EFFECTS OF L1 PROSODIC BACKGROUND AND AV TRAINING ON LEARNING MANDARIN TONES BY SPEAKERS OF CANTONESE, JAPANESE, AND ENGLISH

by

Connie Kwok Lai So

B. Sc., University of Victoria, 1996
M.A., Simon Fraser University, 2000

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the Department of Linguistics

© Connie Kwok Lai So 2006

SIMON FRASER UNIVERSITY
Summer 2006

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.
Name: Connie Kwok Lai So
Degree: Doctor of Philosophy
Title of Dissertation: Effects of L1 prosodic background and AV training on learning Mandarin tones by speakers of Cantonese, Japanese, and English

Examining Committee:
Chair: Dr. Nancy Hedberg
Associate Professor of Linguistics

Dr. Zita McRobbie
Senior Supervisor
Associate Professor of Linguistics

Dr. Murray Munro
Supervisor
Professor of Linguistics

Dr. Trude Heift
Supervisor
Associate Professor of Linguistics

Dr. Alexei Kochetov
Internal/External Examiner
Assistant Professor of Linguistics

Dr. Catherine Best
External Examiner
Professor, MARCS Auditory Laboratories
University of Western Sydney

Date Approved: ________________________________
DECLARATION OF
PARTIAL COPYRIGHT LICENCE

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection, and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada

Summer 2006
STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, BC, Canada
The present study aims to address the theoretical and methodological issues of tonal acquisition in a second language (L2). The effects of native (L1) prosodic background and audiovisual (AV) training on the perception of the four Mandarin tones by groups of naïve listeners were examined.

The experiment employed a pretest-posttest paradigm. Three listener groups of 10 participants each were recruited: Hong Kong Cantonese (a tone language), Japanese (a pitch-accent language), and English (a stress-accent language). They were randomly assigned to receive one of two training approaches: Simple or AV Feedback. With the Simple Feedback training approach the responses were evaluated as being correct or incorrect. The AV Feedback consisted of sound files, animated pitch graphs, and a brief message that, in addition to indicating whether the response was correct or not, directed listeners’ attention to the crucial perceptual cues of tones. Following training, a posttest and three generalization tests were administered at two different times. Percent correct scores, perceptual sensitivities to each tone (A-prime), and tonal confusions were analyzed.

The results indicated substantial differences in the participants’ perception of Mandarin tones after training. With respect to L1 prosodic background, it was found that listeners’ L1 prosodic systems played a significant role in learning
Mandarin tones. The Cantonese tonal system hindered the learning of Mandarin tones, while the Japanese pitch-accent system facilitated the establishment of a new tonal system. The English stress-accent system neither helped nor hindered tone learning. The performance of the English listeners was intermediate between that of the Cantonese and Japanese listeners. These findings are consistent with the Perceptual Assimilation Model (PAM), suggesting that perceptual mapping is not restricted to segments, but can be extended to suprasegmentals ( lexical tones).

With respect to the training approach, learners who received AV Feedback required shorter training periods, and they outperformed learners who received Simple Feedback. These findings imply that the AV training approach employed in the current study facilitates the learning of Mandarin tones and promotes the long-term modification of listeners' tonal properties of L2 tones, thus providing support for its applicability in the training of other tonal languages.

**Keywords:**
L1 prosodic background; AV training; second language acquisition; perceptual assimilation model (PAM); Mandarin tones; cross-language speech perception
To my
beloved parents and Herman
with the appreciation
for their love
and
support

امج
ACKNOWLEDGEMENTS

Hallelujah, thank God for guiding me through all the difficulties and frustrations, so that my thesis could become a reality.

I would like to thank my supervisor Dr. Zita McRobbie for her encouragement and her assistance in many ways throughout the years of my graduate studies, as well as her comments on the thesis. I am especially grateful to Dr Catherine Best for changing her business trip in order to attend my thesis defense, and for her insightful and valuable comments that have led to significant revisions in my thesis. My gratitude also goes to Dr. Alexei Kochetov for his assistance in recruiting subjects for this study and his comments on this thesis. I would like to thank Dr. Murray Munro for his comments on this thesis, especially his suggestions for the selection of the names for listener groups. I also thank Dr. Trude Heift and Dr. Dean Mellow for their questions during my thesis defense.

I would also like to take this opportunity to thank Dr. Wyn Roberts for his help and consistent encouragement throughout the years of my graduate studies. I am sure I will miss the "cuppa-time" that you and we had in these years.

I should express my thankfulness to the Social Sciences and Humanities Research Council of Canada (SSHRC) for granting me the doctoral fellowship, so that I had enough money to pay my subjects, to buy experimental equipment and software, and to cover my conference expenses. Thanks also go to the
Department of Linguistics (SFU), Dr. Zita McRobbie and Dr. Alexei Kochetov for providing me with the financial support in the forms of university fellowships, teaching and research assistantships during my academic years. I would also like to thank Carol, Gladys, Grace, Liz, Rita, and Sheilagh for their exceptional secretarial assistance in the past ten years.

Special thank goes to Germaine Koch for her friendship and her assistance in editing this thesis, and to Penny Simpson (thesis assistant at the SFU library) for her help in fixing some formatting problems. My sincere thank also goes to all the participants in this study.

Finally, my deepest gratitude goes to my parents and family members for their consistent and caring support, to Herman for his unfailing support and his companionship, and to my forever friends: Derek, Ken, Andersen, Esther, Camy, Queenie, and Anita for their endless support, encouragement, and friendships throughout the years.
TABLE OF CONTENTS

Approval................................................................................................................... ii
Abstract................................................................................................................... iii
Dedication............................................................................................................... v
Acknowledgements................................................................................................. vi
Table of Contents................................................................................................. vii
List of Figures......................................................................................................... xii
List of Tables.......................................................................................................... xv

