorisation percentage into Mandarin tone system for each Cantonese tone and its goodness rating in brackets.

In the tone groups with more than one category over the chance level, the percentages of CT1 being categorised as MT1 and MT4 and CT6 being categorised into MT1 and MT3 are significantly different \((p < .001)\), while the differences between CT2 being categorised as MT2 and MT3 are not significant \((p = .183)\).

Table 5.12 illustrates the status of the six Cantonese tones when being categorised into the Mandarin tone system by L2 Mandarin learners: five of the six tones count as ‘categorised’, while CT2 (T25) is ‘uncategorised’. Regarding CT1 (T55), CT3 (T33) and CT6 (T22), which are all mapped onto the same category (MT1 [T55]), a \(t\)-test for the goodness ratings was performed. With goodness ratings significantly different from each other, tone pairs formed by CT1, CT3 and CT6 are considered CG instead of SC. Tone pairs which include T5 will be UC pairs; to be
specific, depending on whether the pair shares any overlaps, UC pairs are further grouped into UC-no overlap and UC-partial overlap.

Table 5.12

Summary of the Categorisations of the Six Cantonese Tones—Mandarin Learners

<table>
<thead>
<tr>
<th>Cantonese tones</th>
<th>Mandarin tones</th>
<th>Status</th>
<th>Percentage; Goodness rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT 1 (T55)</td>
<td>MT 4 (T51)</td>
<td>C</td>
<td>(66%; 3.7)</td>
</tr>
<tr>
<td>CT 2 (T25)</td>
<td>MT 2 (T35)</td>
<td>U</td>
<td>(44%; 3.0)</td>
</tr>
<tr>
<td></td>
<td>MT 3 (T214)</td>
<td></td>
<td>(49%; 3.2)</td>
</tr>
<tr>
<td>CT 3 (T33)</td>
<td>MT 1 (T55)</td>
<td>C</td>
<td>(93%; 3.8)</td>
</tr>
<tr>
<td>CT 4 (T21)</td>
<td>MT 4 (T51)</td>
<td>C</td>
<td>(78%; 3.7)</td>
</tr>
<tr>
<td>CT 5 (T23)</td>
<td>MT 2 (T35)</td>
<td>C</td>
<td>(84%; 3.4)</td>
</tr>
<tr>
<td>CT 6 (T22)</td>
<td>MT 1 (T55)</td>
<td>C</td>
<td>(61%; 2.4)</td>
</tr>
</tbody>
</table>

Table 5.13

Summary of the Assimilation Patterns—Mandarin Learners to Mandarin

<table>
<thead>
<tr>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 2</td>
<td>UC-no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 3</td>
<td>CG</td>
<td>UC-no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 4</td>
<td>TC</td>
<td>UC-no</td>
<td>TC</td>
<td></td>
</tr>
<tr>
<td>Tone 5</td>
<td>TC</td>
<td>UC-o</td>
<td>TC</td>
<td>TC</td>
</tr>
<tr>
<td>Tone 6</td>
<td>CG</td>
<td>UC-o</td>
<td>CG</td>
<td>TC</td>
</tr>
</tbody>
</table>

A summary of the tone contrast pair categorisations is presented in Table 5.13, where T1-T4, T1-T5, T3-T4, T3-T5, T4-T5, T4-T6 and T5-T6 are TC pairs; T1-T3, T3-T6, T1-T6 are CG pairs, T2-T3 and T2-T4 are UC-no overlap, and T2-T5 and T2-T6 are in the UC-overlap category. According to the predictions made by
PAM, PAM-L2 and PAM-S, the discrimination of TC should be good, CG should be moderate, UC-no overlap should be good, and UC-o should be moderate.

If we merge the categorisation patterns for Mandarin learners, derived both from their L1 (English intonation) and their L2 (Mandarin tones), then we arrive at the picture presented in Table 5.14. The conflicting results are reported with a slash: ‘/’. The category before the slash comes from categorisation according to the English intonation system, while that after the slash is derived from the Mandarin tone system. T1-T3 is SC in the L1 system but CG in L2; T1-T6 and T3-T6 are TC in L1 but CG in L2; T4-T6 is a CG pair when categorised into English, and TC when categorised into Mandarin. Two UC pairs have different grouping results as well: T2-T5 and T2-T6 are UC-same set and UC-no overlap in L1 respectively, but are both UC-partial overlap in L2.

Table 5.14

*Combination of English and Mandarin Categorisation—Mandarin Learners*

<table>
<thead>
<tr>
<th></th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 2</td>
<td>UC-no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 3</td>
<td>SC/CG</td>
<td>UC-no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 4</td>
<td>TC</td>
<td>UC-no</td>
<td>TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 5</td>
<td>TC</td>
<td>UC-s/UC-o</td>
<td>TC</td>
<td>TC</td>
<td></td>
</tr>
<tr>
<td>Tone 6</td>
<td>TC/CG</td>
<td>UC-no/UC-o</td>
<td>TC/CG</td>
<td>CG/TC</td>
<td>TC</td>
</tr>
</tbody>
</table>

5.4.3 Discussion

English speakers with intermediate learning experience determined the high-rising tone T25 as uncategorised, regardless of whether they categorised the L2 tones into their L1 English intonation system or the L2 Mandarin tones.
It is very surprising that the exact counterpart of the Cantonese high-level tone—the Mandarin high-level tone—was not the main choice for Mandarin learners. This is still probably due to their preference in L1 for the falling intonation ‘More.’, which is similar to the Mandarin falling tone (T51). By contrast, the categorisation for the mid-level tone (CT33) was the same as with native Mandarin speakers, who chose the Mandarin high-level tone (MT55) as the main category. When focusing on the choice of the low-level tone (CT22), the main choice was the falling-rising tone (MT214) (61%), which is the only Mandarin tone with a similar tonal height to the target low-level tone (T22). This choice is not used at all in L1 Mandarin speakers’ categorisation, which indicates that more Mandarin learners pay attention to tone height. Unlike L1 Mandarin speakers who categorised the low-falling tone into MT3 (T214), English speakers chose the Mandarin falling tone (MT51) as the most similar tone. This could be explained by L2 speakers being less familiar with a major allotone of the falling-rising tone, which is a low-falling tone. Still, this could also be due to their preference for statement intonation. The low-rising tone was categorised into MT2 (T35), while L1 Mandarin speakers showed dual patterns of MT2 (T35) and MT3 (T214).

Apart from analysing the data within the PAM framework, data were also submitted to other two analysis measures: fit index and the mapping-based degree of response diversity. These will be presented in the next section.

5.5 General Comparison

This section summarises and compares the Cantonese tone categorisation results from the three participant groups, who differ systematically in their tone language experience: Mandarin speakers with L1 tone experience; Australian English speakers without any (L1 or L2) tone language experience; and Australian English
speakers with L2 tone language experience from a ‘small’ tone space language (when compared to Cantonese). To compare the similarities and differences between the three speaker groups in categorising Cantonese tones, two additional analyses were undertaken: the fit index and the degree of diversity, following Wu et al. (2014). The fit index combines the mean percentage and goodness ratings and thus provides a clear picture of how each tone is categorised, while the degree of diversity examines how diverse each group’s choices are.

The fit index is the result of multiplying response rates and goodness ratings; the modal response category has the maximum index value. The larger the number, the closer the L2 tone is to the chosen L1 category. The fit index combines the mean percentage and rating score effectively, which makes it comprehensive when choosing the modal response. As usual, the mean percentage is the main focus of comparison.

Further, to determine the assimilation diversity, the degree of response diversity is calculated. This measurement was adopted by Wu et al. (2014). K’ (Koopman, Personal communication; Simpson, 1949), the diversity degree, is computed with the following formula:

$$K' = \frac{1}{\sum_{i=1}^{R} P_i^2}$$

In this formula, $R$ is the number of L1 tone categories. $P_i$ stands for the percentage that an L2 category ($i$) is assimilated. The larger the K’ value, the less similar the L2 category is to the modal category. The minimum K’ value is 1, showing that the L2 tone is consistently mapped onto a particular L1 category. The maximum K’ value is the number of L2 tone categories, which is six in the current case (as Cantonese has six tones).
The results of Mandarin learners categorising Cantonese tones into their L2 Mandarin category are compared with L1 Mandarin speakers’ results; while the way in which Cantonese tones are categorised into English categories by Mandarin learners is compared with L1 English speakers. The comparisons are thus divided as follows: Cantonese into Mandarin and Cantonese into English.

5.5.1 Categorisation by Mandarin speakers and Mandarin learners

A comparison can be made between native Mandarin speakers and the Mandarin learners when they both assimilate Cantonese tones onto the Mandarin tone system. The results show that many similarities exist between the two assimilation patterns with respect to the three level tones and the high-rising tone. For the Cantonese high-rising tone (T25), the two most popular choices were the Mandarin high-rising (T35) and dipping tones (T214). More Mandarin learners chose the dipping tone. The majority assimilated the mid-level tone to the Mandarin high-level tone, although the Mandarin learners were more consistent. Mandarin learners showed more confusion between the high-level tone (T55) and the falling tone (T51) when assimilating the high-level tone, while L1 Mandarin speakers were very consistent with the high-level tone. For the low-level tone, both groups favoured the high-level tone, while quite a few Mandarin learners (33%) chose the dipping tone (T214). This has a more similar F0 onset to the target level tone.

The other two tones (CT4 the low-falling tone and CT5 the low-rising tone) showed different assimilation patterns: for the low-falling tone (T21), most Mandarin speakers chose the Mandarin dipping tone (T214) with a few choosing the high-falling tone (T51); however, the majority of Mandarin learners chose the high-falling tone (T51). This can be explained in the following way: Mandarin learners do not have the option to assimilate tones using their phonological knowledge of allotones,
unlike L1 Mandarin speakers who have a knowledge of phonological assimilation built-in that can help them assimilate an L2 tone into an allotone of their L1 category. The situation with the low-rising tone is similar—the rising tone (T35) and the dipping tone (T214) were chosen equally by Mandarin speakers, while the majority of Mandarin learners assimilated it into the high-rising tone, being unable to discern the allotone (T35) from the dipping tone (T214). The assimilation patterns for these two thus provide evidence for the lack of phonological perception in L2 speakers with a tone language background.

However, CT4 (the low-falling tone) was categorised by Mandarin speakers mainly as MT3 (68%), with the other main category of MT4 (at 30%) having a higher rating score (3.5) than MT3 (3.3). Even with the fit index, the modal response was still MT3 but the rating score was taken into account when comparing them. For CT5 (T23), 40% categorised it as MT2, with a rating score of 3.4; 44% categorised it into MT3, with a lower score of 3.3. It would be even more difficult to choose the modal response in this case, while the fit index gives us an answer—MT3 (1.45), which is slightly higher than MT2 (1.36).

The fit index results for the two groups are given in Table 5.15. Here, the two groups who categorised Cantonese tones into Mandarin tone systems are compared: L1 Mandarin speakers and Mandarin learners. The modal answers (numbers in bold) are mostly different, the only matching answer was that given for CT3 (T33), the mid-level tone. Both groups chose the Mandarin high-level tone as the matching tone. For L1 Mandarin speakers, the modal answers generally share the same pitch contour with the target Cantonese tone. For Mandarin learners, the modal answers for CT1 and CT4 were both MT4. This is very interesting, as CT1 and CT4 do not share the same pitch contour or height, yet they are categorised into the same
Mandarin falling tone. A possible explanation could be participants’ preference for the falling intonation tune. However, when categorising CT1 (T55) into English intonation, these speakers with a L1 English background did not choose ‘More.’ as their primary answer. Instead they chose ‘More…’, showing their capacity to hear the level pitch.

The two Cantonese rising tones reveal exactly contrary answers by the two groups: the high-rising tone (CT2) was categorised into MT2 by Mandarin speakers and MT3 by Mandarin learners, while MT2 was chosen as the main category for the low-rising tone (CT5) by Mandarin learners and MT3 was the primary choice for Mandarin speakers. As discussed in Section 5.2.3, the fact that Mandarin speakers categorised both CT4 (T21) and CT5 (T23) into MT3 (T214), which has allotonic forms of T21 and T35, indicates phonological assimilation. For Mandarin learners, CT4 was categorised into MT4, while CT5 was categorised into MT2, which share similar pitch contours with the perceived Cantonese tones.

Table 5.15
Assimilation Fit of Cantonese Tones to Mandarin Tone Categories—Mandarin Listeners and Mandarin Learners

<table>
<thead>
<tr>
<th>Perceived as</th>
<th>Presented tones</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CT1(T55)</td>
<td>CT2(T25)</td>
<td>CT3(T33)</td>
<td>CT4(T21)</td>
<td>CT5(T23)</td>
<td>CT6(T22)</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>EM</td>
<td>M</td>
<td>EM</td>
<td>M</td>
<td>EM</td>
</tr>
<tr>
<td>MT1(T55)</td>
<td>3.59</td>
<td>1.15</td>
<td>0.27</td>
<td>0.00</td>
<td>2.31</td>
<td>3.53</td>
</tr>
<tr>
<td>MT2(T35)</td>
<td>0.23</td>
<td>0.00</td>
<td>1.78</td>
<td>1.32</td>
<td>0.26</td>
<td>0.00</td>
</tr>
<tr>
<td>MT3(T214)</td>
<td>0.00</td>
<td>0.00</td>
<td>1.09</td>
<td>1.57</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>MT4(T51)</td>
<td>0.00</td>
<td>2.64</td>
<td>0.00</td>
<td>0.00</td>
<td>0.55</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Note: bold numbers indicate fit index values for the modal responses * EM = English learners of Mandarin; M = Mandarin speakers.
As presented in Figure 5.8, the Cantonese tones are mapped differently onto the L1 Mandarin tone system by each participant group. With reference to K’, the most similar counterpart Cantonese category for Mandarin speakers is CT1, which has a value of nearly 1. The most distant Cantonese tone category for the Mandarin speakers is CT5, which has the largest mapping diversity. This partially supports that the prediction from PAM that CT5 is treated as an uncategorised tone, which means it has the most discrepancy from any L1 category. The degree of diversity informs us which L2 tone is perceived as most similar to the L1 category and which has the most diversity of assimilation. In this case, for Mandarin speakers, CT1 is perceived as the most similar L2 tone: it has a value of nearly 1 for K’. This makes sense, as it has almost the same F0 height and F0 contour as the Mandarin high-level tone. CT5 is determined as having the most diversity of assimilation answers, which aligns with the previous finding that CT5 is uncategorised.

![Degree of Diversity - Cantonese vs. Mandarin](image)

*Figure 5.8. Mapping diversity for the six Cantonese tones perceived by Mandarin speakers and Mandarin Learners.*
For the L1 English Mandarin learners, the Cantonese tones with the smallest K’ value is CT3, the mid-level tone. This means that the chosen category is the most similar to the modal answer. The largest K’ value lies in CT2, the high-rising tone, suggesting least agreement on categorising this tone. In general, the (L1 English) Mandarin learners have smaller K’ values than L1 Mandarin speakers, indicating that Mandarin learners are more consistent in categorising the Cantonese tones to the Mandarin tone systems. This finding probably relates to their similar proficiency in Mandarin, as they are all intermediate learners and have similar exposure to Mandarin tones.

For Mandarin speakers, the L2 tones are categorised to the most acoustically similar counterparts in most cases, showing that phonetic assimilation occurs quite frequently. According to the degree of diversity, only CT1 has a perfect counterpart, which means that even partial similarity can stimulate phonetic assimilation. Further, the fact that all three level tones are categorised into the same, and the only level, tone in Mandarin demonstrates that Mandarin speakers rely more on F0 contour as their primary cue. This replicates previous results indicating that listeners with tonal L1s are more sensitive to F0 contours than to other cues (Francis, Ciocca, Ma & Fenn, 2008; Gandour, 1983; Guion & Pederson, 2007; Huang & Johnson, 2010). However, the results also show that Mandarin speakers are able to distinguish between different F0 heights as they give different goodness ratings for the three level tones—3.9, 3.3 and 3.0 respectively—indicating that Mandarin listeners find CT1 the best fit for MT1. That is, even though they also choose MT1 for CT3 and CT6, they are aware that these sounds are less similar to those in their L1 system. According to the fit index scores, CT6 has higher scores than CT3, which means that CT6 is a better fit than CT3. The high score is affected by the greater percentage of
CT6 categorised as MT1 (79% vs. 70%). The degree of diversity results (K’) show that when assimilating rising tones (one of the allotonic variants of MT3), special perceptual difficulties are caused by the acoustic similarities between allotones. In contrast, Mandarin learners share similar patterns when categorising the three Cantonese level tones with MT1 as the chosen answer; however, in general lower goodness ratings were given by Mandarin learners than by Mandarin speakers. More competing choices were found with Mandarin learners and K’ scores are higher with CT1 and CT6 for this speaker group.

5.5.2 Categorisations by English speakers and Mandarin learners

The results from the Cantonese tone-categorisation study show systematic similarities between the monolingual English-speaking participants, and the L2 Mandarin-learning native speakers of English. For example, both groups categorised the high-level (T55) and the mid-level (T33) tones into the same English tune ‘More…’, in 95% and 78% of instances, respectively. The most consistently chosen category for the low-level tone (T22) was ‘More.’ (59%; 3.1), but the secondary category was ‘More…’ (35%; 2.9). For the high-rising tone (T25), the two rising intonations ‘More?’ (43%) and ‘More?!’ (48%) were both chosen, with the latter one having a slightly higher intonation and goodness rating (3.0 over 2.5). The low-rising tone was mainly categorised into ‘More?!’ (63% of the time). ‘More?’ was chosen 29% of the time. The low-falling tone (T21) was categorised into the statement intonation ‘More.’ 65% of the time, and into the exclamation “More!” 31% of the time, with a goodness rating as high as 4.2, higher than that of the main category (3.9).

In general, the three level tones were mostly assimilated into ‘More…’ in Australian English, which has a level pitch contour. English speakers make their
judgement according to pitch shape in relation to level tones. For the high- and mid-level tones, the assimilation patterns were also quite similar—the uncertain ‘More…’ was chosen most of the time. For the low-level tone, both groups shared similar choices, but more Mandarin learners chose the statement ‘More.’. The second biggest chosen category was ‘More…’.

For contour tones, the two speaker groups showed different assimilation patterns. Assimilation results for the high-rising tone by English monolinguals are not unified: ‘More?!’ (31%; 2.3) and ‘More…’ (31%; 2.8) received equal choices (31%), while some chose the falling tune ‘More.’ (19%; 2.0) or the high rising tune ‘More?’ (19%; 2.6). Mandarin learners, by contrast, mainly allocated answers to either ‘More?’ (43%; 2.5) or ‘More?!’ (48%; 3.0), which both have a rising contour. This participant group chose similar patterns but favoured ‘More?!’ (63%; 2.2) when assimilating the low-rising tone. By contrast, English monolinguals’ most frequent choice was the falling tune ‘More.’ (44%; 3.5), followed by the question ‘More?’ (31%; 3.3); a few also chose ‘More…’ (19%; 2.6). The last contour tone, the low-falling tone, seemed to be mostly assimilated into the statement ‘More.’, a simple falling tune; this preference was more robust for English monolinguals. In 31% of the cases, Mandarin learners preferred ‘More!’ (a rise-fall) as the chosen category.

This categorisation pattern leads to some very interesting observations: the most common choices by English monolinguals do not always share their pitch contour with the target tone. Conversely, the choices Mandarin learners make align with the match between pitch contours. For the low-falling tone (CT21), English speakers are very consistent—94% chose ‘More.’. Mandarin learners favoured ‘More.’ over ‘More!’, which also attracted 31% of the choices. This is most likely due to a feature of T21 itself, which has a slightly higher onset. The contour shape
might indeed be more similar to the second choice (‘More!’) perceived by Mandarin learners. In general, both speaker groups chose ‘unknown on several occasions for T33, T23 and T22.

Table 5.16

Assimilation Fit of Cantonese Tones to English Intonation Categories—English Listeners and Mandarin Learners

<table>
<thead>
<tr>
<th>Perceived as</th>
<th>Presented tones</th>
</tr>
</thead>
<tbody>
<tr>
<td>More?</td>
<td>0.00</td>
</tr>
<tr>
<td>More.</td>
<td>0.00</td>
</tr>
<tr>
<td>More!</td>
<td>0.18</td>
</tr>
<tr>
<td>More…</td>
<td>3.00</td>
</tr>
<tr>
<td>More?!</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Note: bold numbers indicate fit index values for the modal responses

Table 5.16 illustrates the fit index results by English monolinguals and English speakers who are Mandarin learners. It is clear that English monolinguals made far less use of the intonation patterns in their modal choices than the Mandarin learners did. The monolinguals used just two categories – ‘More.’ and ‘More...’ – as their modal choices. Unlike the results from Mandarin speakers and Mandarin learners, the modal answers are mostly the same for the two groups here, except for CT2 and CT5, the two rising tones. English monolinguals mainly categorised CT2 into ‘More…’ [H* H-L%] and CT5 into ‘More.’ [H* L-L%]. Neither of the two chosen categories has a rising contour, indicating that pitch contour might not be a significant cue for English monolinguals. By contrast, English speakers with tonal
experience categorised both CT2 and CT5 into ‘More?!’ [H* H-H%], which has an obvious rising contour. This might be evidence for how Mandarin learning experience has tuned English speakers’ way of perceiving L2 lexical tones, in particular their attention to the cue of pitch contour trajectory.

Figure 5.9 illustrates the results for K’ when categorising Cantonese tones into the English intonation system by English monolinguals and English speakers with L2-Mandarin experience. Generally, the K’ scores are higher when compared to those of the Mandarin tone system. The inference is that the Mandarin system is more comparable with the Cantonese system than the English intonation system to the listeners who participated in categorisation tasks. For English speakers, the most consistent categorisation is for CT4 (the falling tone), as CT4 closely resembles the falling tone in English. The tone that causes the most confusion is the high-rising (T25) tone, where the K’ value is almost as high as 4. This could be due to the two
English rising tones being similar to each other—one is $L^* H-H\%$ and the other is $H^* H-H\%$.

Mandarin learners similarly revealed the most disagreement on the high-rising tone, but with a much smaller $K'$ value, which was barely over 2. The smallest $K'$ values existed for the high-level tone with Mandarin learners, meaning that this categorisation result was the most similar to the modal answer. This aligns with Mandarin speakers when they categorise Cantonese tones into their L1 system, finding the high-level tone the easiest-to-categorise Cantonese tone. Mandarin learners have smaller $K'$ results than English speakers. A possible explanation for this might lie in their Mandarin training, which familiarises them with tones and makes them generally more sensitised to pitch variation as a consequence.

To conclude, applying different methods (fit index and diversity degree) leads to similar results. It is not surprising that the three different speaker groups show distinct categorisation methods, indicating that L2 perception is influenced by different linguistic experiences. The comparison between different groups is: 1) based on the categorisation types of PAM; and 2) based on the chosen category between Mandarin speakers and Mandarin learners, and English speakers and Mandarin learners. For Mandarin speakers and Mandarin learners, only one of the Cantonese tones is uncategorised, while for monolingual English speakers, two are uncategorised. For Mandarin speakers, the uncategorised tone is CT5 (the low-rising tone), which is categorised into both MT25 and MT214. For Mandarin learners, the uncategorised tone is CT2 (the high-rising tone), which has the dual categorisation pattern of ‘More?’ and ‘More?!’. This tone is uncategorised for monolingual English speakers as well and both groups show the same confusion pattern. In addition, they determined the low-level tone (T22) as uncategorised at the same time.
5.6 Summary

This chapter has reported how Mandarin and English speakers categorise Cantonese tones onto their native prosodic systems. Mandarin learners map the six tones onto both their native English intonation system and the L2 Mandarin tone system. The percentage of choices, along with goodness ratings, show that both Mandarin and English speakers use their tone/intonation systems to perceive L2 tones. The results also determined whether a Cantonese tone was categorised or uncategorised for different speaker groups. Further, based on the categorisation patterns, two non-native tones can constitute a TC, SC or CG pair if both are categorised. If one of tones in the pair is uncategorised, then the pair can constitute a UC-no overlap, UC-overlap or UCsame; or a UU-no overlap, UU-overlap or UU-same set if both are uncategorised, according to PAM-S. Fit index and degree of diversity were used to compare the categorisations between groups, indicating that Mandarin learners perform differently from monolingual English speakers in categorising Cantonese tones using English tunes, and also differently from L1 Mandarin speakers when categorising using Mandarin tones. On the whole, Mandarin learners generally categorised tones in a more similar way to monolingual native English speakers. Nevertheless, these learners found Mandarin tones to be more comparable to Cantonese tones, compared to English tunes. This chapter has provided predictions for the discrimination results based on PAM and PAM-S, which are the basis for the second part of the perception study, presented in the next chapter. Chapter 6 will present the methodology, results and discussion for the discrimination task.
Chapter 6: Discrimination of Cantonese Tones

The experiment presented in the present Chapter 6 is based on the PAM and PAM-S frameworks, and the model’s extensions to lexical tone perception (see Chapter 4.1). The discrimination study is based on the categorisation results presented in Chapter 5; the discrimination predictions have been made based on these categorisation results.

To briefly summarise the relevant information from Chapter 5, Section 5.1.2 shows that for L1 Mandarin speakers, Cantonese T1-T2, T1-T4, T2-T3, T2-T4, T2-T6, T3-T4 and T4-T6 are TC groups, T1-T3, T1-T6 and T3-T6 form into CG pairs, T1-T5 and T3-T5 are UC-no overlap, T4-T5 and T5-T6 are in the UC-overlap group, and T2-T5 is in the UC-same set group. The categorisation patterns by native English monolinguals are as shown in Section 5.2.2. For speakers with no tone experience, Cantonese tones T1-T4, T1-T5, T3-T4 and T3-T5 are TC groups, T1-T3 and T4-T5 form into CG pairs, T2-T4 and T2-T5 are UC-no overlap, T1-T2, T1-T6, T2-T3, T3-T6, T4-T6 and T5-T6 are UC-overlap, and T2-T5 is in the UC-same set group. Similarly, the results from English speakers with Mandarin learning experience (see Section 5.3.2) indicate that T1-T4, T1-T5, T1-T6, T3-T4, T3-T5, T3-T6, T4-T5 and T5-T6 are TC groups, T4-T6 forms into CG pairs, T1-T2, T2-T3, T2-T4 and T2-T6 are UC-no overlap, and T2-T5 is in the UC-same set group. Detailed descriptions of how Cantonese tone pairs are tagged with these PAM labels are shown in Tables 6.3, 6.7 and 6.10 for Mandarin, English and English speakers with Mandarin experience, respectively.

According to the predictions made by PAM, PAM-L2 and PAM-S (again see Chapter 4.1), the discrimination of TC should be good, CG should be moderate, UC-
no overlap should be good, UC-partial overlap should be moderate and UC-same set should be poor.

The current chapter investigates these speakers’ discrimination abilities and compare the results to the predictions made according to the categorisation results and theoretical frameworks. L1 Cantonese speakers are also included to discriminate their native tones, interpreted as ceiling results. Thus, our four speaker groups can be further divided into two groups: tone (Cantonese and Mandarin) and non-tone (English monolinguals and English speakers with tone experience).

6.1 Discrimination of Cantonese Tones by Tone Language Speakers

6.1.1 Methods

6.1.1.1 Participants

In addition to the same group of Mandarin participants who participated in the categorisation task (see Chapter 5), 20 age-matched Cantonese speakers were recruited to participate in the present discrimination study. These participants were born and raised in Hong Kong and were undergraduate students studying in Australia. No one from this group had been trained as musicians or had a self-reported language difficulty.

6.1.1.2 Stimuli

The stimuli used in the discrimination task were the three carrier syllables /baːp/ /biː/ and /buː/, which included the three most distinct vowels (occupying the corner positions) in the vowel space. These syllables were selected as none of the six tones in these syllables form a real Cantonese word (and thus do not have corresponding characters). One checked syllable (/baːp/) was included as all unchecked syllables with /aː/ can form into real words, according to Cantonese
Character Database\(^2\). This is a departure from most other studies of Cantonese tone perception and production where, typically, the stimuli used are real characters. The use of non-words may arguably provide participants with a more unbiased task, as the ability to access the verbal meaning of some or all syllables if they were real words might influence native Cantonese speakers’ perception (differences in word frequency may be particularly disruptive). Thus, the current experiment follows the common rule in perception studies in which nonsense words are used. The stimuli were recorded under the same conditions as the categorisation study (see Chapter 5.2.1).

Two female native Cantonese speakers were instructed to read the target syllables using the same tone as that for real words. In total, 108 tones (3 syllables × 3 tokens × 6 tones × 2 repetitions) were recorded. These stimuli were screened by another three native Cantonese speakers, who verified that these tones were categorisable.

6.1.1.3 Procedure

Participants were asked to discriminate and detect tones first and then to categorise and rate them (see Chapter 5). The discrimination task was given first to avoid categorisation responses influencing discrimination performance, following the procedure outlined in previous perception studies (So, 2011; So & Best, 2010).

For the discrimination experiment, an AXB forced-choice task was conducted. In an AXB task, listeners are asked to determine whether the tone in the middle token is the same as the first or the last one. In the current study, the participants were instructed to press the keyboard button ‘f’ if the middle was the same as the first and ‘j’ if it was the same as the last one. This procedure follows that

\(^2\) No significant difference has been found across vowels in discrimination accuracy \((p = 0.19)\) or production results where the same stimuli was used \((p = 0.78)\).
conducted for previous perception papers: ‘j’ and ‘f’ were selected as they are central on keyboards and are normally pressed with the index fingers.

The AXB discrimination tasks were presented via a Sony laptop, using the presentation program E-Prime 2.0 (Schneider, Eschman & Zuccolotto, 2007). The AXB discrimination focuses on listeners’ ability to distinguish paired individual tones, while the tone detection task assesses their ability to differentiate tones in context.

This discrimination task consisted of 360 trials in six different experimental blocks corresponding to the three target words (syllables) and two speakers: ‘baap’, ‘bi’ and ‘bu’ (i.e., 60 trials per block, blocked by syllable type and a repetition with a different speaker). Each block consisted of the fifteen combinations of six tone contrasts on the target word (T1-T2, T1-T3, T1-T4, T1-T5, T1-T6, T2-T3, T2-T4, T2-T5, T2-T6, T3-T4, T3-T5, T3-T6, T4-T5, T4-T6 and T5-T6) in four trial formats (AAB, ABB, BAA and BBA). The symbols ‘A’ and ‘B’ represent the two contrastive stimuli (tone categories) of the target word in the sentence, and the four trial formats refer to the order of A and B. Cantonese tones involve three level tones of different F0 registers and two rising tones with different rising ranges. In addition, each speaker has a slightly different formant range; thus, within each trial the three tones were produced by the same speaker, while the speakers were changed between trials.

6.1.1.4 Analysis

The accuracy for every tone pair was compared and analysed with a mixed-factor ANOVA, with the participant group as the between-subjects factor and the tone pair as the within-subjects factor. Post-hoc t-tests with a Bonferroni correction were applied to examine the different accuracy of grouped tone pairs. Discrimination
results were also compared based on the groupings according to categorisation results (see Chapter 5). This potentially worked as an access to test predictions from PAM and PAM-S (see Chapter 4.1).

### 6.1.2 Results

The discrimination results by Cantonese and Mandarin speakers are summarised in Figures 6.1 and 6.2. The Cantonese listeners’ mean percentage correction (92.8%) was significantly higher than that of the Mandarin listeners (77.8%) ($p < .05$). There was a 41-msec difference in the response time between the two speaker groups, which is not significant ($p > .05$) as indicated by a paired $t$ test. In general, for non-native discrimination (Mandarin speakers), discrimination of TC and UC-no overlap was the best, better than for CG and UC-partial overlap, with UC.same set being the worst. This confirmed PAM’s prediction about discrimination by L2 listeners. A mixed-factor ANOVA test was applied to the discrimination data with group as the between-subjects factor, and tone pair as the within-subjects factor. The results showed a significant effect of group, $F(1, 36) = 2359.507, p < .001$; a significant effect of tone pair, $F(14, 504) = 228.988, p < .001$; the group × tone pair interaction was also significant, $F(14, 504) = 70.143, p < .001$. 


Figure 6.1. The mean correct discrimination (in percentages) for each Cantonese tone pair by Mandarin listeners.

The ANOVA for the Mandarin group showed a main effect for tone pair in that the mean percentage correction of some tone pairs was significantly lower than that of others, $F(14, 270) = 174.686, p < .001$. Post-hoc $t$-tests with a Bonferroni correction for multiple $t$-tests further revealed that the mean percentage correction for

Figure 6.2. The mean correct discrimination (in percentages) for each Cantonese tone pair by native Cantonese listeners.
the CG-assimilated T1-T3 (71%), T1-T6 (73%) and T3-T6 (69%) was significantly lower \((p < .001)\) than for TC contrasts. UC-no pairs T1-T5 (87%) and T3-T5 (80%) were not significantly lower than the TC pair. While only one of the UC-partial overlap pairs (T4-T5) was significantly lower, the other one was not. The UC-same set pair T2-T5 (51%) was significantly lower than other tone pairs, \(p < .001\).

The asterisk (*) indicates the difference between the two groups is significant \((p < .001)\).

*Figure 6.3.* Mean discrimination of the category groups.

The discrimination of CGs by Mandarin and Cantonese speakers is presented in Figure 6.3; the discrimination score for TC was 84% (Mandarin) and 93% (Cantonese). This difference increased with the CG contrast (71% for Mandarin and 95% for Cantonese). Within the UC contrast, no overlap, overlap and SC also showed different patterns. For UC-no overlap, Mandarin speakers performed at 84%, while native discrimination was 96%. For the UC-overlap contrast, the discrimination decreased to 73% for Mandarin speakers and 92% for Cantonese speakers. The most difficult case for both groups was the UC-same set contrast, where Mandarin speakers only discriminated the contrast at chance level (51%) and Cantonese speakers achieved 78%. The results confirm our predictions (based on
PAM and PAM-S) that TC contrasts result in excellent discrimination, while CG has moderate to good discrimination accuracy. Further, within the UC groups, the assimilation with no overlap had better discrimination than UC with partial overlap; UC-same set (involving categorisation to the same set of native categories) fared the worst.

6.1.3 Discussion

The present study shows that Mandarin speakers discriminate Cantonese tones moderately well, likely due to their L1 language experience with tone. However, their difficulty with some particular L2 tone pairs also suggests that the L1 system can hinder second language tone perception.

The results support PAM-S’s predictions and a number of previous findings: TC contrasts are better discriminated than CG contrasts, while the discriminability of UC contrasts varies greatly depending on how the contrasts are perceived as overlapping (or not) with L1 categories (PAM-S). According to the categorisation of contrasts from Section 5.1.2, the 15 tone pairs are further grouped on the basis of the similarity/discrepancy between them. Those that are categorised into two L1 categories fall into TC; those that are categorised into one group fall into CG (there is a significant difference between the goodness ratings); those UC are further grouped into no overlap, partial overlap and same set. PAM posits that the more different the two tones are from each other, the better the discrimination will be. This study confirms that the perception of Cantonese by Mandarin speakers supports PAM’s predictions. This is similar to Reid et al.’s (2014) results, where PAM was applied to Mandarin speakers’ perception of Thai tones.