Chapter 1: Introduction............................................................................................ 1
  1.1 Learning Mandarin Tones................................................................................ 1
    1.1.1 Difficulties in Learning Mandarin Tones................................................. 1
    1.1.2 Some Concerns......................................................................................... 2
    1.1.3 Objectives of the Present Study............................................................... 4
    1.1.4 The Organization of this Thesis............................................................... 5
  1.2 Effect of Native Language................................................................................. 6
    1.2.1 Developmental Change: From Infants to Adults...................................... 6
    1.2.2 The L1 effect on Adults’ Perceptions of L2 contrasts............................ 9
      1.2.2.1 Segmental Level.................................................................................. 10
      1.2.2.2 Suprasegmental Level....................................................................... 13
  1.3 Theoretical Models of L2 Speech Perception................................................... 17
    1.3.1 The Speech Learning Model (SLM)......................................................... 17
    1.3.2 The Perceptual Assimilation Model (PAM)............................................. 18
  1.4 Perceptual Training........................................................................................... 22
    1.4.1 Training Segmental Contrasts................................................................. 23
    1.4.2 Training Suprasegmental Contrasts......................................................... 29
    1.4.3 Summary of Previous Perceptual Training Studies.............................. 33
  1.5 The Role of Feedback in Training................................................................... 37
    1.5.1 Types of Feedback................................................................................... 37
    1.5.2 Concerns about Using AV Feedback in Training................................. 41
  1.6 The Four Languages........................................................................................ 44
    1.6.1 Tone Languages....................................................................................... 44
    1.6.2 Prosodic Systems..................................................................................... 45
      1.6.2.1 Mandarin............................................................................................ 45
      1.6.2.2 Cantonese.......................................................................................... 49
      1.6.2.3 Japanese............................................................................................ 52
      1.6.2.4 English............................................................................................. 54
1.7 The Present Study..........................................................................................56
  1.7.1 Research Questions..................................................................................56
  1.7.2 The Outline of the Present Study............................................................59
  1.7.3 Hypotheses..............................................................................................63
    1.7.3.1 H1: The Effect of L1 Prosodic Background......................................63
    1.7.3.2 H2: The Effect of Training Approaches........................................65
    1.7.3.3 H3: Cantonese Listeners and AV Training.................................67

Chapter 2: The Experiment..............................................................................68
  2.1 Participants..................................................................................................68
  2.2 Materials......................................................................................................69
  2.3 Procedure.....................................................................................................74
    2.3.1 Phase 1: Pretest.....................................................................................75
    2.3.2 Phase 2: Training.................................................................................76
      2.3.2.1 Training Procedures......................................................................76
      2.3.2.2 Feedback Type: Audio-visual (AV) vs. Simple (SIM)................78
    2.3.3 Phases 3 and 4: Posttest-Day1 (PTDay1) and Posttest-Day2
      (PTDay2).................................................................................................81
  2.4 Analysis........................................................................................................82
    2.4.1 Analyses for Pretest and Posttest Phases...........................................83
    2.4.2 Analysis for the Training Phase..........................................................84
    2.4.3 Analysis for Training Improvement....................................................84

Chapter 3: Results: Pretest.............................................................................85
  3.1 Percent Correct Scores..............................................................................85
  3.2 A-prime Scores............................................................................................86
  3.3 Tonal Confusions.........................................................................................88

Chapter 4: Results: Training Duration..........................................................94

Chapter 5: Results: Identification Tests on Posttest-Day1 (PTDay1).............98
  5.1 Posttest1 on PTDay1..................................................................................98
    5.1.1 Percent Correct Scores.......................................................................98
    5.1.2 A-prime Scores.................................................................................99
    5.1.3 Tonal Confusions..............................................................................100
  5.2 Gen 1 on PTDay1: Novel Speakers..............................................................103
    5.2.1 Percent Correct Scores.....................................................................103
    5.2.2 A-prime Scores...............................................................................104
    5.2.3 Tonal Confusions............................................................................105
  5.3 Gen 2 on PTDay1: Novel Stimuli.................................................................107
    5.3.1 Percent Correct Scores.....................................................................107
    5.3.2 A-prime Scores...............................................................................108
    5.3.3 Tonal Confusions............................................................................109
  5.4 Gen 3 on PTDay1: Nonspeech (Filtered Speech)......................................111
    5.4.1 Percent Correct Scores.....................................................................111
5.4.2 A-prime Scores .............................................................. 112
5.4.3 Tonal Confusions ...................................................... 114

Chapter 6: Results: Identification Tests on Posttest-Day2 (PTDay2) .......... 118
6.1 Posttest2 on PTDay2 .................................................... 118
   6.1.1 Percent Correct Scores ........................................ 118
   6.1.2 A-prime Scores .................................................. 119
   6.1.3 Tonal Confusions ................................................ 121
6.2 Gen 1 on PTDay2: Novel Speakers .................................. 125
   6.2.1 Percent Correct Scores ........................................ 125
   6.2.2 A-prime Scores .................................................. 126
   6.2.3 Tonal Confusions ................................................ 127
6.3 Gen 2 on PTDay2: Novel Stimuli .................................... 130
   6.3.1 Percent Correct Scores ........................................ 130
   6.3.2 A-prime Scores .................................................. 131
   6.3.3 Tonal Confusions ................................................ 133
6.4 Gen 3 on PTDay2: Nonspeech (Filtered Speech) ................ 136
   6.4.1 Percent Correct Scores ........................................ 136
   6.4.2 A-prime Scores .................................................. 137
   6.4.3 Tonal Confusions ................................................ 139

Chapter 7: Overall Training Improvement & Generalization .................. 143
7.1 Training Improvement: Pretest-Posttest1-Posttest2 ..................... 143
7.2 Tests of Retention of Generalization .................................. 146
   7.2.1 Gen 1: PTDay1 vs. PTDay2 ................................... 146
   7.2.2 Gen 2: PTDay1 vs. PTDay2 ................................... 147
   7.2.3 Gen 3: PTDay1 vs. PTDay2 ................................... 149

Chapter 8: Discussion and Conclusion ........................................... 152
8.1 The Three Hypotheses of the Present Study ............................ 152
   8.1.1 Hypothesis 1: L1 Prosodic Effect ................................ 152
      8.1.1.1 The L1 Prosodic Effect in the Four Phases .......... 153
      8.1.1.2 PAM and Lexical Tone Assimilations ................. 155
   8.1.2 Hypothesis 2: The Training Approach ........................ 163
   8.1.3 Hypothesis 3: Cantonese Listeners and AV Training ....... 166
8.2 Other Issues ............................................................ 169
   8.2.1 The Cross-Language Characteristics of Lexical Tone Perception ........................................ 169
      8.2.1.1 Asymmetrical Patterns in Tonal Perception ........ 169
      8.2.1.2 The Perception of Filtered Speech for Lexical Tones ... 170
   8.2.2 The Influence of Japanese Prosodic Background on Tonal Perception ........................................ 171
   8.2.3 The Effect of Short-term Training on Long-term Tonal Modification ........................................ 172
8.3 Contributions, Limitations, and Future Research ....................... 173
8.4 Conclusion ........................................................................................................ 175

Appendix: Questionnaire .................................................................................. 178

References ........................................................................................................ 179
LIST OF FIGURES

Figure 1-1. Mandarin tonal patterns on the syllable /ta/ spoken by a female native Mandarin speaker (top) and a male native Mandarin speaker (bottom).................................................................................46

Figure 1-2. Cantonese (HK) tonal patterns on the syllable /si/ spoken by a female native Cantonese speaker (top) and a male native Cantonese speaker (bottom). ........................................................................50

Figure 1-3. Japanese pitch-accent patterns on the disyllable ame spoken by a female native Japanese speaker (top) and a male native Japanese speaker (bottom). .......................................................................53

Figure 1-4. English lexical stress patterns on the word subject spoken by a female native English speaker (top) and a male native English speaker (bottom). ........................................................................55

Figure 2-1. A presentation screen in the familiarization block ........................................76

Figure 2-2. A presentation screen in the Pretest ..................................................................76

Figure 2-3. Simple feedback for a correct (left) and an incorrect (right) response........78

Figure 2-4. AV feedback for a correct (left) and an incorrect (right) response....................80

Figure 3-1. Mean percent correct scores (%) in Pretest by the listener in the AV (left) and the SIM (right) groups .................................................................86

Figure 3-2. Tonal confusions in Pretest by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.........................................................91

Figure 3-3. Tonal confusions in Pretest by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................92

Figure 4-1. Mean number of training sessions for the training sets spent by the listeners in the AV groups .........................................................................................95

Figure 4-2. Mean number of training sessions for the training sets spent by the listeners in the SIM groups .........................................................................................95

Figure 5-1. Mean percent correct scores (%) in Posttest1 on PTDay1 by the listener in the AV (left) and the SIM (right) groups .................................................................99
Figure 5-2. Tonal confusions in Posttest1 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 102

Figure 5-3. Tonal confusions in Posttest1 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 102

Figure 5-4. Mean percent correct scores (%) in Gen1 on PTDay1 by the listener in the AV (left) and the SIM (right) groups ......................................................... 103

Figure 5-5. Tonal confusions in Gen 1 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 106

Figure 5-6. Tonal confusions in Gen 1 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 106

Figure 5-7. Mean percent correct scores (%) in Gen2 on PTDay1 by the listener in the AV (left) and the SIM (right) groups ......................................................... 107

Figure 5-8. Tonal confusions in Gen 2 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 110

Figure 5-9. Tonal confusions in Gen 2 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 110

Figure 5-10. Mean percent correct scores (%) in Gen3 on PTDay1 by the listener in the AV (left) and the SIM (right) groups ......................................................... 112

Figure 5-11. Tonal confusions in Gen 3 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 115

Figure 5-12. Tonal confusions in Gen 3 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 116

Figure 6-1. Mean percent correct scores (%) in Posttest2 on PTDay2 by the listener in the AV (left) and the SIM (right) groups ......................................................... 119

Figure 6-2. Tonal confusions in Posttest2 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 123

Figure 6-3. Tonal confusions in Posttest2 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category ......................................................... 124

Figure 6-4. Mean percent correct scores (%) in Gen1 on PTDay2 by the listener in the AV (left) and the SIM (right) groups ......................................................... 125
Figure 6-5. Tonal confusions in Gen 1 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................129

Figure 6-6. Tonal confusions in Gen 1 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................130

Figure 6-7. Mean percent correct scores (%) in Gen2 on PTDay2 by the listener in the AV (left) and the SIM (right) groups..............131

Figure 6-8. Tonal confusions in Gen 2 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................135

Figure 6-9. Tonal confusions in Gen 2 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................136

Figure 6-10. Mean percent correct scores (%) in Gen3 on PTDay2 by the listener in the AV (left) and the SIM (right) groups..............137

Figure 6-11. Tonal confusions in Gen 3 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................140

Figure 6-12. Tonal confusions in Gen 3 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.................................................................141

Figure 7-1. Mean percent correct scores (%) in the Pretest, Posttest1, and Posttest2 by the listeners in the AV (left) and SIM (right) groups.................................................................144

Figure 7-2. Mean percent correct scores (%) in Gen 1 (Novel speakers) on PTDay1 and PTDay2 by the listeners in the AV (left) and the SIM (right) groups.................................................................146

Figure 7-3. Mean percent correct scores (%) in Gen 2 (Novel stimuli) on PTDay1 and PTDay2 by the listeners in the AV (left) and the SIM (right) groups.................................................................148

Figure 7-4. Mean percent correct scores (%) in Gen 3 (Nonspeech) on PTDay1 and PTDay2 by the listeners in the AV (left) and the SIM (right) groups.................................................................149
LIST OF TABLES

Table 2-1. Mean durations (ms) and F0 values (Hz) for the stimuli in the Pretest .................................................................72
Table 2-2. Mean durations (ms) and F0 values (Hz) for the stimuli in the training sessions .......................................................72
Table 2-3. Mean durations (ms) and F0 values (Hz) for the stimuli in Gen1. ...........................................................................73
Table 2-4. Mean durations (ms) and F0 values (Hz) for the stimuli in Gen2. ...........................................................................74
Table 3-1. Mean A' scores for the Mandarin tones in Pretest for the listener groups. .................................................................87
Table 3-2. Confusion Matrices for the Responses in Pretest Performed by the Listener Groups ...............................................88
Table 4-1. Mean Number of Training Sessions for the Training Sets during Training ...................................................94
Table 5-1. Mean A' scores for the Mandarin tones in Posttest1 on PTDay1 for the listener groups ............................................100
Table 5-2. Confusion Matrices for the Responses in Posttest1 on PTDay1 Performed by the Listener Groups ..........................101
Table 5-3. Mean A' scores for the Mandarin tones in Gen 1 on PTDay1 for the listener groups ....................................................104
Table 5-4. Confusion Matrices for the Responses in Gen1 on PTDay1 Performed by the Listener Groups .............................105
Table 5-5. Mean A' scores for the Mandarin tones in Gen 2 on PTDay1 for the listener groups ....................................................108
Table 5-6. Confusion Matrices for the Responses in Gen2 on PTDay1 Performed by the Listener Groups .............................109
Table 5-7. Mean A' scores for the Mandarin tones in Gen 3 on PTDay1 for the listener groups ....................................................113
Table 5-8. Confusion Matrices for the Responses in Gen3 on PTDay1 Performed by the Listener Groups .............................114
Table 6-1. Mean A' scores for the Mandarin tones in Posttest2 on PTDay2 for the listener groups ............................................120
Table 6-2. Confusion Matrices for the Responses in Posttest2 on PTDay2 Performed by the Listener Groups ........................................ 122
Table 6-3. Mean A’ scores for the Mandarin tones in Gen 1 on PTDay2 for the listener groups ................................................................. 126
Table 6-4. Confusion Matrices for the Responses in Gen1 on PTDay2 Performed by the Listener Groups ............................................... 128
Table 6-5. Mean A’ scores for the Mandarin tones in Gen 2 on PTDay2 for the listener groups ................................................................. 132
Table 6-6. Confusion Matrices for the Responses in Gen2 on PTDay2 Performed by the Listener Groups ............................................... 133
Table 6-7. Mean A’ scores for the Mandarin tones in Gen 3 on PTDay2 for the listener groups ................................................................. 138
Table 6-8. Confusion Matrices for the Responses in Gen3 on PTDay2 Performed by the Listener Groups ............................................... 139
Table 8-1. Summary of listener Groups’ score differences between AV and SIM groups in the identifications tests on PTDay1 and PTDay2 ................................................................. 167
CHAPTER 1: INTRODUCTION