This study also highlights the fact that the chosen categorisation criterion of a given study systematically affects the predictions for perception, as different criteria
result in differences in the categorisation patterns (e.g. the differences in the %
categorised threshold, or the use of modal categories without a minimum
categorisation requirement beyond that). For instance, the present results indicate
that TC has better discrimination than CG, while the UC pair has significant within-
group variation (see also So & Best [2010]). Contrasts with a TC assimilation pattern
have excellent discrimination and so do those from UC-no overlap, as the two sounds
are quite different from each other. CG and UC-partial overlap have moderate to
good discrimination, while UC-same set has poor discrimination. The
suprasegmental extension of PAM, PAM-S (So & Best, 2014) provides detailed
predictions for UC pairs; thus, we can compare the results within the model in a
more detailed fashion. However, as stated above, the method of categorising the
contrasts systematically influences the discrimination. For example, So and Best
(2010) classified the Mandarin T2 (T35)- T3 (T214) contrast as a CG, while Hao
(2012) classified it as a TC, even though poor discrimination was found in both
cases. In the current study, Cantonese T25-T23 is classified as a UC
same set
contrast; thus, poor discrimination is expected, which contradicts Qin and Mok’s
(2011) results where CT2 (T25)-CT5 (T23) was classified as CG assimilation, and
thus a moderate to very good discrimination was expected. This discrepancy is due to
different ways of addressing the assimilation pattern, and this has a profound effect
on the predictions for the discrimination of the tone pairs that based on these
patterns.

When comparing L2 discrimination with L1 ones, the two speaker groups
share some similar patterns. First, the most difficult pair for both groups is T2-T5,
the two rising tones, which is a classic confusable pair in Cantonese. Previous studies
indicate that L1 adult speakers sometimes find this confusing; even children who are
learning Cantonese find it difficult and acquire it last (e.g., Ciocca & Lui, 2003; Mok, Zuo & Wong, 2013; Qin & Mok, 2014; Wong, Ciocca & Yung, 2009). This suggests that the T2-T5 confusion might be more difficult than other pairs regardless of listeners’ language backgrounds. This position finds some support in Mandarin T2-T3 confusion in studies with English, Cantonese and French speakers (Hao, 2012; So & Best, 2010).

The second most difficult pair for both tone language participant groups (L1 Mandarin and L1 Cantonese) is T3-T6, which can be explained acoustically, as T3 (T33) and T6 (T22) have the smallest F0 difference. This aligns with previous findings on Cantonese perception (Qin & Mok, 2014). The other two easily confused tone pairs are T4 (T21)-T5 (T23) and T4 (T21)-T6 (T22), as they share similar F0 onsets with slightly different F0 offsets. This may be explained as Mandarin listeners are still more sensitive to F0 contours than F0 heights and use fewer F0 offsets as the primary cue to discriminate. The preferred perceptual cue differs among speakers with different native languages. Tonal language speakers tend to pay more attention to tone contours; this includes Thai, Mandarin and Cantonese listeners.

6.2 Discrimination of Cantonese Tones by Non-tone Language Speakers

This section focuses on the discrimination results by English speakers with no tone experience and those with Mandarin learning experience.

6.2.1 Methods

The same groups of English monolinguals and Mandarin learners who participated in the tone categorisation study (see Chapter 5) took part in this task. All stimuli and procedures followed the same procedure as in discrimination experiment with tonal speakers (see Sections 6.1.2 and 6.1.3).
6.2.2 Results

The discrimination results for English monolinguals and English speakers are presented in Figure 6.4. Those for English speakers with Mandarin L2-learning experience are presented in Figure 6.5. There was a significant difference between the two groups with respect to the task reaction time measure ($p < .01$), such that English speakers with tonal experience were much quicker in responding to the experiment than monolingual English speakers. For general accuracy, English speakers’ mean accuracy in discriminating Cantonese tones was 71.9%, which was significantly worse than that of Mandarin learners (79.1%) ($p < .05$). A mixed-factor ANOVA test was applied to the discrimination data with group as the between-subjects factor and tone pair as the within-subjects factor. The results showed a significant effect of group, $F(1, 209) = 467.32, p < .001$. A significance effect of tone pair was indicated as well, $F(16, 367) = 764.44, p < .001$, the group × tone pair interaction was also significant, $F(16, 367) = 187.23, p < .001$.

For English speakers (see Figure 6.4), the discrimination results align with PAM’s prediction about discrimination by L2 listeners. When the two tones in a pair were both categorised, the predictions roughly supported PAM, as the mean accuracy for TC was 74.3%, with 69% for CG. However, tone pairs from TC were not always easier to discriminate than were those from CG: T1-T3 had an accuracy of 71%, which is the same with T3-T4 from TC. Within-group variation was larger when one of the tone pair is uncategorised. In the UC-no overlap group, T2-T4 was easy to discriminate (80%) while T2-T5 was difficult in comparison; the accuracy here was as low as 64%. According to PAM-S, contrasts classified as UC-no overlap should have good accuracy, as the two tones share no overlap. In relation to the UC-overlap, the difference was even larger: the accuracy of T1-T2, T1-T6, T2-T3 and T3-T6 was
above 70%; this number dropped to 68% on T4-T6, and lowered dramatically with T5-T6 (52%). The discrimination for the only UU pair—T2-T6—was the easiest for English speakers: 81% of the time, this pair was discriminated accurately.

The discrimination results by Mandarin learners are summarised in Figure 6.5. As two categorisation patterns are involved (from the English intonation system and the Mandarin tone system), the conflicted groups are expressed with a slash: ‘/’. These categorisations are discussed in detail in Chapter 5 (see Table 5.14). The group before the slash originates from the categorisation into English intonation and the group name after the slash comes from the categorisation into the Mandarin tone system. The accuracy of the tone pairs in TC ranged from 89% to 69%. Following the English categorisation results (the name before the slash), the SC (T1-T3) had a higher accuracy than several tone pairs from TC. Following the Mandarin categorisation, then tone pairs from CG (T1-T3, T1-T6) were still easier than TC.

Figure 6.4. The mean correct discrimination (in percentages) for each Cantonese tone pair by English listeners.
Results for the UC-no overlap group were generally excellent, all having an accuracy of above 80%. For tone pairs T2-T5 and T2-T6, UC-same set had a poor discrimination of 69% and UC-no overlap had excellent accuracy of 80%, following the categorisation into English intonation. Following the groups after the slash, then the two groups were both from UC-overlap and showed a significant variation of discrimination accuracy within the same group ($p < .001$).

![Discrimination of Cantonese Tones by Mandarin Learners](image)

**Figure 6.5.** The mean correct discrimination (in percentages) for each Cantonese tone pair by Mandarin learners.
**Figure 6.6.** Mean discrimination of the category groups.

The results of the tone discrimination task by English monolinguals and Mandarin learners, grouped by contrast type, is presented in Figure 6.6. The discrimination scores for TC were 74% (English), 79% (Mandarin learners, categorisation pattern into English), 79% (Mandarin learners, categorisation pattern into Mandarin); the difference was larger with the CG group (69% for English and 78% for Mandarin learners—English categorisation pattern, 77% for Mandarin learners—Mandarin categorisation pattern). There was one SC for Mandarin learners, with an accuracy of 76%. Within the UC group, the no overlap, overlap and same set categories also showed different patterns. For UC-no overlap, English speakers performed 72%, while Mandarin learners had 82% for categorising into English and 83% into Mandarin tones. For UC-overlap, the discrimination decreased to 70% for English speakers and 75% for EM into Mandarin tones. The UC-same set from EM into English tones was 69%. One UU group existed for English speakers, with an accuracy of 81%. The results confirm PAM and PAM-S’s predictions that TC assimilation has excellent discrimination for both groups. However, CG for
Mandarin learners had similarly excellent results, instead of moderate to good as predicted. Further, within the UC groups, it is only for Mandarin learners that the assimilation with UC-no overlap had better discrimination than UC-partial overlap, with UC-same set faring the worst. For English speakers, no great difference existed between UC-no overlap and UC-partial overlap. Finally, for English speakers, UU-partial overlap had excellent accuracy, confirming the predictions that the discrimination of UU is independent of a native system and should have fair to good accuracy.

For the English monolinguals, post-hoc $t$-tests with a Bonferroni correction for multiple $t$-tests further revealed that the only significantly different group comparisons are between UU and all other groups ($p < .001$). TC was not significantly higher than CG, nor was the UC-no overlap significantly higher than the UC-partial overlap. Similar procedures were repeated with Mandarin learners for both categorising into English and Mandarin. Under both conditions, no significant difference existed between TC, CG and SC. The UC-no overlap group was significantly than UC-same set ($p < .001$) when categorising into English intonations. Likewise, a significant difference was found within UC groups when categorising into Mandarin tones ($p < .001$).

6.2.3 Discussion

This task has revealed that English speakers with Mandarin experience outperform participants without any tone experience. This advantage of learning a second language is a significant contribution to our understanding of tone perception as it has not been systematically examined previously: only one previous study (Qin & Jongman, 2015) has demonstrated a similar L2 advantage when investigating the discrimination of just three Cantonese tones by Mandarin learners. The current study
confirms this finding and extends it by including all six tones with a larger corpus size.

For English monolinguals, the most difficult tone pair was T5-T6, where they passed with an almost chance rate (52%), followed by T2-T5 and T4-T5. As discussed in Section 6.1.3, the difficulty associated with T2T5 may be universal, supporting So and Best (2010), Hao (2011) and Qin and Mok’s (2011) findings. Interestingly, T5-T6 and T4-T5 are pairs with a similar pitch height but a different pitch contour (T23-T22, T21-T23), indicating that English monolinguals experience more problems when distinguishing pitch contours. This finding confirms previous findings in Gandour (1983, 1984), So and Best (2010), Hao (2011) and Ding et al. (2011) (the last regarding German speakers’ performance). For Mandarin learners, their discrimination ability for these three pairs was significantly better than that of English speakers ($p < .01$). However, these pairs were still the most difficult to discriminate for Mandarin learners.

While it is clear from the present study that L2 Mandarin learning experience improves English speakers’ ability in the discrimination of another tone system, interestingly, this L2 tone learning does not change English listeners’ perception difficulty patterns: despite their L2 Mandarin experience, L1 English participants continue to struggle with those tonal contrasts that are difficult to English monolinguals, but to a lesser degree. In contrast, Qin and Mok (2014) found that every tone pair involving the low-falling tone T4 (T21) had excellent discrimination. According to their categorisation, these pairs fell into TC contrasts. The researchers (Qin & Mok, 2014) concluded that the pattern confirmed the predictions from PAM.

For English speakers both with and without Mandarin learning experience, the discriminations results follow PAM-S’s predictions roughly within categorised
tone pairs: for English speakers, TC was better discriminated than CG, with SC being the most difficult subset. A significant group variance existed within UC pairs: UC-no overlap, UC-partial overlap and UC-same set had quite different discrimination scores. The fact that the distinction within the categorised group (CG, SC, TC) was not significant does not strictly support PAM-S. Unlike the suggestion in PAM that UC pairs always have moderate discrimination, great variance existed within the UC group. The discrimination accuracy depends on the similarity between the pair’s counterparts and how each of these relates to an L1 prosodic system. In general, PAM/PAM-S’s predictions are supported more by Mandarin speakers than English ones.

6.3 Summary

This chapter has presented results from a Cantonese tone discrimination study with participants from four language backgrounds: native Cantonese speakers, Mandarin speakers, monolingual English speakers and L2-Mandarin learning English speakers. The results show that native Cantonese speakers discriminate their L1 tones the best (91.9%), followed by Mandarin learners (79.0%), Mandarin speakers (77.8%) and English speakers (71.9%). Having learned a tone language improves the ability to perceive L2 tones, but the difficulty pattern remains unchanged. Discrimination results for Mandarin speakers give more support to the predictions from PAM/PAM-S. The UC pairs need further categorisation, as not all UC pairs are equally well perceived. Along with Chapter 5, this chapter has presented a comprehensive investigation of L2/L3 speakers’ ability in perceiving Cantonese tones. Chapter 7 will focus on the production of Cantonese tones, including the four analysing methods that examine L2 production thoroughly.
Chapter 7: Production of Cantonese Tones

As discussed in Chapters 2 and 3, the influence of a native prosodic system on the perception of pitch is supported by several studies (Grabe et al., 2003; Ulbritch, 2008). Gandour (1983), for example, found that English speakers focus on pitch height when perceiving non-native tones, while Cantonese listeners pay attention to both pitch height and pitch contour. As a consequence, we might predict that English speakers would have difficulty in perceiving tones with similar pitch height but with different contours. This prediction has been confirmed by two studies (Hao, 2011; So, 2006), which found that English speakers had trouble discriminating a Mandarin tone pair differing in pitch contour but with similar pitch height. Similar results have been found for German listeners (Ding et al., 2011). It is also well known that general psychoacoustic features universally influence speakers’ perceptions, regardless of language background; for example, the similarity and distance between the two L2 tones (Burnham et al., 2014). Further, it is suggested that having a tonal language background does not automatically make L2 perception of another tone language easier, although the error patterns are steadier (Peabody & Seneff, 2009).

The perception study results presented in Chapters 5 and 6 confirm the influence from the native prosodic system as well as from L2 learning experience. For Mandarin speakers, their way of perceiving Cantonese tones was very similar to their own language, but they had more problems with pitch height as it is not a salient cue in their native language (Qin & Mok, 2011; Wang et al. 2003). For English speakers, they could perceive tones in the same way they perceive intonation (Gandour, 1983). English listeners have experience with pitch via post-lexical
accentuation and intonation so they categorised tones onto their intonation system by interpreting them as post-lexical pitch accents.

In addition to the influence from L1, L2 experiences (English L1 and Mandarin learning as L2) contributed to L3 perception (Cantonese tones). The categorisation results (Chapter 5) suggest that Mandarin learners’ perception of Cantonese tones is influenced by both their native English intonation system and their Mandarin learning experience. The discrimination study (Chapter 6) indicates that this learning experience led to a better performance than that of either the Mandarin or the monolingual English speakers. We thus agree with the proposal that when the L2 and L3 belong to the same language typology, L2 modulates L3 perception (Qin & Mok, 2015). However, little is known about the influence of linguistic experience on L3 production.

While it is established that L2 perception and production are related (Chapter 4.1.3 and 4.2.2), the exact nature of this relationship needs to be determined. This chapter attempts to extend perception findings to production and seeks to uncover whether 1) tone production is influenced by L1 in the same way as in perception; 2) L2 experiences assist L3 production; and 3) the relationship between perception and production. Production of Cantonese tones by the four speaker groups will be compared across the following four aspects: 1) tone differentiation, as reported through scatterplots of F0 onsets and offsets; 2) tone contour, which is F0 movement over time; 3) tone duration; and 4) native auditory judgements.

7.1 Method

7.1.1 Participants

The same four speaker groups that participated in previous tasks took part in this task: native Cantonese speakers, native Mandarin speakers, native English
speakers with no tonal experience, and native English speakers with Mandarin learning experience. For detailed description, see Section 6.1.1.1.

7.1.2 Stimuli

The same non-word stimuli /baːp/, /biː/, /buː/, as recorded in the discrimination study (see Section 6.1.1.2), were applied in the current production task.

7.1.3 Procedure

An imitation task was conducted to investigate speakers’ production of Cantonese tones. In this study, participants heard one of the Cantonese target syllables and then produced the target. The experiment was conducted in E-Prime 2.0 on a laptop computer, and all speakers were recorded at the MARCS Institute recording studio, with a head-mounted microphone (Sennheiser SC230ML). The recording order of the 54 tokens (3 syllables × 3 repetitions × 6 tones) was randomised.

7.1.4 Data analysis

As discussed in Chapter 2.2, the primary phonetic correlates of tone are F0 height, F0 movement and duration. A number of analytical methods were thus applied to include all of the three phonetic correlates. An important step to be undertaken before analysing the tones is normalisation: both F0 and duration must be normalised.

7.1.4.1 Normalisation

To establish a model of production by speakers with different language backgrounds, certain procedures should be performed beforehand. As discussed previously in Chapter 2, F0 is the primary cue for Cantonese tones. However every speaker has his or her unique pitch range, which makes it almost impossible to have
identical F0 patterns produced by different speakers. As such, any type of inter- or intra-speaker variation in F0 must be eliminated first. Even with these differences, L1 speakers can still recognise tonal speech as produced by different people. The aim of normalisation is thus to imitate the perceptual process, removing the variant individual differences as much as possible without losing the invariant acoustic features. In the current study, two types of normalisation were applied.

7.1.4.1.1 Duration normalisation

For duration normalisation, the longest F0 contour in each category was first identified and others were lengthened to this duration. This was done to preserve all F0 information. The lengthening technique adopted is the enhanced pitch-synchronous overlap-and-add (Boersma & Weenink, 2009), which alters the duration without changing the pitch values. This procedure will affect the investigation of duration, but it enables the possibility of linking observed perceptual patterns with the F0 dimension. It also provides us the possibility to investigate tone movements over time, which will be presented in Section 7.1.4.4.

7.1.4.1.2 F0 normalisation

i. Intra-speaker normalisation

Tones are generally assumed to be divisible into three parts: an onset, a central element (nucleus) and an offset. A tone nucleus model has recently been proposed by Zhang and Hirose (2004) and Wang et al. (2008) to perform intra-speaker normalisation. Under this model, the full F0 of a syllable can be divided into three parts: onset trajectory, tone nucleus and offset trajectory. The tone nucleus is the essential element, which includes the tone’s main pitch contour. The onset and offset courses carry articulatory transitions, which depend greatly on context. The tone nucleus is indicated as being quite stable, and is barely influenced by
neighbouring elements or stress and intonation. Thus, focusing on the tone nucleus can help extract a tone’s crucial F0 information.