1.1 Learning Mandarin Tones

Mandarin Chinese is one of the most popular Asian languages learned by thousands of speakers of other languages around the world. Some of them study Mandarin because of personal interests, while others learn the language for business needs. For speakers of non-tone languages, learning a tone language can be a difficult task. The learners need to learn not only the sound segments, but also the four lexical tones of Mandarin. This thesis reported a study about using a novel Audio-visual (AV) approach to train three groups of naïve listeners from different first language backgrounds to learn the tonal system of Mandarin. It primarily focuses on the influences of the native (L1) prosodic backgrounds and of the AV training approach on the performance of the trainees.

1.1.1 Difficulties in Learning Mandarin Tones

The difficulty experienced by non-native speakers of Mandarin Chinese in learning the lexical tones has been well-documented (Gottfried & Suiter, 1997; Kiriloff, 1969; Miracle, 1989; Shen, 1989). The American, Australian, and Dutch learners in these studies found it particularly difficult to produce and to perceive the Mandarin tone pairs, Tone 2 (mid rising) vs. Tone 3 (falling rising), and Tone 1 (high level) vs. Tone 4 (high falling), because the tones in each pair share similarities in their contour patterns.
According to these studies, learners may misidentify or mispronounce one tone as another one. The consequences could be serious: the incorrect use of tones may lead to miscommunication between interlocutors, because using a different tone on the same syllable will change the meaning of the word. For example, the Mandarin syllable (word) *ma* means “mother” when it is pronounced with Tone 1, but it means “horse” if it is produced with Tone 3.

### 1.1.2 Some Concerns

Learners’ difficulties raise at least two concerns about learning a new tonal system. One concern is that our understanding about the influences of the learners’ native languages on the learning of lexical tones is limited. As mentioned earlier, most previous studies examined the performance of native speakers of a non-tone language. It is still unclear how well native speakers of a tone language will perform when they learn new lexical tones or an entire tonal system. A common belief concerning tone learning is that learners whose native language (L1) is a tone language can learn other tone languages faster and easier than those from non-tone languages because they have experience in using pitch to distinguish lexical items. Recent findings by Wayland & Guion (2004) seemed to support this view. The authors found that native speakers of Mandarin outperformed native speakers of English in distinguishing a pair of Thai level tones (mid vs. low) after one week of auditory training. They explained that Mandarin trainees might tune “their existing perceptual system to guide their perception of the acoustic properties of new tones” (Wayland & Guion, 2004, p. 707). However, it is uncertain if adults with tonal experience will always learn
second language (L2) tones better than those who do not have tonal experience in their native languages. In addition, there is considerable previous research showing that adults’ difficulties in producing and perceiving L2 segments are largely attributable to their L1 phonological systems (e.g., Best, 1995; Flege, 1995). Since tone is part of the phonology of native speakers of a tone language, it is reasonable to expect that their L1 prosodic systems (e.g., lexical tonal systems) may also influence their L2 tone learning. As far as the author knows, the effects of learners’ L1 prosodic systems on mature adults’ difficulties in learning L2 tones has not yet been examined.

Another concern is how to assist learners to learn the tones more effectively. Some researchers in second language acquisition (SLA) have recognized that directing learners’ attention to noticing linguistic forms/features is important for learning an L2 (Ellis, 1997; Grass, 1988; Schmidt, 1990, 1995; Sharwood Smith, 1993; and Tomlin & Villa, 1994). Features noticed by learners will become the intake for L2 learning (Schmidt, 1995). Previous laboratory studies have already demonstrated that “lengthy periods of auditory training” (Ortega-Llebaria, Faulkner & Hazan, 2001, p. 40) improve learners’ abilities to perceive and to produce non-native sound contrasts at both the segmental and suprasegmental levels (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Jamieson & Morosan, 1986, 1989; Lambacher, 2001; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991; Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2003; Wayland
& Guion, 2004). More importantly, many of these studies have shown that employing appropriate procedures¹ during training (e.g., using identification tasks and natural stimuli) can shift learners’ attention to particular acoustic cues for target L2 segments. As a result, the learners’ performance can be improved after training. Some studies have also indicated that the learners’ new knowledge gained from training could be transferred to novel words and novel speakers, and that the long-term effects of training were retained months later. However, the procedures employed in the above studies have neither included nor focused on the type of feedback employed during training. In fact, learners have been found to benefit more from feedback types with dual- or multi-channel input than those with a single channel input (e.g., AV_{audio-pitch \ graph} vs. Audio-only, and AV_{audio-pitch \ graph-text} vs. Simple_{text(correct-incorrect)}), when learning some suprasegmental features, such as intonation (de Bot, 1983), and lexical tones (So, 2003). However, the long-term effects of this kind of AV feedback need further investigation, because the findings of the studies were based on the data obtained from an immediate posttest.

1.1.3 Objectives of the Present Study

The present study aims to explore the theoretical and methodological issues of learning a tonal system by naïve listeners. Theoretically, it intends to determine the extent to which listeners’ L1 prosodic backgrounds affect their learning performance. Methodologically, an attempt is made to verify if using a novel AV training approach, which directs the listener’s attention to the essential

¹ These procedures will be discussed in a later section.
perceptual cues of lexical tones during training sessions, will facilitate the process of learning lexical tones. To accomplish these goals, the present study will examine the effects of (i) L1 prosodic backgrounds and (ii) a novel AV training approach on the perceptual performance of four Mandarin tones by three naïve listener groups (Cantonese, Japanese, and English) after training.

1.1.4 The Organization of this Thesis

The organization of this thesis is as follows. The following sections of this chapter will provide background information about the influence of native language on human speech perception from infancy to adulthood (with particular attention being paid to L1 constraints on adults’ non-native sound contrasts), models for L2 perception, perceptual training, and the role of feedback in training (Sections 1.2- 1.5, respectively). Next, the prosodic systems of the four languages in this study (Mandarin: the target language; Cantonese: a tone language; English: a stress-accent language; and Japanese: a pitch-accent language) will be described for later comparison of listeners’ performance before and after training (Section 1.6). Finally, the outline of the present study together with research questions and hypotheses will be provided (Section 1.7). In Chapter 2, the four phases of the experiments of the present study (Pretest, Training, Posttest-Day1, and Posttest-Day2) will be outlined, followed by the description of the methodology, the procedures, and the methods of analyses. Then, Chapters 3 - 6 will report the results for the four experimental phases. Chapter 3 will focus on the listeners’ performance in the Pretest. Chapter 4 will report the training durations used by the listeners in the two training approach
groups. Chapter 5 will centre on the results of the four identification tests on Pottest-Day1 (PTDay1): a posttest (Posttest1, which was identical to the Pretest), and three generalization tests (Gen1, Gen2, and Gen3). Chapter 6 will report the results of the four identification tests on Pottest-Day2 (PTDay2): a posttest (Posttest2 that was identical to the Pretest and the Posttest1) and three generalization tests (Gen1, Gen2, and Gen3). Chapter 7 will present the results of comparisons of the listeners' performance in the identification tests across time. Finally, discussions and a conclusion will be provided in Chapter 8.

1.2 Effect of Native Language

The effect of native language has a long-term influence on the process of human language development. From as young as 6 months old, the ambient language models and shapes infants' abilities to perceive human speech sounds. As a result, adults often experience difficulties in producing and perceiving phonemic contrasts that do not exist in their native languages² (Aslin & Pisoni, 1980; Best & Strange, 1992; Burnham, 1986; Lisker & Abramson, 1970; Werker & Tees, 1984a). More importantly, the listeners' native languages could also affect their learning of a new language.

1.2.1 Developmental Change: From Infants to Adults

As early as the first year of life, native language starts playing an important role in the language development of infants, directing their attention to

² However, Best, McRoberts, & Goodell (2001) showed that mature English adults were able to perceive Zulu clicks, which were not contrastive in English. The authors suggested that the clicks were nonspeech, rather than phonological elements.
the sounds of their ambient language. Young infants under 6 months of age are able to discern a wide range of universal phoneme contrasts that do not exist in the native languages. For instance, English-learning infants can discriminate non-native contrasts in French and Czech (Trehub, 1976), Hindi (Werker, Gilbert, Humphrey, & Tees, 1981), Nthlakampx (Werker & Tees, 1984a), Spanish (Aslin, Pisoni, Hennessy, & Percy, 1981), Zulu (Best, McRoberts, & Sithole, 1988), German (Polka & Werker, 1994), and Mandarin (Tsao, Liu, Kuhl, & Tseng, 2000).

However, infants’ abilities to discriminate non-native contrasts decline towards the end of the first year of life (Best & McRoberts, 2003; Werker & Tees, 1984a). Their perceptual abilities change from language-universal to language-specific (i.e., to their ambient language). Specifically, studies have demonstrated that infants around 10 months of age have difficulties discriminating non-native consonant contrasts that are similar to those experienced by adults. For example, Werker et. al. (1981) compared English infants aged 6 to 8 months, English adults, and Hindi adults on their abilities to discriminate both the Hindi unaspirated retroflex vs. dental stop /t̪a-ːt̪a/ and the Hindi dental voiceless aspirated vs. the voiced aspirated voicing distinction /tʰa-ːdʰa/. Their results showed that English infants’ discrimination abilities are comparable to those of Hindi adults, but that English adults discerned poorly the contrasts in question. In a follow-up study, Werker & Tees (1984a) examined groups of English infants’ perceptual abilities to discriminate two non-English contrasts: the Hindi unaspirated retroflex vs. dental stop, /t̪a-ːt̪a/, and the Nthlakapmx (Salish) glottalized velar vs. uvular stop /ki-ːqi/. They found that the English infants at 10-
12 months of age no longer discriminated the non-English contrasts, and that their performance was similar to that of the English adults in Werker, et al. (1981). For vowels, Polka & Werker (1994) examined the abilities of a group of English infants in discriminating German vowels. They found that the beginning of perceptual declination for vowels was around 6 months of age. By 10-12 months of age, the discrimination of the non-native vowels was even poorer. Similar findings on vowels were also reported by Kuhl, Williams, Lacerda, Stevens, & Lindblom (1992). For lexical tones, Harrison (2000) found that 6 to 8 months old Yoruba infants discriminated between different Yoruba tone categories (a language which has a three-way contrast for tone) better than did English infants. Mattock & Burnham (2006) found that English infants’ discrimination of lexical tone declined between 6 and 9 months of age, but the same pattern was not found in Chinese infants.

Concerning the perceptual abilities of young children and mature adults, Werker & Tees (1983) found that children at ages 12, 8, and 4 years performed as poorly as English adult listeners when discriminating between Hindi voicing contrasts (the Hindi dental voiceless-aspirated vs. breathy voiced distinction /htha/-dha/). This suggests that the perceptual declination has begun some time before 4 years of age. Other studies also found that native English speakers had difficulties in discerning the difference between certain non-native sound segments, such as the Hindi voiceless dental-retroflex stops (Werker, et al., 1981), the Farsi voiced velar-uvular stop place contrast /g-G/ (Polka, 1991), the French front rounded vowels /y-â/ (Gottfried, 1984) and the nasal vowels /a-á/
(Trehub, 1976), and the Czech palatal fricative and fricative trill /3-7/ (Trehub, 1976). As for non-native English speakers, the phenomenon that Japanese speakers have difficulties in perceiving the differences between English /r/ and /l/ has attracted a great deal of attention from researchers over the past three decades (e.g., Goto, 1971; MacKain, Best, & Strange, 1981; Yamada, 1995; Logan, et al., 1991).

Thus, the findings of the previous studies clearly indicate that language-particular categorical patterns of perception begin to emerge in the second half of the first year of life, when infants start losing their sensitivity to non-native sound contrasts. Further, Werker (1995) suggested that developmental change may be caused by an attentional or perceptual reorganization that occurs some time before 4 years of age. This kind of language-particular attunement or perceptual learning from infancy to adulthood involves a shift of attention (i) to the relevant information (e.g., perceptual properties) and (ii) away from irrelevant stimulus information from the native language (Best, 1994; Jusczyk, 1994; Pisoni, Lively, & Logan, 1994).