Tone nuclei were identified manually during the segmentation process with Praat 5.3, where only the vowel production is regarded as the nucleus. This enables intra-speaker normalisation, as it avoids F0 variations arising from different aspects of a language.

ii. Inter-speaker normalisation

After duration normalisation, the pitch values were extracted with Praat 5.3 and R 2.15, using the autocorrelation method, with ranges set differently for female and male speakers (70–400Hz for female, 50–300Hz for male). To obtain a relative value for better comparison, each F0 value was converted from Hz to a logarithm-based T-value, using the formula stated below (Ladd, Silverman, Tolkmitt, Bergmann & Scherer, 1985, Peabody & Seneff, 2009, Rose, 1987; Wang et al., 2003):

\[ T = \frac{\lg X - \lg L}{\lg H - \lg L} \times 5 \]

In this formula, \( \lg \) means the log value. X is the log pitch value at the measurement point, L the lowest and H the highest pitch produced by the speaker. The T-value ranges from 0 to 5, corresponding to Chao’s (1930) tone system. In the current formula, 0 represents the lowest pitch (when X = L) and 5 is the highest (when X = H). This method of transforming pitch values into numbers enables easier comparison between speakers.

7.1.4.2 Plots of F0 onsets and offsets

To describe the perceptual differences between tones, five dimensions may be used (Gandour, 1978): 1) average pitch, 2) direction, 3) length, 4) extreme endpoint and 5) slope. A plot of the F0 offset and onset of the nucleus can include all
information excluding length. Rising tones would be expected to cluster closer to the y-axis, while falling tones would cluster closer to the x-axis. Ellipses around each tone type can be calculated by determining the distribution of points around a mean for each tone. F0 offset versus F0 onset for all speech tokens were plotted and grouped according to tone type into ellipses. The relative pitch values at the onset and offset time points have been extracted from Praat 5.3 and plotted with R 2.1.5 (package ggplot2 [Wickham, 2016]). These ellipses encompassed approximately 95% of the projections on to each axis. They provide a visual summary of the degree of differentiation between the six tonemes. The more different these ellipses are from each other, the better the production accuracy is. This kind of approach to tone production study can be used to observe the differences between groups of speakers. A variance test was also performed with R 2.1.5 on F0 onsets and offsets by all four speaker groups to investigate the production consistency.

In addition to the visual results, three numerical analyses were provided to illustrate how tones are differentiated within the tonal space and among tonemes, following the analyses in Barry and Blamey (2004), where they investigated tone production by children with cochlear implants. Here, the parameters calculated are the lengths of the axes, the areas of the tonal ellipses, and the distances between the centre points of each ellipsis. Based on these data, two indices are proposed for measurement.

7.1.4.3 Measuring the tonal space

To perform this analysis, the three most differentiated tones are first identified. Drawing a line to link the centre points of these tones will result in a triangle representing approximately the speaker’s F0 range. In Cantonese, the most differentiated tones are CT1, CT2 and CT4 (T55, T25 and T21 in Chao numbers,
respectively). Thus, the tonal space being compared is the area formed by these three centre points.

7.1.4.3.1 Index 1—Measuring tone differentiation within the tonal space

Tonal differentiation across the tonal space is a function of the area of the total tonal space and the span of the triangle \((A_{e1,2,4})\) mentioned above, and represented in the following formula:

\[
\text{Index 1} = \frac{At}{Ae1,2,4}
\]

\(At\) represents the area of the tonal space for each tone category, \(A_{e1,2,4}\) represents the area of the triangle. When the result is > 2, an overlap among these tone ellipses is unlikely. If the result is \(\leq 2\), an overlap is likely to occur. The higher the number, the more differentiated the tones.

7.1.4.3.2 Index 2—Measuring differentiation among tonemes

All the ellipses have different lengths on the x- and y-axis; these can be used to describe the degree of variation in pitch used for each tone. Thus, the distances between the ellipses’ centres determine the difference of the average pitch between each tone. Index 2 is the result of the average \((\text{Ave})\) of the lengths of the two axes \((x1+2 = x\text{ axis} + y\text{ axis})\) for the six tones against the average distance of the centres of the six tone ellipses from each other \((\text{Ave Dist.})\), which can be represented as:

\[
\text{Index 2} = \frac{\text{Ave Dist.}}{\text{Ave Ax1} + 2}
\]

Index 2 exhibits several differences from Index 1: the speaker’s pitch range relies on all six tones rather than three; additionally, it is sensitive to differences in pitch height and contour individually.

The results of these two indices can then be analysed statistically to summarise the observed differences between groups and calculate the statistical significance of these differences. The strength of this methodology is that it makes
no pre-assumption about whether speakers can produce tones correctly or not, making it suitable for comparing production by different groups of speakers, especially by L2 speakers and pre-linguistically deafened people whose tonal production abilities are unknown. This can provide answers to questions such as whether ellipse plots are significantly different from each other, or whether L2 speakers differentiate tones as well as do native speakers. However, as this method only examines F0 onsets and offsets, it overlooks how pitch moves over time. Thus, a further measure to investigate the dynamic features of tones is crucial, which will be presented in the next section (Section 7.1.4.4).

7.1.4.4 F0 at different time points

The other approach in the current production analysis is to examine the F0 height at different time points, which is crucial for contour tones, as in the Cantonese tone system. This study followed the traditional analysis of measuring the F0 of every ten percentage points of duration, providing 11 data points for each token. This method also includes duration normalisation for better comparison of the F0 height and contour among different speakers and syllables. A two-way ANOVA was performed on all four groups’ tone values at the 11 timepoints. ‘Timepoint’ is the within-subject factor, while ‘group’ is the between-subject factor. Further Tukey HSD tests were performed to investigate the difference between speaker groups in relation to individual tones. Another series of ANOVAs was performed on each speaker group, with ‘timepoint’ and ‘tone type’ as the two factors. Similarly, Tukey HSD tests were performed to investigate the tone differentiation by different groups.

7.1.4.5 Duration

As duration is the secondary perceptual cue for listeners, most previous papers have only studied F0 information with normalised tone durations. However,
the current study will investigate whether tones are associated with durational variation. Before duration normalisation, the duration of the nucleus vowel was extracted with Praat 5.3 and further boxplotted with R 2.1.5 (packages emuR [Winkelmann et al., 2017] and ggplot2 [Wickham, 2016]) in three vowel types, four participant groups and six tones. A number of post-hoc t-tests with Bonferroni corrections were performed between the three non-native groups against the native one. For native production, a correlation test was also performed between the midpoint of F0 and duration to test the relationship between F0 height and duration.

7.1.4.6 Auditory analysis

All L2 tone production data were ultimately examined by native-speaker judges. Two native Cantonese speakers (one male and one female, both born in Hong Kong and who had completed their undergraduate education with a linguistics major from the Chinese University of Hong Kong) were invited to provide perceptual judgements. As noted above, they had received formal linguistic training and were familiar with Cantonese tone labels. They were provided with 58 sound files (20 native Mandarin speakers, 20 native English monolinguals and 18 Mandarin learners) and an answer sheet to record their perception of each tone. No participant identity was released to the judges. They were instructed to listen to all the tokens in each file and label the tone number. They could re-listen to the production as many times as they wanted to, and were paid at an hourly rate.

7.2 Results

A few tokens were manually eliminated due to a creaky or an over-breathy voice quality. The token numbers for different speaker groups are given in Table 7.1. For Cantonese, Mandarin and English speakers, the total token number was 360 (20
speakers x 6 tones x 3 repetitions); for Mandarin learners, the total token number was 324 (18 speakers x 6 tones x 3 repetitions).

Table 7.1

<table>
<thead>
<tr>
<th></th>
<th>Cantonese speakers</th>
<th>Mandarin speakers</th>
<th>English speakers</th>
<th>Mandarin learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>353</td>
<td>342</td>
<td>342</td>
<td>314</td>
</tr>
<tr>
<td>/i:/</td>
<td>340</td>
<td>354</td>
<td>352</td>
<td>311</td>
</tr>
<tr>
<td>/u:/</td>
<td>339</td>
<td>351</td>
<td>337</td>
<td>318</td>
</tr>
</tbody>
</table>

7.2.1 Tone differentiation

Figures 7.1 to 7.4 illustrate vowel /a/, and no significant difference exists between vowels across groups ($p = 0.78$). Detailed F0 onsets and offsets results can be found in in Appendix D, Figures D.1 to D.6. Figure 7.1 shows that even L1 Cantonese speakers have some within-tone category variation—they do not produce tones at the same position every time. However, minimal overlap exists between the tone ellipses, except for a small portion between T33 and T22, the mid- and low-level tones. By contrast, the ellipses in Figures 7.2 and 7.3 indicate a significant overlap for both Mandarin speakers and English speakers. It is quite difficult to separate the six tone ellipses for English speakers’ tone productions. The ellipses in Figure 7.4 (English speaking Mandarin learners) are more separate than are those for either English or Mandarin speakers. The Mandarin learners’ ellipses are not as discrete as are those of L1 Cantonese speakers, but most of the six tones are recognisably distinct.

Another very interesting finding arises from the data variance: Cantonese speakers have the least variance in both F0 onsets and offsets, meaning that they are
quite consistent in producing the tones. This makes sense, as they are the L1 speakers ($\sigma = 0.62$ and $1.01$ for onsets and offsets respectively). English speakers have much greater variance ($\sigma = 0.93$ and $1.61$ for onsets and offsets) than all other groups, making them least consistent in reproducing tones. Mandarin learners have a slightly more stable performance ($\sigma = 0.86$ and $1.50$). Mandarin speakers have the most consistent pattern apart from L1 speakers, with a variance result of $0.74$ and $1.19$. We can infer from this that tonal language speakers are better at repeating the same tone category than non-tone language speakers, although this proposal would need further investigation to be conclusive.

![Figure 7.1. Tone production by Cantonese speakers.](image)
Figure 7.2. Tone production by Mandarin speakers.

Figure 7.3. Tone production by English speakers.
7.2.1.1.1 Tonal space

The tonal spaces (as defined earlier in Section 7.1.4.3) formed by the three most distant tones (T55, T25 and T21) were calculated based on the relative onset and offset values of the centre points. Cantonese speakers had the largest tonal space at 3.89, followed by English speaking Mandarin learners (3.06), which was slightly larger than for Mandarin speakers (3.0). English speakers had the smallest tonal space (1.35).

7.2.1.1.2 Tone differentiation within the tonal space

Table 7.2 presents the results of Index 1, where the smaller the number is, the more differentiated the tones are. L1 speakers have the smallest number across all six tones, indicating that each tone takes up quite a small part of the entire tonal space, leading to the least amount of tonal confusion. For non-native speakers, English speakers learning Mandarin have smaller values than both Mandarin and English speakers, except for T21, where Mandarin speakers have the smallest value. All non-
native speaker groups present the least amount of tonal confusion on T55, which is probably due to the high-level tone being the most consistently categorised tone to regardless of language background. Mandarin speakers differentiate tones more effectively than do English speakers across all categories except for T23, where English speakers (just) outperform Mandarin speakers. This may be because of the fact that Mandarin only has one high-rising tone; thus, Mandarin speakers tend to produce the low-rising tone with a higher F0 offset.

Table 7.2

Results of Index 1

<table>
<thead>
<tr>
<th></th>
<th>Cantonese</th>
<th>Mandarin</th>
<th>English</th>
<th>EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>T55</td>
<td>0.35</td>
<td>0.87</td>
<td>0.73</td>
<td>0.41</td>
</tr>
<tr>
<td>T33</td>
<td>0.29</td>
<td>1.11</td>
<td>1.58</td>
<td>0.84</td>
</tr>
<tr>
<td>T22</td>
<td>0.21</td>
<td>1.74</td>
<td>2.02</td>
<td>0.99</td>
</tr>
<tr>
<td>T25</td>
<td>0.37</td>
<td>1.32</td>
<td>2.07</td>
<td>0.57</td>
</tr>
<tr>
<td>T23</td>
<td>0.23</td>
<td>2.03</td>
<td>1.99</td>
<td>1.49</td>
</tr>
<tr>
<td>T21</td>
<td>0.24</td>
<td>0.65</td>
<td>1.76</td>
<td>1.27</td>
</tr>
</tbody>
</table>

7.2.1.1.3 Tone differentiation between tonemes

As shown in Figure 7.5, the higher the number for Index 2, the greater the difference between tonemes. This index represents the distances between ellipse centres. Clearly, L1 speakers have the best tone differentiation for this measure, while English speakers show the least differentiation between tonemes, which suggests significant overlaps between tonemes. English speakers with Mandarin learning experience have a slightly higher Index 2 score than Mandarin speakers, indicating that they differentiate tones slightly better than do Mandarin speakers.
Figure 7.5. Results of Index 2

7.2.2 Tone movements

The tone trajectories of the four speaker groups are given in Figures 7.6 to 7.9. The patterns of the production of Cantonese tones by L1 Cantonese speakers (Figure 7.6) are very similar to those found in previous studies and are consistent with the representations from Chao’s system for Cantonese tones. Some overlap can be found for T25, T21, T23 and T22 before 20% into the syllable, as they share similar F0 onsets. The two rising tones T25 and T23 are almost identical up until 30% of the duration, after this, T25 rises higher. There is greater difference between T55 and T33 than between T33 and T22, though the latter two tones are still easy to separate. T21 starts to drop from about the 20% time point. From offsets, we can see that T23 and T33 have a lot of overlap from 70% of the duration to the end. T25 and T55 have similar F0 offsets, which are around 4.5. Numeric T-values are given in Appendix E, Table E.1.

Mandarin speakers (see Figure 7.7) have quite different production patterns. For example, T55 has a lower pitch level than for L1 speakers and the discrepancy between T55, T33 and T22 is much smaller compared to L1 speakers, especially for
T22 and T33, which are fairly close to each other. Such a small difference could potentially cause perceptual confusion. The two rising tones T25 and T23 have different F0 onsets where they should be similar. However, the Mandarin speakers clearly make the distinction between the rising slopes, although these still differ from L1 speakers’ production. The falling contour has a less sheer fall before 70% and then drops sharply from there until the end. However, it has a very high F0 onset, at around 3, possibly due to the L1 falling tone having quite a high onset. In general, Mandarin speakers are quite accurate in terms of contour, but much less accurate in their sensitivity to pitch height.

![Diagram of tonal contour by Mandarin speakers.](image)

*Figure 7.7. Tonal contour by Mandarin speakers.*

As shown in Figure 7.8, English speakers behaved differently and tended to produce every tone in a level shape—their production of the three contour tones T25, T23 and T21 all have an F0 change range of less than 2. However, the three levels are very L1-like: they had a similar level shape and pitch height, and the difference between T33 and T22 is still recognisable, with a discrepancy of about 1. However, the onset area around 2 is very crowded—it is even difficult to distinguish between T21, T23 and T22 before 30% point of the duration. The high-rising tone T25 has a higher onset (about 3); however, the low-rising tone is produced in a more L1-like way, partly due to low-rising tone having fewer F0 changes. T21, interestingly, is
produced with a sharp drop at around 70% of the duration. In general, English speakers exhibit more sensitivity to pitch height than do Mandarin speakers, as they have better separation of the three level tones. However, their performance for contour tones is much poorer in terms of both pitch height and contour.

Figure 7.8. Tonal contour by English speakers.

Figure 7.9 illustrates the fact that English-speaking Mandarin learners have fewer problems than either Mandarin or English speakers. Surprisingly, they have the most L1-like trajectory of the six tones. In terms of level tones, their high-level tone is higher than the maximum of L1 Cantonese speakers. Their T33 is right above the 3 value and T22 is a bit lower than 2, quite close to those produced by native speakers. Among the other three contour tones, T25 and T21 have similar F0 onsets but they proceeded in opposite directions. The other rising tone T23 has a lower onset but finally finishes with the same offset as T33. Roughly speaking, this production map is quite robust in terms of tone distinctions, as each tone has a clear path and little overlap with other tones, though the earlier parts of T23 and T22 are still very difficult to separate.
A two-way ANOVA was performed on all four groups’ tone values at eleven timepoints. Timepoint is the within-subject factor, while group is the between-subject factor. The results reveal that for all tones, group is a significant influencing factor ($p < .001$). For contour tones (T25, T21 and T23), timepoint is a significant influencing factor ($p < .001$), where significant tone movement is expected.

Further, Tukey HSD tests revealed the difference between speaker groups in relation to individual tones. For T55, except for English and Mandarin speakers, all groups are significantly different from each other ($p < .001$). For T25, Mandarin and Cantonese speakers shows no significant difference. The biggest difference can be found for English and Mandarin speakers, where the $p$-value is $< .05$. For T33, except for English speakers and English Mandarin learners, all other groups differed from each other, with Mandarin and Cantonese speakers having the biggest difference of 0.36. For T21, the English speakers had the only significant difference to Cantonese speakers: a difference of 0.44 ($p < .05$). For T23, significant differences were found between all three L2 groups and the L1-speaker groups ($p < .001$). For T21, significant differences were limited to those between Cantonese and English speakers, and Cantonese and Mandarin speakers.

Figure 7.9. Tonal contour by English speakers with Mandarin learning experience.
Another series of ANOVAs was performed on each speaker group, with timepoint and tone type as the two factors. For Cantonese and Mandarin speakers and English learners of Mandarin, tone type was a significant factor: each tone was different from the other. Timepoint, tone type (and its interaction Tone × Type) were all influencing factors \((p < .001)\). For Mandarin speakers, \(F(5, 45) = 40.575\), T33 and T25, T22 and T25, T22 and T23 were not significantly different from each other. For English speakers, timepoint was not a significant influencing factor, indicating that they failed to show significant pitch movement along time.

Tukey HSD tests indicated that for Cantonese speakers, all tones were significantly different from each other at \((p < .001\), except for T22-T21: \(p < .05)\). The only non-significant pair was T23-T33 \((p = .35)\). For Mandarin speakers, half of the tone pairs were significantly different from each other \((p < .001)\), with the most similar pairs being T22-T33 \((p = .89)\) and T23-T25 \((p = .42)\). For English speakers, most tones were significantly different from each other \((p < .01)\). For them, the most difficult pairs were T22-T25, T22-T21 \((p = .03)\), and T22-T23. For English learners of Mandarin, most tones can be differentiated \((p < .001)\), although they found T22-T21, T22-T23, and T33-T25 slightly more difficult to differentiate in production.