1.2.2 The L1 effect on Adults' Perceptions of L2 contrasts

Throughout the course of language development, human perception of non-native sounds is strongly constrained by the phonological system of one’s native language (Best, 1995; Flege, 1995; Polka, 1991; Strange, 1995). Thus, when learning a new language, one’s L1 phonological system may impose constraints on L2 learning, and hinder the establishment of new L2 sound
categories. Previous studies have demonstrated examples at both the segmental and suprasegmental levels.

1.2.2.1 Segmental Level

Miyawaki, Strange, Verbrugge, Liberman, Jenkin, & Fujimura, (1975) compared the /r/-/l/ perception difference between native Japanese listeners of English and native English listeners. Since Japanese does not have the liquid contrast between /r/ and /l/, it was expected that Japanese listeners would experience difficulties perceiving the English /r/-/l/ contrast. The authors created synthetic English /r/ and /l/ stimuli with various third formant frequency (F3) patterns. Listeners performed both identification and discrimination tasks. Results revealed that English listeners showed typical categorical perception patterns for /r/ and /l/, but that Japanese listeners did not show any consistent pattern in the discrimination task.

However, other studies showed that given more L2 experience (e.g., intensive training in English conversation), the ability of learners to discern the /r/-/l/ contrasts could be improved. For instance, MacKain et al. (1981) replicated and extended the study of Miyawaki et al. (1975) by (i) using two groups of Japanese learners of English (experienced vs. inexperienced) and a group of native English speakers (as a control group), and (ii) including a synthetic stimuli series that had variations in both the spectral patterns of the second and third formant frequencies, and the temporal patterns of the first formant frequency. All listeners were asked to complete three tasks: identification, and two discriminations (AXB and Oddity). The results showed that the native English listeners produced
typical identification and discrimination patterns for the /r-l/ contrast. The inexperienced Japanese listeners performed as poorly as those in the study by Miyawaki et al. (1975), while the performance of the experienced Japanese listeners were compatible to those of the native English listeners.

Flege & Wang (1989) examined how L1 phonotactic constraints affected the L2 learners' sensitivity to English /t/ and /d/ in word-final position, and investigated the effect of training on improving the sensitivity to the two non-native segments. They compared the performance in identifying English /t/ and /d/ contrasts (closure voicing and release burst cues had been removed) by three groups of L2 learners (Cantonese, Mandarin, and Shanghainese) before and after a brief training session. These languages differed in their L1 phonotactic constraints with respect to the word final stops: Cantonese allows unreleased stops /p, t, k/; Mandarin does not allow any stop in word final position; and Shanghainese has only glottal stops /ʔ/ in word final position. The Cantonese listeners were expected to have the best performance, while the Mandarin listeners would have the worst results, and the Shanghainese performance was predicted to be intermediate to that of the other two groups. Stimuli were presented in three 6-minute blocks as pre-training, training, and post-training. Feedback indicating a correct or an incorrect response was given after every participant's answer during the training phase. The results indicated that training improved the performance of all three groups of listeners. In addition, the Cantonese listeners outperformed the Shanghainese listeners whose score in turn was higher than that of the Mandarin listeners. Further, the Cantonese
listeners' sensitivity to the English /t-d/ contrast in the word-final position was significantly higher than that of the Mandarin listeners. Flege & Wang concluded that Cantonese phonotactic constraints might help Cantonese learners of English to pay more attention to the ends of stimuli, and therefore learning the non-native contrast (/t-d/) was easier. Moreover, the fact that the listeners benefited from the training suggests that using an appropriate sensory stimulation can modify learners' strategies of attentional allocation rooted in their L1 acquisition.

Ingram & Park (1997) examined the production and perception of Australian English vowels (/i: ɛ æ æ:/) by Japanese and Korean learners of English. Four groups of Japanese and Korean speakers (from 24-38 years old) were recruited according to their experience with English-speaking environments (experienced vs. inexperienced). The authors further indicated that younger Korean adults, unlike older Korean speakers, did not have the /e-ɛ/ contrast in their L1 system. Consequently, a new group of older native Korean speakers (over 50 years of age) was also employed to serve as a control group. It was found that Japanese learners tended to make use of their phonemic vowel length feature when producing and perceiving the Australian English vowels. This pattern was not observed in the Korean learners' performance, presumably because Korean does not have the vowel length feature. For the /e-ɛ/ contrasts, the Korean learners' perceptions and productions of the Australian /e-æ/ contrast were poorer than those of the Japanese learners and the older Korean listeners.
The findings indicate that learners’ L1 phonological systems do exert influences on their L2 performance.

Gottfried & Suiter (1997) examined the effects of linguistic experience on the perception of Mandarin vowels and tones. Both native Mandarin speakers and native English learners of Mandarin (who had studied Mandarin for 5.11 years on average) participated in the study. Stimuli were sixteen syllables, di, da, du, dou\(^2\) (4 vowels x 4 tones). Comparisons were made of the listener groups’ correct identification scores in four testing conditions: intact syllables, of syllables with the initial and final portions removed (center-only), of syllables with the centers removed (silent-center), and of syllables with only the initial transition presented (initial-only). As expected, the native English learners made more errors overall than did the native Mandarin speakers. The native English speakers consistently misidentified the diphthong -uo as -u, and performed poorly in tonal identification. Gottfried & Suiter concluded that both the native Mandarin speakers and the English learners of Mandarin use similar acoustic cues in perceiving Mandarin vowels and tones. However, because English does not employ pitch variation to convey meaning in its system (i.e., tones are not phonemic), the native speakers of English had more difficulty in perceiving Mandarin tones than did their Mandarin counterparts.

1.2.2.2 Suprasegmental Level

Evidence for L1 constraints on L2 suprasegmental features has also been documented. However, most of the studies focused on stress or intonation

\(^2\) These are in Mandarin Pinyin. In IPA, they are [ti, ta, tu, tuo]
patterns learned by L2 learners (e.g., Archibald, 1992; Jun & Oh, 2000; Guion, Harada, & Clark, 2004).

Archibald (1992) investigated the acquisition of English stress patterns by adult L2 learners by examining adult native Polish speakers' productions and perceptions of English stress patterns. For the production task, participants were instructed to read English words in isolation and in sentences. For the perception task, they were asked to listen to English words and to put stress placements on the words. Archibald found that the Polish speakers transferred their L1 metrical stress patterns (main stress always falls on the penult) to the production and perception of English words. For example, the Polish learners tended to stress on the penult position of English bisyllabic words. For English words with final stress, such as maintain, and appear, the Polish learners produced as maintain, and appear, respectively. In a later study, Archibald (1993) found that as was the case with the Polish learners of English, Spanish learners of English transferred the stress patterns of their L1 when producing and perceiving English words.