The observation from perception studies (Chapter 5 and 6) that non-tonal speakers are more sensitive to pitch height and that tonal speakers pay more attention to pitch contours is supported by current production findings. Further, speakers from a tonal language background still have better production ability than those with no previous tonal experience, given the evidence from tonal space and tone differentiation indices. However, the current study establishes the fact that L3 speakers with L2 tonal experience (as with the English learners of Mandarin in this study) perform better than both English and Mandarin speakers. This indicates that
L2 experience can be transferred as well as L1 experience. In this case, the L1 English experience helped with participants’ sensitivity to pitch height; at the same time, their L2 experience with Mandarin tones tuned their ability towards tonal contours.

7.2.3 Tone duration

The duration of the tone is the time scale on the horizontal axis, measured in milliseconds (Bauer & Benedict, 1977). In the current chapter, measurements of the time span of vowels are regarded as the tone duration. The vowels and tones were segmented and labelled using Praat 5.3. Analysis was performed by R 2.1.4 with the emuR package. Production of the six Cantonese tones in three vowels /a i u/ by four groups of speakers are summarised in Table 7.3. The duration in the tables and figures is given in milliseconds. The boxplots of the duration differences are given in Appendix F, Figures F.1 to F.4.

The data from the four groups showed some consistency: /i/ had the longest duration (510ms for Cantonese speakers, 725ms for Mandarin speakers, 672ms for English and 621ms for Mandarin learners), which was followed by /u/ (490ms for Cantonese speakers, 667ms for Mandarin speakers, 642ms for English and 588ms for Mandarin learners), with /a/ being the shortest (437ms for Cantonese speakers, 577ms for Mandarin speakers, 595ms for English and 542ms for Mandarin learners). Further, regardless of vowel differences, Cantonese speakers always produced tones with the shortest duration and Mandarin learners the second shortest. By contrast, tones produced by Mandarin and English speakers were much longer than the other two groups. Mandarin speakers were the longest on vowels /i/ and /u/ whereas English speakers performed a longer duration than Mandarin speakers on /a/.
Table 7.3

Mean Duration of the Produced Tones by Different Speakers

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Speaker Groups</th>
<th>Tone55</th>
<th>Tone25</th>
<th>Tone33</th>
<th>Tone21</th>
<th>Tone23</th>
<th>Tone22</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>C</td>
<td>419.48</td>
<td>456.70</td>
<td>433.70</td>
<td>358.37</td>
<td>502.68</td>
<td>451.08</td>
<td>437.00</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>521.49</td>
<td>602.66</td>
<td>585.74</td>
<td>519.86</td>
<td>600.65</td>
<td>630.27</td>
<td>576.78</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>489.95</td>
<td>637.18</td>
<td>616.17</td>
<td>518.20</td>
<td>646.77</td>
<td>663.58</td>
<td>595.31</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>487.48</td>
<td>551.61</td>
<td>602.29</td>
<td>424.27</td>
<td>550.19</td>
<td>637.38</td>
<td>542.20</td>
</tr>
<tr>
<td>/i/</td>
<td>C</td>
<td>487.89</td>
<td>538.54</td>
<td>523.67</td>
<td>440.22</td>
<td>554.59</td>
<td>514.35</td>
<td>509.88</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>678.48</td>
<td>745.66</td>
<td>733.65</td>
<td>704.24</td>
<td>733.71</td>
<td>755.24</td>
<td>725.16</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>633.88</td>
<td>694.13</td>
<td>673.99</td>
<td>650.18</td>
<td>711.75</td>
<td>667.24</td>
<td>671.86</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>581.43</td>
<td>681.74</td>
<td>595.21</td>
<td>592.21</td>
<td>633.53</td>
<td>644.22</td>
<td>621.39</td>
</tr>
<tr>
<td>/u/</td>
<td>C</td>
<td>476.96</td>
<td>536.22</td>
<td>518.03</td>
<td>381.55</td>
<td>528.58</td>
<td>498.94</td>
<td>490.05</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>646.82</td>
<td>659.29</td>
<td>673.24</td>
<td>660.08</td>
<td>685.36</td>
<td>678.55</td>
<td>667.22</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>649.54</td>
<td>639.06</td>
<td>641.84</td>
<td>618.71</td>
<td>661.63</td>
<td>640.93</td>
<td>641.95</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>616.41</td>
<td>596.42</td>
<td>605.33</td>
<td>545.27</td>
<td>577.99</td>
<td>588.13</td>
<td>588.26</td>
</tr>
</tbody>
</table>

Note: C = Cantonese speakers, M = Mandarin speakers, E = ES, EM = English speakers who are Mandarin learners, numbers in bold are the longest and shortest values in each row.

To compare the duration of L2 production with native ones, a number of post-hoc t-tests with Bonferroni corrections were performed between the three L2 groups against the L1 one (see Table 7.4). The results suggest that both Mandarin and English speakers produced Cantonese tones significantly longer than Cantonese speakers ($p < .001$). By contrast, Mandarin learners only produced T21 significantly differently from L1 speakers ($p < .001$), yet they still had the shortest T21 of the three L2-speaker groups, which was the closest to L1 production. Separate analyses were then performed between Mandarin and English speakers to see whether their productions differed from each other. The results showed no significant difference between the duration produced by Mandarin and English speakers. Regarding duration, English speakers with Mandarin learning experience produced the tones in the most L1-like way. No significant difference was found between the production
by English and Mandarin speakers—both groups tended to produce tones longer than did L1 speakers, especially the falling tones.

Table 7.4

*Mean Duration for Each Tone Type and t-scores with Bonferroni Corrections*

*between Multi-group Comparisons*

<table>
<thead>
<tr>
<th></th>
<th>T55</th>
<th>T25</th>
<th>T33</th>
<th>T21</th>
<th>T23</th>
<th>T22</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Mean</td>
<td>461.44</td>
<td>510.49</td>
<td>491.80</td>
<td>393.38</td>
<td>528.62</td>
</tr>
<tr>
<td></td>
<td>t-scores</td>
<td>615.60</td>
<td>669.20</td>
<td>664.21</td>
<td>628.06</td>
<td><strong>673.24</strong></td>
</tr>
<tr>
<td>M</td>
<td>Mean</td>
<td>591.12</td>
<td>656.79</td>
<td>644.00</td>
<td>595.70</td>
<td><strong>673.38</strong></td>
</tr>
<tr>
<td></td>
<td>t-scores</td>
<td>4.48*</td>
<td>6.41*</td>
<td>5.32*</td>
<td>7.07*</td>
<td><strong>4.01</strong></td>
</tr>
<tr>
<td>E</td>
<td>Mean</td>
<td>561.77</td>
<td>609.92</td>
<td>600.94</td>
<td><strong>520.58</strong></td>
<td>587.24</td>
</tr>
<tr>
<td></td>
<td>t-scores</td>
<td>3.06</td>
<td>3.44</td>
<td>3.38</td>
<td>3.85*</td>
<td>1.45</td>
</tr>
<tr>
<td>EM</td>
<td>t-scores</td>
<td>1.244</td>
<td>0.688</td>
<td>1.153</td>
<td>1.33</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Note: numbers in bold are the longest and shortest values in each row, asterisk* means p < .001).

C = Cantonese speakers, M = Mandarin speakers, E = ES, EM = English speakers who are Mandarin learners, M&E = Mandarin and English speakers

Upon merging the vowel groups and calculating the mean values of each tone type, we can see that the longest tones were either T23 (Cantonese and English speakers) or T22 (Mandarin speakers and Mandarin learners), and the shortest were either T21 (Cantonese speakers and Mandarin learners) or T55 (Mandarin and English speakers). For the native production by Cantonese speakers, the low-rising tone (T23) had the longest duration (529 ms), slightly longer than the other rising tone (T25, 510 ms), while the falling tone was the shortest (393 ms). The three level tones had medium duration, with the mid-level tone being the longest, the low-level tone the second and the high-level tone the shortest. A comparison with previous
studies is given in Table 7.6. The rank of the last three tones is as in Kong (1987), whereas some contradictions can be found in the three tones with the longest duration. However, the longest tone (T33) in the current study is the same as in Fok (1974), and is longer than the high-rising tone, which is the opposite of Kong (1987). The rank of the level tones aligns with Kong: T33>T22>T55, which does not follow Gandour’s (1977) conclusion about the inverse relationship between F0 and duration. Table 7.5 illustrates the comparison of the duration ranking from the three studies.

Table 7.5

Summary of the Duration Rank

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st (longest)</td>
<td>T23</td>
<td>T25</td>
<td>T23</td>
</tr>
<tr>
<td>2nd</td>
<td>T25</td>
<td>T33</td>
<td>T25</td>
</tr>
<tr>
<td>3rd</td>
<td>T33</td>
<td>T23</td>
<td>T22</td>
</tr>
<tr>
<td>4th</td>
<td>T22</td>
<td>T22</td>
<td>T33</td>
</tr>
<tr>
<td>5th</td>
<td>T55</td>
<td>T55</td>
<td>T21</td>
</tr>
<tr>
<td>6th (shortest)</td>
<td>T21</td>
<td>T21</td>
<td>T55</td>
</tr>
</tbody>
</table>

Note: T = tone

Further, a correlation test was performed between the midpoint of F0 and duration, revealing a correlation coefficient of -.143 (p-value = .423), meaning that in terms of level tones, F0 and duration are not significantly related. As the two rising tones start at similar pitch heights (T23 and T25), we applied the F0 value at the endpoint to investigate whether F0 and duration are inversely related in the case of rising tones. A negative relationship was thus confirmed (r = -.637, p = .004). In L1 Cantonese tone production, duration varies significantly between tones. In general, rising tones have the longest duration, followed by level tones, and falling tones have
the shortest duration. Roughly speaking, higher pitched tones have shorter duration. This relationship is more consistent with rising tones than with level tones. As for level tones, T22, which has relatively lower F0, has a shorter duration than T33.

7.2.4 Auditory analysis

All L2 production tokens were perceptually analysed by two L1 judges. The judgement results were then compared with the intended tone. The error rates, along with the best-produced tones and the worst tokens, are summarised in Table 7.6. In this table, the heading was the intended tone label. The misperceived tone was presented followed with the number standing for error rates, separated by groups. Generally, the performance of Mandarin learners was the best according to auditory judgement: the mean error rate was 33%, which is better than Mandarin speakers (38%). English speakers were considered the most difficult to identify by the two L1 judges—only 56% of the produced tones were accurately identified. The easiest tone for all three groups was the high-level tone (T1): the error rates are as low as 8% for Mandarin speakers and Mandarin learners. On the basis of previous results, this tone is most consistently categorised, discriminated and produced by all groups. The most difficult tone to be identified was the low-rising tone, regardless of the speaker group. A tonal confusion pattern is further summarised in Table 7.7, according to the most easily confused tones in Table 7.6.
### Table 7.6

**Auditory Analysis of Non-native Productions**

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>T3,8</td>
<td>T5,20;</td>
<td>T6,31;</td>
<td>T6,37;</td>
<td>T2,73;</td>
<td>T3,20;</td>
</tr>
<tr>
<td>MS</td>
<td>/i/</td>
<td>T3,17;</td>
<td>T5,21;</td>
<td>T6,43;</td>
<td>T6,42;</td>
<td>T2,67;</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>T3,13;</td>
<td>T5,26;</td>
<td>T6,38;</td>
<td>T6,53;</td>
<td>T2,65;</td>
</tr>
<tr>
<td>mean</td>
<td>13</td>
<td>30</td>
<td>48</td>
<td>44</td>
<td>68</td>
<td>26</td>
</tr>
<tr>
<td>/a/</td>
<td>T3,21;</td>
<td>T5,41;</td>
<td>T6,31;</td>
<td>T6,39;</td>
<td>T4,35;</td>
<td>T5,29;</td>
</tr>
<tr>
<td>ES</td>
<td>/i/</td>
<td>T3,17;</td>
<td>T5,39;</td>
<td>T3,26;</td>
<td>T6,29;</td>
<td>T6,39;</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>T3,19;</td>
<td>T5,31;</td>
<td>T6,21;</td>
<td>T6,31;</td>
<td>T4,43;</td>
</tr>
<tr>
<td>mean</td>
<td>31</td>
<td>55</td>
<td>37</td>
<td>44</td>
<td>67</td>
<td>26</td>
</tr>
<tr>
<td>/a/</td>
<td>T3,8</td>
<td>T5,23;</td>
<td>T6,22;</td>
<td>T6,35;</td>
<td>T4,31;</td>
<td>T4,9</td>
</tr>
<tr>
<td>EM</td>
<td>/i/</td>
<td>T3,9;</td>
<td>T5,19;</td>
<td>T6,13;</td>
<td>T6,25;</td>
<td>T4,37;</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>T3,14;</td>
<td>T5,31;</td>
<td>T6,19;</td>
<td>T6,21;</td>
<td>T4,44;</td>
</tr>
<tr>
<td>mean</td>
<td>14</td>
<td>37</td>
<td>30</td>
<td>39</td>
<td>61</td>
<td>14</td>
</tr>
</tbody>
</table>

Note: all numbers stand for percentage (%) of incorrectly perceived tones

In Table 7.7, the intended tones are listed in the first row, with the most common mis-identifications by the native judges listed according to participant group in the following rows. Interestingly, the mis-identified tones are quite similar for all three L2 groups. Regardless of the speakers’ background, the high-level tone was misperceived as the mid-level tone, the high-rising tone was misperceived as the low-rising tone, the mid-level tone as the low-level tone and the low-falling tone was mostly misperceived as the low-level tone. This situation might be due to a
phonological/allophonic relationship between the target tones and the misidentified tone categories, for native speakers.

The other tone targets showed a different pattern. For the low-rising tone, participants’ backgrounds seemed to influence their productions: Mandarin speakers’ T23 was mostly misheard as the rising tone, while for English speakers (regardless of tone experience), this was mostly misperceived as the low-falling tone. The confusion patterns were more diverse for the low-level tone: Mandarin speakers tended to produce it more as a mid-level tone; English monolinguals’ productions were mostly misperceived as the low-rising tone; Mandarin learners tended to insert a falling shape on this level tone.

Table 7.8

Tone Confusion Patterns

<table>
<thead>
<tr>
<th>Intended Tone</th>
<th>T1(T55)</th>
<th>T2(T25)</th>
<th>T3(T33)</th>
<th>T4(T21)</th>
<th>T5(T23)</th>
<th>T6(T22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>T3(T33)</td>
<td>T5(T23)</td>
<td>T6(T22)</td>
<td>T6(T22)</td>
<td>T5(T25)</td>
<td>T3(T33)</td>
</tr>
<tr>
<td>ES</td>
<td>T3(T33)</td>
<td>T5(T23)</td>
<td>T6(T22)</td>
<td>T6(T22)</td>
<td>T4(T21)</td>
<td>T5(T23)</td>
</tr>
<tr>
<td>EM</td>
<td>T3(T33)</td>
<td>T5(T23)</td>
<td>T6(T22)</td>
<td>T6(T22)</td>
<td>T4(T21)</td>
<td>T4(T21)</td>
</tr>
</tbody>
</table>

7.3 Discussion

Production of Cantonese tones by L1 Cantonese speakers, L1 Mandarin speakers, monolingual English speakers, and L1 English Mandarin learners were investigated with four different analytical methods. In terms of three dimensions examined in the tone differentiation analysis: tonal space, tone differentiation within tonal space, and tone differentiation between tonemes, the six Cantonese tones are best differentiated by L1 Cantonese speakers, Mandarin learners, followed by L1
Mandarin speakers, and English monolinguals (Section 7.2.1). The examination of tone dynamics over time suggests that L1 Mandarin speakers tend to exaggerate the pitch range for the low-falling tone (CT21) and low-rising tone (CT23), while English monolinguals produce contour tones in a level fashion. L1 English Mandarin learners produce the six tones in the most native-like way (Section 7.2.2). Similar results are found with duration analyses (Section 7.2.3) and native judgement (Section 7.2.4), that Mandarin learners are better than Mandarin speakers, with English speakers being the least able to produce accurate Cantonese tone contrasts. The current study’s results support the observations from perception studies that Mandarin speakers are more sensitive to pitch contour while English speakers pay more attention to pitch height, confirming previous findings. L1 prosodic systems influence L2 tone production greatly, as well as they do in perception.

Speakers coming from a tonal language background can produce tone contrasts more accurately than speakers with no prior tonal experience, according to the evidence from tonal space and tone differentiation indices presented in this study. However, the fact that Mandarin learners perform better than both English and Mandarin speakers indicates that L2 experience can be transferred as well as L1 experience. The L1 English experience of Mandarin learners helps with their sensitivity to pitch height; at the same time, their L2 experience with Mandarin tones tunes their ability towards tonal contours and possibly to tonal height as well, which yields possibilities for future investigation. Future work should extend the investigation to speakers with other language backgrounds (e.g., L2 English learners) to see whether a non-tonal L2 language assists L3 production.

L1, as well as L2, experience with tonal languages has a great effect on tone production. The results here extend the findings of L2/L3 perception to the
production domain. Mandarin speakers are less accurate in their production of Cantonese tones that share the same tonal contour but have different heights for the L1 tone system compared to the L2 system. Mandarin speakers tend to exaggerate pitch movement and have more problems with tones of medium pitch height. For the two rising tones, Mandarin speakers perform the best on T25; this is likely due to the pitch range and pitch height being the closest to native production. T25 is more similar to Mandarin speakers’ L1 rising tone, which is a high-rising T35. The low-rising tone as produced by Mandarin speakers has a more dramatic rise than it should have, which could be due to Mandarin speakers only dealing with rising tones that have large movements. The falling tone shows a dramatic change for Mandarin speakers as well—their L1 falling tone has a much steeper fall than the Cantonese falling tone, which may explain their Cantonese production.