Jun & Oh (2000) examined the intonation of Seoul Korean produced by American English speakers of different levels of proficiency of Korean. Four English speakers of Korean (2 advanced learners vs. 2 less advanced learners) were asked to produce forty Korean sentences with different intonation patterns. Results showed that the advanced English learners of Korean produced more native-like intonation patterns than did the less advanced learners of Korean. However, it was found that the advanced learners still had difficulties producing Korean Wh-questions and yes/no-questions. The authors found that their L1
phonological properties of intonation influenced the English learners' Korean sentential productions. Thus, the findings suggested that an L1 effect can still be observed in advanced English learners of Korean.

Guion, Harada, & Clark (2004) examined early and late Spanish-English bilinguals' stress placement on English words. Three groups of participants (early and late Spanish-English bilinguals, and native English speakers) were asked to produce and to make perceptual judgments on a list of non-words of varying syllabic structures in noun and verb sentence frames. They found that all participant groups showed significant effects of stress patterns of phonological similarity and lexical class (nouns vs. verbs). The late bilinguals exhibited more L1 transfer in their stress patterns than did the early bilinguals. Although the early bilinguals' performance was comparable to that of the native English speakers, the effect of syllabic structure on these early bilinguals was slightly different from that of the native English speakers and the late bilinguals. Thus, the findings of this study reveal an L1 effect not only in monolingual adults' perceptions of non-native contrasts, but also in both early and late bilinguals' stress assignment patterns.

Wayland & Guion (2004) examined the performance of naive native speakers of English and Mandarin (from Taiwan and mainland China) in learning a contrastive pair of Thai tones (mid vs. low) before and after training. The effect of L1 background of the participants was examined. Native speakers of Thai served as a control group. Eight minimal pairs of syllables (CV, CVV, and CVN)

---

4 The stress assignments on non-words were affected by the listeners' native stress patterns of known (real) words in their dominant language (English or Spanish).
were used as stimuli in discrimination tasks for pretest, training, and posttest. The training phase lasted for 5 days, and involved five 30-minute sessions. They found that participants' performance in discriminating the mid-low tone contrasts improved significantly after training. Further, the Chinese listeners outperformed the native English speakers. Wayland & Guion suggested that the Chinese speakers' prior experience with the Chinese tonal system might have assisted their learning of the two Thai tones.

A number of issues associated with the aforementioned L2 speech perception studies deserve further discussion. First, the impact of an L1 system on learning new L2 contrasts at the initial stage of learning is still unclear. It is unclear because the majority of the above studies employed bilingual learners who had both L1 and L2 knowledge and experience. The findings of these studies in fact suggest that the learners' L1 systems are still affecting their learning of a second language. The influence of an L1 system will be stronger on the performance of the late bilinguals (the less advanced or inexperienced learners) than on that of the early bilinguals (the advanced or experienced learners). Thus, in order to have a better understanding of the influence of an L1 system on learners' performance at the initial stage, the focus should be on naïve listeners (see Wayland & Guion, 2004), because this will exclude the influence of L2 knowledge and experience on the listeners' learning performance.

Second, while the majority of previous studies focused on segments, little is known about the L1 effect on non-native suprasegmentals. Some studies have examined the transfer of stress patterns (e.g., Jun & Oh, 2000; Guion et al.,
2004), but only a few have attempted to explore the L1 effect on lexical tone training (Wayland & Guion, 2004).

1.3 Theoretical Models of L2 Speech Perception

Among existing working models of L2 speech perception, the Speech Learning Model (SLM) (Flege, 1995, 1999, 2002, 2005) and the Perceptual Assimilation Model (PAM) (Best, 1994, 1995; Best et al., 2001; Best & Tyler, 2006) are commonly used to account for an adult's difficulties when perceiving non-native segments. Despite the fact that both models recognize and predict the effects of a L1 system on learning L2 sound segments, the two theoretical frameworks approach the issue from different perspectives.

1.3.1 The Speech Learning Model (SLM)

Flege's SLM (1995, 1999) focuses on the issue of L2 learning, aiming to explain "age-related limits on the ability to produce L2 vowels and consonants in a native-like fashion" (Flege, 1995, p. 237). Its main concern is "the ultimate attainment of L2 pronunciation" (Flege, 1995, p. 238). Therefore, studies employing the SLM have focused on "bilinguals who have spoken their L2 for many years, not beginners" (Flege, 1995, p. 238). Although the SLM has been slightly modified over the years, the central idea of the model remains unchanged. According to Flege (2005), the aim of the SLM is to account for how individuals learn or fail to learn to produce and perceive segments in a second language at the phonetic level.
The SLM proposes that our capacity for speech learning remains intact across the life span. For bilingual speakers, the model posits that the phones of their L1 and L2 systems exist in a common phonological space (at a position-sensitive allophonic level), interacting with each other (Flege, 1995, 1999, 2002). The SLM posits that learners perceive L2 and L1 phonetic categories through the (cognitive) mechanism of equivalence classification. If the L2 and L1 phonetic categories are very similar, the existing L1 categories may block the formation of the new L2 categories. However, the greater the phonetic dissimilarities between the closest L1 and L2 categories perceived by learners, the greater the chance that a new phonetic category will be formed.

1.3.2 The Perceptual Assimilation Model (PAM)

Best's PAM (1994, 1995) focuses on the issue of developmental change from infancy to adulthood. It also focuses on how the adult's language experience has resulted in systematic influences from both the phonological and phonetic properties of the native language(s) on perception of non-native contrasts. Human perception of non-native speech sounds is gradually constrained by acquired native phonology as a function of experience. The PAM aims to account for the wide range of listeners' performance when perceiving diverse non-native sound contrasts (Best & McRoberts, 2003). For adults who are learning a second language, the central premise of the PAM is that naïve adult listeners have a strong tendency to perceptually assimilate non-native phones to the native phonemes that they perceive as the most similar (Best, 1994, 1995; Best, McRoberts & Goodell, 2001). Assimilations are based on the
perception of articulatory gesture information of speech. During this perceptual process, a non-native phone may be perceptually assimilated either as a *Categorized* exemplar for some L1 phonemes, an *Uncategorized* segment that falls somewhere in between two native phonemes, or as a *Nonassimilable* nonspeech sound that has no identifiable similarity to any native phoneme.

The PAM (Best, 1994, 1995) further posits that listeners’ abilities to discern the differences between non-native phones are predictable and dependent on how the non-native phones assimilate to native segments. There are several possible types of assimilations.