English monolinguals with no experience with tonal languages are quite sensitive to pitch height and perform better on level than on contour tones. They tend to produce tones with less pitch movement. English speakers tend to produce all tones relatively level; in general, they exhibit small pitch movements for all tones, regardless of the tonal shapes. A possible reason for this is that they are much less sensitive to tonal contours; as such, they are less capable of producing them.

English-speaking Mandarin learners, in contrast, can combine their L1 sensitivity to pitch height and L2 experience with pitch contour. They exhibit a quite stable performance across all six tones: they are not as good as Mandarin speakers at pitch change on level tones, or as English monolinguals on pitch height, but they are better than these two speaker groups in the more challenging tone contrasts. That is, they have the most L1-like production for the low-falling tone.
The tone movement analyses in this thesis support the conclusion that L1, as well as L2, experience with tonal languages has a great influence on tone production. The present study shows that Mandarin speakers are less accurate in their production of Cantonese tones sharing the same tonal contour but with different heights for the L1 tone system compared to the L2 system. In addition, English speakers who have no experience with tonal languages are quite sensitive to pitch height and perform better on level than on contour ones. English-speaking Mandarin learners, in contrast, can combine their L1 sensitivity to pitch height and L2 experience with pitch contour. This study contributes to the field of L2 tone production and the influence of tonal experiences on producing L2 and in addition, L3 tones. More research is required to make solid conclusions regarding how L1 and L2 experiences tune L3 production at the same time.

### 7.4 Combining Tone Perception and Production

Combined with perception results I will firstly discuss the relationship between non-native tone perception and production; secondly the existing individual differences in both perception and production will be discussed.

#### 7.4.1 Relationship between Tone Perception and Production

The performances by each individual speaker are illustrated in Figure 7.10. In general, the percentage of correctly discriminated tones was higher than the tones produced correctly, as judged by L1 speakers for all three groups. This is undoubtedly influenced by the nature of the tasks: the discrimination task was digital, while the production task was analogue. Additionally, this could be possible evidence of perception preceding production.
The perception and production performances by the three non-native speaker groups are summarised in Figure 7.11 to 7.13. Mean perception and production scores for Mandarin speakers are 77.8% ($SD = 6.65$) and 61.8% ($SD = 5.21$) respectively. English speakers achieved 71.9% ($SD = 5.13$) and 55.8% ($SD = 8.74$) for the perception and production of Cantonese tones. English learners of Mandarin perceive and produce L3 at 79% ($SD = 7.76$) and 67.5% ($SD = 4.87$). For both perception and production, English learners of Mandarin exhibit the best performance, followed by Mandarin and then English speakers. All three groups show great individual differences in both tasks.

According to a series of Pearson’s correlation analyses, a strong positive link between perception and production is apparent with Mandarin speakers ($r = .71$, $p < .001$) and English learners of Mandarin ($r = .84$, $p < .001$). The correlation is stronger with English learners of Mandarin. No direct correlation exists between the perception and production by English speakers ($r = .05$, $p > .5$).
Correlations between the perception and production of non-native lexical tones can be found then with listeners who have had no contact with the target tone system but have had experience with tones in either L1 or L2. By contrast, this performance is uncorrelated for participants without tonal experience. This study suggests that tonal experiences (either from L1 or L2) influence the perception and production of a new tone system.

Figure 7.11. Correlations between perception and production by Mandarin speakers.

Figure 7.12. Correlations between perception and production by English speakers.
Another method of examining the link between perception and production is to investigate the most difficult tone pairs in perception and production respectively, and determine whether a correspondence exists between the two modalities.

The most badly discriminated pairs are those with the lowest discrimination scores. For Mandarin speakers, the three most difficult tone pairs are T2-T5, T3-T6 and T1-T3; for English speakers, T5-T6, T2-T5 and T4-T5; for Mandarin learners, T2-T5, T5-T6 and T4-T5. The most difficult to discriminate tone pairs are thus the same for English speakers and Mandarin learners, but the order of difficulty is different in the two cases.

The most difficult tones to produce will be based on the error rates of each tone, and the counterpart with which it was misperceived by native judges. For Mandarin speakers, the most three difficult produced tones are T2-T5, T1-T3 and T4-T6; for both English speakers and Mandarin learners, these are T4-T5, T4-T6 and T2-T5. Table 7.9 illustrates these results, with the different tone pairs in italics. For each group, one production tone pair cannot be explained by perceptual difficulty. For Mandarin speakers, T4-T6—a poorly produced pair—has an intermediate discrimination rate, 79%. The poorly discriminated T1-T3 (71%) has a good
production result (mis-identified for only 17% times by native judges). For English speakers, the T5-T6 pair was not one of the most difficult to produce. But T4-T6 was correctly discriminated just 68% of the time. For Mandarin learners, it was discriminated well at 78%. In the production task, T5 (T23) is not the most confusing counterpart for T6 (T22); neither is this the case when the situation is reversed. This kind of comparison could provide potential evidence for the idea that perception and production are not directly linked—some difficulties in tone production cannot be explained by perceptual performance. However, since some tone pairs exhibit difficulties in both perception and production tasks, perception and production are still evidently linked in some way, as supported by the previous discussion.

Table 7.8

Tone Difficulty by Different Speaker Groups

<table>
<thead>
<tr>
<th>Most difficult</th>
<th>MS</th>
<th>ES</th>
<th>EM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived</strong></td>
<td>T2-T5, T3-T6, T1-T3</td>
<td>T5-T6, T2-T5, T4-T5</td>
<td>T2-T5, T5-T6, T4-T5</td>
</tr>
<tr>
<td><strong>Produced</strong></td>
<td>T2-T5, T3-T6, T4-T6</td>
<td>T4-T5, T4-T6, T2-T5</td>
<td>T4-T5, T4-T6, T2-T5</td>
</tr>
</tbody>
</table>

Note: MS = Mandarin speakers, ES = English Speakers, EM = English speakers who are Mandarin learners.

Table 7.8 highlights the fact that, given the differences in perception and production by the three participant groups, some difficulties are shared by all participants (e.g., T2-T5), although ranked differently by each. This supports Burnham et al.'s (2014) contention that universal and language-specific factors combine during the L2 perception process. The current findings extend this notion to production.
7.4.2 Individual differences

As discussed earlier, great individual variance is apparent within each speaker group, as the standard deviations are quite high. The variance across speaker groups in perception and production tasks is compared, and the summary of these values is given in Table 7.9. Interestingly, for Mandarin speakers and Mandarin learners, more variance exists in perception than in production, while the opposite is true for English speakers. This observation supports the point that familiarity with tone influences L2 tone production: tone experience could be a key component in maintaining stable production.

Table 7.9

<table>
<thead>
<tr>
<th>Variance (σ²)</th>
<th>Perception</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>44.24</td>
<td>27.14</td>
</tr>
<tr>
<td>ES</td>
<td>26.26</td>
<td>76.35</td>
</tr>
<tr>
<td>EM</td>
<td>60.14</td>
<td>23.67</td>
</tr>
</tbody>
</table>

Note: MS = Mandarin speakers, ES = English speakers, EM = English speakers who are Mandarin learners

The fact that all speaker groups’ variances are over 20 indicates that great individual differences are present. Figures 7.14 to 7.16 illustrate the individual speakers’ performances on both perception and production tasks. Both Mandarin speakers and Mandarin learners show a positive correlation between the two modalities as a group, as discussed before. It is interesting that when speakers’ perception performances are ordered from lowest to highest, their production performances vary. For example, for Mandarin speaker 10 to Mandarin speaker 13, their perception scores fell in the middle range, while their production scores were
quite low. Mandarin speakers 14, 17 and 20 are the three most successful learners, as both their perception and production exhibit the best range. However, it should be noted that Mandarin speaker 20, the highest discrimination scorer, did not perform the best on production. Likewise, Mandarin speaker 14, the best producer, did not perform the best on perception.

Figure 7.14. Mandarin speakers’ individual performances on perception and production.

Figure 7.15. Mandarin learners’ individual performances on perception and production.
For Mandarin learners (Number 9 in particular) perception is moderate but production is very low. Mandarin learners 14 to 18 are effective learners: their perception exceeded 85% and their production was higher than 70%. Still, the best perceiver and producer are different speakers. Even within these excellent learners, some fluctuation can be observed: Mandarin learner 16 has a higher discrimination score but a lower production compared to Mandarin learner 15. In the lower discrimination score range, Mandarin learner 4 is an interesting case: not perceiving well, but being well perceived by L1 speakers regarding production. Thus, in some cases, a participant’s ability to correctly perceive non-native tones is not entirely matched by to his/her ability to also produce the tone correctly.

Figure 7.16. English speakers’ individual performances on perception and production.

For English speakers who show no correlation between perception and production as a group, their individual performances have greater variance and the mismatch between perception and production is more obvious. For instance, English
speakers 12 and 16 have great perception, but their productions are extremely poor. This is the opposite to the perceptions of English speakers 3 and 5, who discriminate tones poorly but can produce them very well. However, a few speakers can be defined as super learners, as they perform equally well on both perception and production (e.g., English speakers 15 and 19). Likewise, there are speakers who are generally bad with Cantonese tones (e.g., English speakers 6 and 7).

Even in the two groups showing strong positive correlations between perception and production, individual speakers’ performances do not match the correlation all the time. A good perceiver does not have to be a good producer and vice versa. Sometimes, an overall correlation is found between perception and production, but when examining an individual speaker’s performance, no relationship is established.

7.5 Summary

This chapter has described the production study comprehensively, first overviewing the methodology and data preparation, then presenting and discussing the results in four sections: tone differentiation, tone movements, duration and L1 auditory judgement. Mandarin learners were found to have better differentiation, larger tonal space, the most L1-like duration and the highest auditory judgement. English monolinguals had the smallest tonal space and the most overlap between tone categories. For tone movements, Mandarin speakers behaved in a quite different way to speakers from participants with English-language backgrounds: they exaggerated tonal contours with less attention to pitch height. By contrast, English speakers attended more to pitch height but tended to lose pitch contour at some time. The results have consistently demonstrated the Mandarin learners’ superiority in production. L2 Mandarin learning experience tuned this cue weighting; thus, they
had a better balance of pitch height and contour. Chapter 8 will summarise the results found so far and answer the research questions raised in Chapter 4. It will also discuss the theoretical implications.
Chapter 8: Discussion and Conclusion

This thesis has examined the perception (categorisation: Chapter 5; discrimination: Chapter 6) and production (Chapter 7) of Cantonese tones by L1 Mandarin speakers, and by English speakers with and without tone learning experience. This final chapter reviews the main findings from the previous chapters and discusses the results in relation to the L1/L2 influence on perception and production, the correlation between them, as well as extensions to the theoretical frameworks.

8.1 Summary

After an introduction to the study in Chapter 1, followed by an overview of tone and intonation, as well as the prosodic systems of Cantonese, Mandarin and English in Chapter 2, Chapter 3 separately reviewed the relevant literature on perception and production by speakers of tone and non-tone languages. Chapter 4 introduced two of the most influential theoretical frameworks, with suggestions for extension, then outlined the experimental program and the research questions. Chapter 5 reported the results of speakers categorising Cantonese tones onto the Mandarin tone system (Mandarin speakers), the English intonation system, or both (Mandarin learners). The six Cantonese tones were identified as either categorised or uncategorised according to the assimilation patterns by different speaker groups. The tone pairs formed by any two tones were further tagged as SC, CG and TC if both were categorised tones; or UC-same set, UC-partial overlap, UC-no overlap if one tone was categorised and the other uncategorised; or UU-same set, UU-partial overlap, UU-no overlap if both were uncategorised. This further categorisation was based on the assimilation patterns that determined whether the two tones were assimilated into the same category and also the goodness ratings given to the target
category. For Mandarin speakers, five of the six Cantonese tones were categorised; the exception was Cantonese T5 (T23), which had the competing choices of Mandarin T2 (T35) and T3 (T214). For English speakers, four tones were categorised and two tones were un categorised (T2 [T25] and T6 [T22]). For Mandarin learners, the results were unified when categorising the Cantonese tones into English intonations and Mandarin tones: only T2 (T25) was un categorised. However, the detailed assimilations were quite distinct. English speakers with and without L2 tone experience categorised Cantonese tones into the same intonation categories, with the exception of the two rising tones. In contrast, the Mandarin speakers and Mandarin learners (with L1 English) only shared a single modal category (see Section 5.4 for a detailed description of modal categories) for CT3 (T33), where both groups chose MT1 (T5).

Chapter 6 discussed the discrimination results of the Cantonese tones and compared the percentage of correctly discriminated tones according to PAM-S, based on the categorisation patterns from Chapter 5. Overall, Cantonese speakers discriminated tones with the highest degree of accuracy (91.9%), followed by Mandarin learners (79.0%), Mandarin speakers (77.8%) and English speakers (71.9%). The results from Mandarin speakers confirmed the predictions from PAM-S/PAM-L2: TC > CG, UC-no overlap > UC-overlap > UC-same set. For English speakers, TC > CG, UC-no overlap > UC-overlap, and UU-overlap were the most easily discriminated pairs. Curiously, and despite the fact that the mean accuracy of TC pairs was higher than CG with English speakers, a few TC pairs showed lower accuracy than CG ones. For English-speaking Mandarin learners categorising into English, the accuracy ranking of the tone contrasts was TC ≥ CG > SC, UC-no overlap > UC-same set; and for English-speaking Mandarin learners categorising
into Mandarin, TC ≥ CG, UC-no overlap > UC-overlap. Not all TC pairs were better discriminated than the CG pairs. Additionally, for all speaker groups, UC did not always have moderate to excellent discrimination, contradicting what is proposed in PAM-S/PAM-L2.

In Chapter 7, the tone productions by four speaker groups were acoustically analysed in three dimensions: tone differentiation (Section 7.2.1), tone movement (Section 7.2.2) and tone duration (Section 7.2.3). In addition, L1 judges provided auditory assessment (Section 7.2.4). Tone differentiation analyses suggested that Cantonese speakers had the most differentiated tone productions, followed by Mandarin learners, Mandarin speakers and English speakers. With respect to tone identity, all speaker groups found level tones easier to produce than contour tones. Mandarin speakers tended to exaggerate tone movements on contour tones, while English speakers produced contour tones in a flattened way. Mandarin learners had the most L1-like contour productions. In relation to duration, Mandarin and English speakers produced tones that were significantly longer than those of the Cantonese speakers and Mandarin learners. These performance ranks were consistent with the auditory analysis—English Mandarin learners had the best productions (67.5%), followed by MS (61.8%) and ES (55.8%).

In the remainder of this chapter, key findings will be discussed in relation to the research questions raised in Chapter 4:

**RQ 1.** How are tones from a large tone inventory mapped to tones in a small inventory?

**RQ 2.** How do non-tone language speakers assimilate tones to their L1 prosodic system?

**RQ 3.** Does L1 and L2 tonal experience help in perceiving and producing
another tonal language?

**RQ 4.** What is the relationship between tone perception and production?

### 8.2 How Tone and Non-tone Speakers Assimilate Cantonese Tones

As reviewed in Chapter 2, together with the four types of word prosody—stress, tone/lexical pitch accent, both of these, and none of these—languages can be re-grouped into 15 different types. The current study involves English, Mandarin and Cantonese, which all belong to head-prominent languages, according to Jun (2014). Both Cantonese and Mandarin are lexical tone languages while English has intonation as its prosodic system, which also involves the use of different F0 patterns (however, they function post-lexically). English has medium macro-rhythm and stress, while Mandarin and Cantonese share a similarly weak macro-rhythm. However, Mandarin has tone and stress at the same time, but Cantonese has only tone (Jun, 2014). Given the different prosodic systems in these three languages, the ways in which speakers assimilate complex Cantonese tones to their L1 system is of great interest.

The categorisation results from Chapter 5 indicate that, in most cases, L2 tones are categorised as their most acoustically similar L1 counterparts, regardless of whether it is a lexical tone or intonation pattern. Mandarin speakers have a smaller tone system with a distinction based primarily on pitch contour. That all three Cantonese level tones are categorised as the only level tone in Mandarin demonstrates that even partial similarity can stimulate phonetic assimilation. Mandarin speakers’ L1 transfer negatively influences their perception, as they are confused by tones with the same contour but different pitch height. However, differences in the goodness ratings suggest that Mandarin speakers are indeed able to
differentiate F0 height: Mandarin listeners found CT1 (T55) to be the best fit for MT1 (T55), even though they also chose MT1 (T55) for CT3 (T33) and CT6 (T22).

In the case of the two rising tones, CT2 (T25) and CT5 (T23), listeners chose MT2, which is a rising tone (T35), but also the falling-rising MT3 (T214), which has an allophonic rising form with the tone pattern (T35). When the Cantonese rising tone is categorised as the rising tone in Mandarin, this again suggests that assimilation occurs at the phonetic level. However, when the rising tone is assimilated to the Mandarin falling-rising (T214) tone, this indicates phonological assimilation. This is because Mandarin listeners apparently apply their L1 phonological knowledge—that the falling tone and rising tone are allophonic variants of the falling-rising tone—to categorise the Cantonese rising tone (T23) and the Cantonese falling tone (T21). These results align with those of So and Best (2010), but contradict the findings of one study where no phonological assimilation occurred with naïve listeners (Wu et al., 2014).

L1 English speakers compared the target Cantonese syllables carrying six Cantonese tones with five provided monosyllabic words with different English intonation patterns and one unknown category. They most often chose ‘More.’ (H*L-L%) or ‘More…’ (H* H-L%). Participants showed the most agreement on the low-falling tone (T21). Two of the three level tones (T55 and T33) were mainly categorised into ‘More…’, which has a level contour. For the low-rising tone, the category chosen the most was still ‘More.’, which has an opposing pitch contour (falling vs. rising). Both the low-level and the high-rising tones were uncategorised for English monolinguals. For the low-level tone, these participants found two matching intonations: ‘More…’ and ‘More.’ for this target level tone, with ‘More.’ being category chosen most. Interestingly, the ‘More.’ pattern has a falling pitch
contour while ‘More…’ is more of a level pitch, yet English speakers chose the falling pitch contour to align with the low-level Cantonese tone. For the high-rising tone, where an even number of choices were made for ‘More…’ and ‘More?!’, with the level pitch contour chosen as the counterpart of a rising tone.

English speakers sometimes chose an unmatched contour for the intended tone type, supporting the notion that English speakers are more attentive to pitch height than pitch contour (Gandour, 1983, 1984). Even though it was not indicated by the main choice for the two Cantonese rising tones, a greater number of participants chose ‘More?!’ (H* H-H%) for the high-rising tone (T25). ‘More?’ (L* H-H%) was the secondary choice (aside from the statement intonation) for the low-rising tone (T23). This indicates that English monolinguals can distinguish the different rising ranges between the low- and high-rising tones. English monolinguals’ strong preference for ‘More.’ or ‘More…’ indicates that they favour intonation patterns with less pitch movement. This is in line with previous studies showing that the Mandarin falling tone is the most ‘normal’ tone to English ears (Broselow, Hurtig & Ringen, 1993; Chiang, 1979). This also supports the notion that English intonation has level pitches as underlying; intonation contours are the combinations of high- and low-level pitches (Liberman, 1978; Pierrehumber 1980; Pike, 1945).

8.3 The Influence from Native as well as Non-native Experiences

That linguistic background is a determining factor in participant performance is not in question: all speaker groups performed differently from one another across the tasks. The effect of L1 language backgrounds can be clearly seen in the different performances between Mandarin and English speakers. English speakers, coming from a non-tone language background, are less familiar with pitch information that has lexical meaning (Lee et al., 1996; Wayland & Guion, 2004; Wayland & Li,
This may be the cause of their less competent performance in the tone discrimination task.

The finding that Mandarin speakers outperform English speakers in AXB discrimination tasks also indicates that coming from a tone language background still exerts certain advantages when discriminating L2 tones. Their familiarity with lexical tone as a perceptual cue may be a contributing factor in their better performance. This supports previous findings from Wayland and Guion (2004) and Qin and Mok (2011). However, it differs from results reported by Hao (2011), who observed that English speakers outperformed Cantonese speakers on both Mandarin tone identification and reading tasks. That English speakers are better than Mandarin speakers at discriminating level tones can be seen as a negative influence affecting Mandarin speakers, as they have only one level tone in their L1 tone inventory. This supports the findings of Chiao et al. (2011), who concluded that listeners from non-tone background were better at perceiving level tones than speakers with only one level tone in their L1 system. In the case of the English monolingual participants, their categorisation results can be used to predict their discrimination performance, indicating that these non-tone speakers are still using their L1 intonation system to perceive these seemingly unfamiliar tones, which is supportive of proposals put forward by So and Best (2011).

The production results presented here show that Mandarin speakers have a bigger tonal space and better tone differentiation than do English speakers. For tone movements, it is difficult to define which speaker group performed ‘better’. English and Mandarin speakers show quite different patterns in production: Mandarin speakers exaggerate pitch movement on contour tones and have a shrunken space for level tones, while English speakers tend to pronounce contour tones with a more
level shape. In their L1 languages, speakers weight perceptual cues differently—
Mandarin speakers pay more attention to pitch direction, as they are used to
differentiating their L1 tones by pitch direction alone. The English speakers are
instead more sensitive to pitch height (Wang et al., 2003). This indicates that their L2
productions are highly moulded by L1 production patterns. English speakers’
particular difficulty can be explained by their lack of familiarity with tones, influence
from intonation and smaller pitch range (White, 1981). Tone language speakers’
advantage ensures that their tone productions are robust, a notion supported by

The combined findings from the perception and production tasks provide
detailed evidence for the influence from L1 prosodic systems. As suggested by both
PAM and SLM (the two major speech theories reviewed in Chapter 4), the
perception and production of a new language is dependent on the discrepancies
between native tonal/intonational system and the L2 tone system. The difference in
performance by English and Mandarin speakers can be explained from a
neuroimaging perspective: that the brainstems of speakers from a tone language
background have a more accurate pitch-tracking ability when processing lexical
tones than those from non-tone language backgrounds (Krishnan, Gandour &
Bidelman, 2010).

The results of the current study also point towards different cue weighting by
speakers with different language backgrounds during L2 tone perception. Both
discrimination and production studies show that Mandarin speakers are more
sensitive to pitch contour and have more problems with tones that share the same
contour but have different pitch heights. English speakers, on the contrary, pay more
attention to pitch height. When producing contour tones, they tend to flatten the
shape. This distinction confirms previous observations from Gandour (1983) regarding the perception by tone and non-tone language speakers: the two groups differed in the extent to which they were attentive to F0 direction and height. Specifically, non-tone language speakers pay more attention to pitch height.

In addition to investigating how L1 languages influence L2 perception and production, this thesis offers an important and innovative extension to assessing the role of language experience by examining how L2 experience influences L3 perception and production. The discussion of how linguistic experience influences L2 perception and production usually focuses on experience with one language, the first language. However, as foreign language learning becomes an increasingly important part of education, many people also learn a third or fourth language. It is essential to understand how L2 learning experience interacts with L1 experience when a speaker is learning another new language—will it be disruptive? The current results clearly demonstrate that L2 learning experience shapes L3 perception and production as well. Even when coming from the same L1 language backgrounds, English speakers with Mandarin learning experience categorise Cantonese tones differently onto English intonation systems. The low-falling tone is mainly categorised into ‘More.’ by English monolinguals; however, quite a few Mandarin learners found it more like ‘More!’ English monolinguals mostly categorised the low-rising tone into ‘More.’, an intonation with falling shape, while Mandarin learners debated between ‘More?’ and ‘More?!’, both of which carry rising pitches. This suggests that their attention to pitch contour has been tuned by their Mandarin learning experience. In addition, Mandarin learning experience offers a significant benefit to English speakers in terms of their discrimination and production accuracy of Cantonese tones. In discrimination studies, English-speaking Mandarin learners
outperform Mandarin speakers by 1.2% and English speakers by 7.1%. When comparing the discrimination results between Mandarin speakers and Mandarin learners, English monolinguals and Mandarin learners, the experience strengths are striking. Mandarin learners outperform Mandarin speakers on a few tones: T3-T4, T4-T5, T1-T3, T3-T6, T1-T6, T2-T4 and T2-T5, which are either tones with the same contour different height, or pairs involving T4. By contrast, Mandarin learners outperform English monolinguals on most tones. The pairs with the biggest differences are T3-T4, T2-T3, T4-T6 and T5-T6, which all result from confusion between a level and a contour tone. Thus, Mandarin learners have better judgement about pitch height than do Mandarin speakers and they pay much more attention to pitch contour than do English speakers. The production findings suggest that Mandarin learners have a slightly larger tonal space than do Mandarin speakers, which in turn is twice as large as the tonal space of English speakers. These learners’ tone differentiations within their tonal space are better than those of Mandarin speakers on almost every tone type, except for T21. This observation accords with the findings for toneme differentiation (against other tones), that Mandarin learners are slightly better than Mandarin speakers. Each of these groups is also much better than English speakers. In previous research, the influence of L2 experience on L3 tone perception and production has not been reported extensively. A perception study examining the same population found a similar positive influence from L2 Mandarin experience (Qin & Jongman, 2015). Findings from the current thesis have confirmed this observation by examining the categorisation and discrimination of all six Cantonese tones; Qin and Jongman’s (2015) study limited the stimuli to only three of the tones. Further, this study has extended this positive influence to L3 tone production, having shown that tone language experience greatly improves non-tone
language speakers’ ability to produce a new tone language. The current findings also support those of Burnham et al. (2014), who determined that universal and language-specific factors work together during the L2 perception process and provide strong evidence that this view can be extended to production.

8.4 Correlation between Perception and Production

The complexity of the link between perception and production has never been in dispute. However, controversy regarding the nature of the relationship between speech perception and production has long existed, and has led to numerous investigations on multiple populations with perceptual training (Bradlow et al., 1997; Huensch & Tremblay, 2015), or without perceptual training (Flege, MacKay & Meador, 1999; Kosky & Boothroyd, 2003; Sheldon & Strange, 1982; Wode, 1996).

This section discusses the relationship between participants’ L2 (for Mandarin and English speakers) or L3 (for Mandarin learners) perception and production. The current study design enables the comparison of different groups’ performance in perception and production tasks. The discrimination results presented in Chapter 6 can be regarded as perception performance, and the auditory judgement of non-native production presented in Chapter 7 correspond to production performance. The results show that the perception and production abilities of both Mandarin speakers and English speakers with Mandarin learning experience are highly positively correlated. By contrast, English monolinguals show no correlation between their perception and production of Cantonese tones.

The positive perception-production link for L2 speakers has been reported mostly for L2 learners (Ding et al., 2011; Flege, et al., 1999; Kosky & Boothroyd, 2003; Sheldon & Strange, 1982; Smith, 2001). In the current case, neither the Mandarin speakers nor the Mandarin learners were L2 learners of Cantonese,
indicating that the roots of the link do not necessarily originate in learning experience. Wang et al. (2003) have previously indicated that the correlation was present after only a short training period.

The no-correlation case of English speakers is not unique: several other studies have found similar results (de Jong et al., 2009; DeKeyser & Sokalski, 1996) or merely a partial correlation (Hattori & Iverson, 2010). The result does align with findings reported by Bent (2005), but contradicts those of Leung (2007), Hao (2011) and Yang (2014), who found English speakers’ tone production ability was always limited by their perception.

Interestingly, the results of the present study suggest the possibility that the correlations found with L2 learners do not stem from learning a second language, but in a more general way relate to tone experience. As neither of the two groups showing correlations in this study had learned Cantonese, it could be inferred that as long as speakers have tone experience (no matter whether in an L1 or L2), their perception and production are positively correlated. As highlighted in tone production studies indicating that familiarity with tone has a great influence on non-native tone production (Leung, 2008; White, 1981), the correlation’s existence could be more dependent on production performance as practice is required for production.

Further, even though the best individual perceivers and producers are not the same under some circumstances, there is still an apparent trend for people who perceive well to also have better production. This supports a previous vowel discrimination and production study (Bent, 2005) in which speakers with higher auditory acuity are held to produce a more precise representation. Here (Bent, 2005), it is claimed that the more precise representations can be seen as smaller target areas in acoustic space. Thus, there will be less variation in production, as speakers with
better perception ability will notice the outlying produced tokens and self-repair the productions. The current findings that English speakers have the most variation in production support this, as they perceive Cantonese tones poorly.

In sum, this thesis indicates great individual differences in both perception and production data. More variance is found with Mandarin speakers and Mandarin learners in relation to perception than to production, while the opposite is true for English monolinguals. Apart from language backgrounds, several factors were controlled when recruiting participants: age, gender, education and musical background. Every participant passed a pure tone-screening test at the beginning of the study as well. After second review of the information collected through the language background questionnaires, still nothing could be linked to the better and poorer performances of the participants directly. Previous studies have indicated a wide range of possible explanations for individual differences, with linguistic aptitude a popular suggestion, and this ‘aptitude’ supposedly gives a person enhanced ‘phonetic coding ability’, which in turn helps L2 learning (Carroll, 1981; Sparks et al., 1997). This ability can either be innate, ‘a residue of L1 aptitude’ or dependent on experience with other languages (McLaughlin, 1990). Besides aptitude, L1 skills and general cognitive abilities also contribute to the variance of performance (Darcy, Park & Yang, 2011; Sparks & Ganschow, 1993; Sparks, Ganschow & Patton, 1995; Sparks et al., 1997). A few studies have also found a relationship between L1 phonological ability and L2 learning success (Díaz et al. 2008; Sparks et al., 1997). This indicates that linguistic pitch ability and musicality are better predictors of tone learning success than general cognitive ability and L2 aptitude (Bowles, Chang & Karuzis, 2015; Cooper & Wang, 2012; Gottfried, 2007; Sleve & Miyake, 2006). Our results support these results: that a domain-related
ability—here pitch ability—works as a better predictor of tone learning success than general L2 aptitude.

8.5 Implications for Current Frameworks

This section will first review the proposed extension of current frameworks (PAM and SLM) outlined in Chapter 4, followed by a discussion of the insights gained from the current study.

This thesis has tested PAM-S in the domain of tone perception and provide further extension of PAM into the domain of tone production. The implications are noted in the following paragraphs. For non-tone language speakers, most tones are likely to be perceived as speech, although not categorisable according to an L1 phonological entity (e.g., an intonation system). Thus, the L2 tones will be either uncategorisable or categorisable, with contrasts formulated as UC and UU. For a UU contrast, the L1 system should exert little influence on discrimination and the goodness should be fair to good, depending on the distance between the L2 phonemes and the closest L1 ones. A UC contrast, however, should have excellent discrimination results, as the two tones differ significantly from each other.

For English speakers, four tones are categorisable and the other two are uncategorisable (CT2 and CT6). However, not all UC contrasts are equally easy to discriminate; for example, UC-no overlap and UC-partial overlap are the most poorly discriminated pairs. Categorised pairs follow the PAM predictions well—that TC>CG>SC—although not all TC pairs have higher discrimination accuracy than the CG ones. For English speakers with Mandarin experience, only one tone is uncategorised, with tones mapped onto either Mandarin or English prosodic systems. Generally UC pairs are well discriminated, apart from one tone pair that is categorised as either UC-same set (English) or UC-partial overlap (Mandarin), which
is poorly discriminated.

Thus, the present thesis suggests that PAM-based UC predictions still need some refinement: even with the distinctions of categorised and uncategorised, two tones with overlap are more difficult to discriminate than those that have no overlap. A reason for this overlap between categorised and uncategorised is that the definition of uncategorised does not limit itself to a no-matching category, but does not match one specific category. In addition, as for both groups TC pairs are not always better discriminated than CG pairs (even though the mean accuracy of TC is higher than that of CG), further investigation could help in terms of this extension.

For tone language speakers, L2 tones will most likely be perceived as categorisable with respect to speakers’ L1 tone inventory, and it is likely that some tone pairs will constitute TC pairs, while others will be CG, and in rare cases perhaps even SC. If the tone contrast is categorised as TC, the two tones should be quite different in both L1 and L2. Hence, this contrast will be easy to discriminate. If the two tones fall into the CG pair, the level of discrimination difficulty is predicted by the articulatory, acoustic and perceptual distance between the two members from the L1 category. If these two tones differ greatly from each other as well as the L1 tone they will still be easy to discriminate. However, if they are both close to the L1 category, discrimination will be more difficult. When two tones form a SC pair, it will be extremely difficult to discriminate them as they are assimilated to the same L1 category with the same distance to the L1 tone.

The results presented in this thesis also show that not all tones are categorisable to Mandarin speakers: indeed, we saw in Chapter 5 that CT5 (T23) is uncategorised. Apart from that, the extension is well supported by the current findings, as most tones are TC or CG and no tone pair here is SC for Mandarin
speakers. Discrimination results are as predicted: TC are mostly very easy to
discriminate, while CG are moderate. A UC pair formed with CT5 does not always
have excellent discrimination, but it follows the general rule that when the two pairs
have more similarity with each other, discrimination is more difficult.

An extension of PAM into tone production is much needed. Indeed, as PAM
predicts that perception leads production, and they are intimately connected as they
rely on the same perceptual system, listeners’ perception and production must be
closely linked—if a learner perceives L2 tones well, he or she should also be able to
produce them reasonably accurately. Moreover, the errors one makes in production
should be directly relatable to perception errors. For example, if a learner
misperceives a particular tone, he or she should also have problems when producing
it.

Importantly, however, the findings reported in this thesis do not support a
direct connection of the sort proposed under a strict PAM framework; not all tone
difficulty in production has a perceptual basis and not all poorly perceived tones are
produced poorly. For speakers with no tone language background, the two modalities
are not even correlated. There are individual speakers who perceive and produce
equally well, but as a group, their performances are not linked. Thus, the direct link
between perception and production proposed by PAM is not supported by this study.
SLM, on the other hand, explicitly proposes the relationship between perception and
production, but it does not have a detailed explanation for L2 perception and
production performances.

In Chapter 4.2.2, I proposed an extension for tone language speakers’
perception: L2 speakers will map L2 tones to the L1 categories, according to a
similarity effect, just as with vowels and consonants. An L2 tone from a completely
different category than the L1 one might be easier to perceive and produce than one perceived as being in the same category. The categorisation results reported in Chapter 5 support the suggested extension that Mandarin speakers map Cantonese tones onto their L1 tone systems. If two tones are mapped onto different L1 categories, they are easier to discriminate. However, if the uncategorised tone (T23) is seen as different from the L1 category, the discrimination difficulty is also determined by how similar or different it is with the counterpart in a pair, ranging from poor to excellent. The production accuracy for T23 is relatively low, according to the auditory judgement. Thus, the distance from the L1 category does not guarantee success in perception and production.

The present thesis also puts forth an extension of SLM to account for non-tone language speakers: it is likely that they will make use of their L1 prosodic patterns to perceive L2 tones. Tones similar to existing prosodic patterns might be more difficult to perceive and produce for such learners, while tones with no overlap might be easier. Findings from the current thesis indicate that all English speakers successfully mapped Cantonese tones onto their intonation systems. English speakers with and without Mandarin learning experience have slightly different categorisation patterns. For English monolinguals, T25 and T22 are both uncategorised, while for Mandarin learners, only T25 is uncategorised. Similarly with Mandarin speakers, if the two tones are mapped onto different L1 systems with no overlap, the discrimination is easier. The production accuracy fluctuates: for English speakers, their T25 production is the second worst but T22 is moderate. For Mandarin learners, T25 has moderate production. The production difficulty might be more related to the acoustical difficulty of the tones (T23 is the most difficult tone to accurately produce for all speaker groups) and the cues that listeners are familiar with in their L1
prosodic systems (English speakers are more sensitive to pitch height while Mandarin speakers pay more attention to contour).

Taken together, these results show that L2 tone perception and production are not directly linked, and some representations are different. Indeed, the results may be taken to suggest that tone as perception is rooted in psychoacoustics, while production is founded on articulatory elements. Further, the results suggest that perceptual learning precedes production learning and a problematic perception will lead to imperfect production, but importantly this does not mean that all production errors have a perceptual basis.

The link between perception and production observed in the current study is more consistent with SLM’s indirect relationship, as some poorly produced tones are well perceived. The correlation between the two modalities is sometimes not present when investigating individual speakers’ performances. General better discrimination in perception could support SLM’s assumption that perception precedes production.

The extensions of PAM and SLM into tone perception and production can be summarised as follows:

1. speakers from either tone or non-tone language backgrounds will recruit their L1 prosodic system to perceive and produce L2 tones
2. the specific predictions for categorisation and discrimination from PAM are well supported by the results, except for the UC pair
3. production is less predictive in both theories
4. perception and production are correlated when speakers have some experience with tone, but the link is indirect, and the SLM makes better sense here.
However, neither of the frameworks provides an explanation for L3 perception and production; thus, a model combining L1 and L2 experiences is vital. According to the assimilation fit index (see Section 5.4, Tables 5.15 and 5.17), the modal answers are mostly the same for English speakers with and without Mandarin experience, except for the two rising tones. This indicates that L2 learning experience tunes English speakers mostly on rising tones. The discrimination results show that they have a better ability to distinguish contour tones from level tones. The production findings indicate that Mandarin learners have a larger tonal space and bigger pitch movement on contour tones. Therefore, a model incorporating L3 tone perception and production is proposed based on the current findings: when L2 falls into the same prosodic typology as L3, L1 and L2 will both be drawn on to assist perception and production. The cues which speakers are less accustomed to in their L1 perception will be under-practised during L2 training. This cue attention will change speakers’ assimilation of the L2 tones into their own prosodic system, and thus enhance their performance in discriminating and producing L2 tones.

8.6 Strengths and Limitations

The current study explores the influence of L1 and L2 linguistic experiences on the perception and production of Cantonese tones comprehensively. The study’s strengths lie in the chosen populations, and the carefully-considered experiments and analyses. The recruitment of intermediate Mandarin learners from English-speaking backgrounds is innovative—this is the first study testing the production of tones by speakers from non-tone backgrounds who have received intermediate tone training in their L2 studies. This is particularly important in the Australian context, a multicultural society in which people have a range of first and additional language experiences. The influence from linguistic background is no longer limited to an L1;
instead, it encompasses cumulative linguistic experiences, regardless of whether this is L1 or L2 experience. As such, conducting this study with this specific group has great practical importance: learning a second language will change the way in which someone perceives and produces a new language, in the same way as one’s L1 influences perception and production. The study design enables a comprehensive investigation of every step in the early stages of exposure to L2 tones. The categorisation develops a picture of how speakers from different language backgrounds categorise Cantonese tones onto their L1/L2 prosodic systems and provides detailed predictions for the discrimination task. The fit index and mapping diversity analyses summarise how distinct or similar the L2 and the L1 systems are.

The stimuli applied for the categorisation task are fine-tuned: the similar syllable in three languages makes the task more comparable. When asking English speakers to categorise the use of the five English intonations (‘More.’, ‘More?’, ‘More!’, ‘More…’ and ‘More?!’), participants were able to listen to these choices linked with pre-recorded sounds produced by L1 Australian speakers. This is an innovative method, improving the previous categorisation into simple language descriptions of intonation (e.g., statement or question). Mandarin learners were asked to categorise Cantonese tones onto their L1 English intonation systems and Mandarin tone systems separately, to ensure that the influence from both systems could be traced. A comparison with English and Mandarin speakers’ categorising patterns led to the observation that having learned Mandarin as an L2, their mapping onto the L1 English system has also been changed. Their categorisation has more similarities with L1 English speakers than with Mandarin speakers.

Further, the production analyses are comprehensive: they include total tonal space, tone differentiation against other tone type and the total tonal space, tonal
contour across time, duration and native auditory judgement. These analytical methods cover almost all the important cues in tone, and lead to a robust conclusion that Mandarin learners outperform both Mandarin speakers and English speakers.

As with all research, this study has its limitations. Firstly, though Mandarin speakers typically claim that they cannot understand Cantonese, it is very common for them to have experienced Cantonese culture while growing up. For example, they would most likely have had exposure to Cantonese music and television drama. Thus, the credit for Mandarin speakers’ better performance may be partially due to some level of familiarity with Cantonese.

Secondly, the language background questionnaire did not include a test of general intelligence, which is an important variable in explaining language-learning aptitude. No specific reasons were established for the individual differences observed in the present study, and though the participants in this study were all university students and therefore had much in common in terms of educational background, it is possible that intelligence may have offered some insights into these differences. Further, the Mandarin learners’ performance in learning Mandarin tones was not recorded. A skilled learner of Cantonese can be excellent at learning Mandarin as well. With this information, it might have been possible to ‘connect the dots’ regarding some speakers’ higher learning ability being due to universal tone ability.

Thirdly, this study attempted to extend the current theoretical frameworks to L2 perception and production. With PAM, the definition of categorised/uncategorised is that the chosen category must have more choices than chance level and significantly more choices than other competing categories. Thus, it is quite possible for an uncategorised tone to have two competing categories, which could always have some overlap with the other tone in the pair. Therefore, it is less
realistic for a UC pair to have a generally better discrimination; the accuracy will depend on whether the two tones in a pair are categorised onto L1 categories with or without overlap.

8.7 Future Directions

This study has offered a comprehensive set of results, following a carefully-considered methodological approach, which together provide a strong foundation for future work on ways in which the perception-production link is mediated by language experiences. The influence of L1 and particularly L2 linguistic experiences is more frequently investigated at the segmental level, and while this study’s focus on tonal patterns makes a valuable contribution to the literature in this area, it is clear that more investigations of suprasegmental phenomena are necessary in order to arrive at a more comprehensive understanding of the perception-production link. The question of whether L2 experience can mould the perception and production of a new system in the same way as L1 experience is also under-explored; the findings of the present study provide compelling evidence that it can, and suggest a need for linguistic experience to be considered in a range of different ways.

As has been demonstrated, there is scope for existing models to continue to be improved and extended based on experimental data. Different instrumental approaches may further develop understandings; given that neuroimaging studies provide effective explanations for the cue weighting by tone and non-tone language speakers (Wang et al., 2003), it would be helpful to examine the advantage of Mandarin learners via neurological methods like PET (Positron Emission Tomography) or fMRI (Functional magnetic resonance imaging) to see whether having learned a tone language changes the way the brain functions when processing lexical tone information.
The lack of a correlation between tone perception and production for English speakers is posited as being linked with tone experience in general, as well as a lack of tone experience influencing tone production. Further studies can be conducted to test this hypothesis and determine whether it is correct that tone familiarity is the key component, influencing tone production more than perception.
References


Appendices

Appendix A: Language Background Questionnaires

The appendix includes language background questionnaires for the participant recruitment in the current study. It includes the English version for native English speakers and the Chinese version for native Mandarin and Cantonese speakers.

---

Speaker Language Background Questionnaire

1. Your name: _____________________________

2. Gender: □ Male □ Female

3. Your age in years: _____________________________

4. Country of birth: _____________________________

5. Where did your father come from? _____________________________

6. Where did your mother come from? _____________________________

7. Please indicate below which countries and towns you have lived in from birth till now:

<table>
<thead>
<tr>
<th>Age Range (e.g. birth - 5yrs)</th>
<th>Country</th>
<th>Region &amp; Town(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Please tell us all the languages and/or dialects you know, at what age you began learning each one, and how well you them (circle the corresponding number in the column).

<table>
<thead>
<tr>
<th>Language/Dialect</th>
<th>Age when you began learning it</th>
<th>From whom or where</th>
<th>Indicate your level of ability by circling number 1=very little, 5=very well</th>
<th>To whom do you speak it</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>
9. Have you received any professional musical training (musical instruments, or singing) before?
   □ Yes  □ No
   If yes, how long have you learned it? ______
   Are you still playing/singing it? ______

10. Did/do you, your parents, or any of your siblings have any special problems with language development (e.g. delayed language onset, serious difficulties in learning new words or remembering the names of objects)? Yes
    No  □  □
    If YES, who has/had the problem and what is/was the nature of the problem? At what age did it occur? Did it/does it require a language therapist?
    __________________________________________
    __________________________________________

11. Did/do you, your parents, or any of your siblings have any special problem with speaking (e.g. stuttering, lisping, etc.)? Yes  □ No  □
    If YES, who has/had the problem and what is/was the nature of the problem? At what age did it occur? Did it/does it require a language therapist?
    __________________________________________
    __________________________________________

Please indicate us an e-mail address where we can contact you

__________________________________________

Signature:_________________________________
语言背景调查

1. 你的姓名：__________________________

2. 性别： ☐ 男 ☐ 女

3. 年龄：__________________________

4. 出生地：__________________________

5. 你何时来的澳洲？__________________________

6. 你父亲来自哪？__________________________

7. 你母亲来自哪？__________________________

8. 请标明你从出生到现在居住过的国家以及城市

<table>
<thead>
<tr>
<th>年龄阶段（例如 出生至五岁）</th>
<th>国家</th>
<th>城市</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. 请告诉我们你会说的语言以及方言，你何时开始学，你说得怎么样。

<table>
<thead>
<tr>
<th>语言以及方言</th>
<th>你几岁时开始学</th>
<th>你从谁或者哪儿学的</th>
<th>请圈出你的水平 1=只会一点儿，5=非常好</th>
<th>你和谁说这个语言/方言</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
10. 你接受过连续三年以上的音乐训练吗？包括乐器以及唱歌。

□ 有  □ 无

如果有，你学了多久？________
你现在还会演奏/唱歌吗？________

11. 你或者你父母、你兄妹有过任何语言发展的问题吗？（例如很晚开始说话，没法学会新词之类）

□ 有  □ 无

如果有，谁有？是什么问题？什么时候开始出现的？需要语言治疗师介入吗？

__________________________________

__________________________________

12. 你或者你父母、兄妹有任何发音障碍吗？（例如口吃问题等）

□ 有  □ 无

如果有，谁有？是什么问题？什么时候开始出现的？需要语言治疗师介入吗？

__________________________________

__________________________________

请告诉我们你的电子邮箱地址

__________________________________

签名：__________________________________
Appendix B: Experiment screen for categorisation study in Chapter 5

This appendix includes the screenshots for the categorisation task: categorising Cantonese tones into Mandarin tones and English tunes.

Cantonese vs Mandarin

Which is the most similar?

1

How similar are they?
1 – 5 scale
1 least similar; 5 almost the same
1 2 3 4 5

Figure B.1. Cantonese into Mandarin tones

Cantonese vs English

1

More?  More!  More?!  Unknown

Figure B.2. Cantonese into English tunes
Appendix C: Illustration of English stimuli in Chapter 5

This appendix provides the Praat screenshots of the five English tunes.

Figure C.1. Illustration of English stimuli ‘More?’ (L* H-H%).

Figure C.2. Illustration of English stimuli ‘More!’ (L+H* L-L%).

Figure C.3. Illustration of English stimuli ‘More.’ (H* L-L%).
Figure C.4. Illustration of English stimuli ‘More…’ (H* H-L%).

Figure C.5. Illustration of English stimuli ‘More?!’ (H* H-H%).
Appendix D: Scatterplots for the F0 onsets and offsets results in Chapter 7

This appendix includes the detailed F0 onsets and offsets production results for each speaker group and each tone category.

Figure D.1. F0 onsets and offsets results for Tone 55
Figure D.2. F0 onsets and offsets results for Tone 25
Figure D.3. F0 onsets and offsets results for Tone 33
Figure D.4. F0 onsets and offsets results for Tone 21
Figure D.5. F0 onsets and offsets results for Tone 23
Figure D.6. F0 onsets and offsets results for Tone 22
**Appendix E:** T-values for tone movement results in Chapter 7

This appendix includes the table detailing the T-values at every quarter timepoints. It is a part of the production results presented in Chapter 7.

Table E.1. T-values at quarter timepoints for all speaker groups

<table>
<thead>
<tr>
<th>Tones</th>
<th>Speakers</th>
<th>T-value at 0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
<th>Max. F0</th>
<th>Min. F0</th>
<th>Avg. F0</th>
<th>Change Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HL55</td>
<td>C</td>
<td>4.31</td>
<td>4.34</td>
<td>4.30</td>
<td>4.36</td>
<td>4.33</td>
<td>4.36</td>
<td>4.30</td>
<td>4.33</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>4.03</td>
<td>4.02</td>
<td>4.06</td>
<td>4.02</td>
<td>4.01</td>
<td>4.06</td>
<td>3.99</td>
<td>4.03</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.96</td>
<td>3.98</td>
<td>4.06</td>
<td>3.99</td>
<td>4.02</td>
<td>4.06</td>
<td>3.96</td>
<td>4.01</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>4.22</td>
<td>4.23</td>
<td>4.19</td>
<td>4.16</td>
<td>4.18</td>
<td>4.23</td>
<td>4.16</td>
<td>4.20</td>
<td>0.07</td>
</tr>
<tr>
<td>ML33</td>
<td>C</td>
<td>2.87</td>
<td>2.87</td>
<td>2.83</td>
<td>2.84</td>
<td>2.87</td>
<td>2.82</td>
<td>2.85</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>3.18</td>
<td>3.21</td>
<td>3.23</td>
<td>3.19</td>
<td>3.22</td>
<td>3.23</td>
<td>3.18</td>
<td>3.21</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.82</td>
<td>2.88</td>
<td>2.93</td>
<td>2.96</td>
<td>2.91</td>
<td>2.97</td>
<td>2.82</td>
<td>2.91</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>2.88</td>
<td>2.89</td>
<td>2.91</td>
<td>2.93</td>
<td>2.93</td>
<td>2.94</td>
<td>2.87</td>
<td>2.91</td>
<td>0.07</td>
</tr>
<tr>
<td>LL22</td>
<td>C</td>
<td>1.97</td>
<td>1.96</td>
<td>1.97</td>
<td>1.98</td>
<td>2.00</td>
<td>2.00</td>
<td>1.96</td>
<td>1.98</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>2.83</td>
<td>2.84</td>
<td>2.81</td>
<td>2.77</td>
<td>2.79</td>
<td>2.85</td>
<td>2.77</td>
<td>2.81</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.16</td>
<td>2.23</td>
<td>2.15</td>
<td>2.18</td>
<td>2.20</td>
<td>2.23</td>
<td>2.15</td>
<td>2.18</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>1.80</td>
<td>1.82</td>
<td>1.85</td>
<td>1.87</td>
<td>1.87</td>
<td>1.89</td>
<td>1.80</td>
<td>1.84</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Rising</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR25</td>
<td>C</td>
<td>2.08</td>
<td>2.57</td>
<td>3.12</td>
<td>3.76</td>
<td>4.31</td>
<td>4.31</td>
<td>2.08</td>
<td>3.18</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>2.14</td>
<td>2.35</td>
<td>3.18</td>
<td>3.85</td>
<td>4.14</td>
<td>4.14</td>
<td>2.14</td>
<td>3.13</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.93</td>
<td>3.12</td>
<td>3.46</td>
<td>3.69</td>
<td>4.34</td>
<td>4.34</td>
<td>2.93</td>
<td>3.51</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>2.41</td>
<td>2.84</td>
<td>3.26</td>
<td>3.97</td>
<td>4.35</td>
<td>4.35</td>
<td>2.41</td>
<td>3.38</td>
<td>1.94</td>
</tr>
<tr>
<td>LR23</td>
<td>C</td>
<td>2.02</td>
<td>2.29</td>
<td>2.57</td>
<td>2.83</td>
<td>3.02</td>
<td>3.02</td>
<td>2.02</td>
<td>2.56</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>1.64</td>
<td>1.98</td>
<td>2.38</td>
<td>2.78</td>
<td>3.15</td>
<td>3.15</td>
<td>1.64</td>
<td>2.38</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.97</td>
<td>2.35</td>
<td>2.42</td>
<td>2.51</td>
<td>2.80</td>
<td>2.80</td>
<td>1.97</td>
<td>2.42</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>1.68</td>
<td>1.96</td>
<td>2.37</td>
<td>2.66</td>
<td>2.91</td>
<td>2.91</td>
<td>1.68</td>
<td>2.32</td>
<td>1.23</td>
</tr>
<tr>
<td><strong>Falling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF21</td>
<td>C</td>
<td>2.08</td>
<td>1.89</td>
<td>1.47</td>
<td>1.21</td>
<td>0.99</td>
<td>0.99</td>
<td>2.08</td>
<td>0.99</td>
<td>1.53</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>2.65</td>
<td>2.31</td>
<td>2.12</td>
<td>1.55</td>
<td>0.80</td>
<td>0.80</td>
<td>2.65</td>
<td>0.80</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.30</td>
<td>2.16</td>
<td>1.78</td>
<td>1.81</td>
<td>1.92</td>
<td>2.30</td>
<td>1.62</td>
<td>1.97</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>2.28</td>
<td>1.98</td>
<td>1.64</td>
<td>1.33</td>
<td>1.03</td>
<td>1.03</td>
<td>2.28</td>
<td>1.03</td>
<td>1.65</td>
</tr>
</tbody>
</table>

C = Cantonese M = Mandarin E = English EM = English speakers with Mandarin experience
Appendix F: Boxplots for duration results in Chapter 7

The appendix presents the duration boxplots for all four speaker groups in six Cantonese tones and three vowels.

Figure F.1. Boxplots of duration—production by Cantonese speakers.
Figure F.2. Boxplots of duration—production by Mandarin speakers.
Figure F.3. Boxplots of duration—production by English speakers
Figure F.4. Boxplots of duration—production by Mandarin leaners