- **Two Category assimilation (TC):** Two non-native phones assimilate to two phonetically similar native phones separately. Discrimination between the non-native contrast is expected to be excellent.
- **Single Category assimilation (SC):** Two non-native phones assimilate to a single native phoneme. Poor discrimination is predicted.
- **Category Goodness difference in assimilation (CG):** When both non-native phones assimilate to a single category, one tends to assimilate better than the other. Moderate to excellent discrimination is anticipated.
- **Uncategorized-Categorized assimilation (UC):** One non-native sound is assimilated as a native category, while the uncategorized one falls in the phonetic space, outside any particular native categories. Discrimination is expected to be good.
- **Uncategorized-Uncategorized assimilation (UU):** If both non-native phones are not categorized, they will fall within the phonetic space but
outside of any particular native categories. Discrimination is still affected by L1 phones for uncategorized assimilations, but less so than for categorized ones. Moreover, the L1 effects are spread across the several L1 phonemes that are perceived as similar to the non-native phone. Discrimination should range between fair to good, depending on how similar the non-native phones to each other and to the closest L1 phonemes are perceived.

- **Non-Assimilable (NA):** If the phone is completely dissimilar to any native phoneme, it will be perceived as a nonspeech sound. In this condition, native phonology does not have any influence on it. Discrimination is expected to be good.

In addition, the PAM predicts that there is a gradient of discrimination levels for the assimilations. Discrimination of TC is better than CG which in turn is better than SC (i.e., TC > CG > SC). In other words, adults will have very good discrimination for the non-native phones of the TC type, but they will have very poor perception for the phones of the SC type. This prediction was supported in Best et al. (2001) in which American English-speaking adults' were asked to perceive three isiZulu consonant contrasts: voiced vs. voiceless lateral fricatives (TC), voiceless aspirated versus ejective velar stops (CG), and plosive versus implosive bilabial stops (SC). They found that the listeners' discrimination performance for the three pairs of contrasts ranged from excellent (TC) to poor (SC), while their scores for the CG type of assimilation fell between the two (i.e., good).
Existing L1 phonology may hinder, enhance, or exert no effect on the discrimination of non-native contrastive sounds. Native phonology will enhance discrimination when two non-native phones are assimilated separately. On the other hand, native phonology may hinder the ability to distinguish the phones when they are assimilated into a single L1 phoneme. In the case that the non-native phones are perceived as non-speech sounds, native phonology will neither help nor hinder discrimination.

It should also be mentioned that the PAM does not assume that the effect (or constraint) of native language on perception of non-native contrasts is "absolute or permanent" (Best, 1994, p. 173). Indeed, adults' difficulties in perceiving non-native contrasts could be improved through the naturalistic experience of the L2 sounds. For instance, MacKain et al. (1981) found that experienced Japanese listeners who had been living in the United States for 28 months outperformed inexperienced Japanese listeners (9 months of residence) in identifying and discriminating synthetic English contrasts for /r/ and /l/. In addition, L2 experience can also be gained through laboratory training. Supporting evidence can be found in several training studies (e.g., Logan et al., 1991; Pisoni, Aslin, Perey, & Hennessy, 1982), as well as in those that shifted trainees' attention to salient acoustic cues (e.g., Francis & Nusbaum, 2002; Jamieson & Morosan, 1986, 1989) (see the discussion in Section 1.4).

In sum, while the SLM centres on L2 learning by bilinguals at varying stages (e.g., early vs. late bilinguals), the PAM focuses on naïve listeners' first-time encounter with completely unfamiliar non-native sound contrasts (Best &

---

5 See Section 1.2.2.1 for the details of the study.
Tyler, 2006). Since one of the purposes of the present research is to examine the effect of L1 prosodic backgrounds on learning a new L2 (Mandarin) tonal system, only naïve listeners without any prior L2 tonal experience before training will be recruited. Thus, analyses of the present study will be placed within the framework of the PAM. In addition, although the PAM originally focused on the perception of speech at the segmental level, Hallé, Chang & Best (2004) have recently suggested that the PAM could be applicable at the suprasegmental level as well. This study will further explore this possibility.

1.4 Perceptual Training

As discussed in Section 1.2, the L1 effect on L2 learning is inevitable, especially for late bilinguals. However, it is widely believed that “the mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning” (Flege, 1995, p. 239). A method to assist L2 learners to improve the production and perception of non-native sound contrasts is to provide them with perceptual training. The ultimate goal of perceptual training is to modify the mature adults’ perception of non-native contrasts by shifting their attention to phonetic information that is previously unattended. As a result, a learner can establish the L2 sound categories (see Logan et al., 1991). In terms of methodology, previous studies have used various approaches to accomplish this goal. The results of these studies have enriched our knowledge of administering perceptual training so that training supports the learning of non-native contrasts more successfully.
1.4.1 Training Segmental Contrasts

Pisoni, Aslin, Perey, & Hennessy (1982) employed synthetic stimuli to examine the effect of training on native English speakers’ discriminations of three contrastive categories of Voice Onset Time (VOT): (pre)voiced, voiceless unaspirated, and voiceless aspirated. The experiment lasted for four consecutive days (1 hour per day) employing a pretest-posttest paradigm. A single block training was conducted on Day 1. A total of 240 trials were administered as identification tasks. Feedback indicating the correct response was provided after each trial. Half of the subjects (6 out of 12) who reached a predetermined criterion of at least 85% correct identification were selected for subsequent tests (identification and discrimination tests) administered over the following three days. The authors found that the naïve English listeners successfully perceived the differences of the three VOT categories, demonstrating that adults’ perceptual mechanisms could be modified easily by employing simple laboratory techniques over a short period of time.

Tees & Werker (1984) examined the impact of linguistic experience and training on the perception of non-native voicing and place contrasts. Two groups of native English speakers (15 listeners each) were trained to perceive the voicing and place contrasts from two pairs of contrastive Hindi segments: voiceless aspirated vs. breathy voiced, /tʰa/-/dʱa/ and retroflex vs. dental /ʈa/-/ʈa/. They used a multiple natural token paradigm, and all speech samples were spoken by a native Hindi speaker. For the pretest, training, and posttest phases, a predetermined criterion of 30 change trials was used to determine if a listener
was able to discriminate the contrast or not. In the pretest, trainees were tested in a simple category-change discrimination paradigm. The listeners were instructed to press a button when they heard a change in the given stimuli. Training was done by way of the self-monitored presentation of sounds. The trainees listened to a sequence of CV exemplars from one category followed by a sequence of CV exemplars from the contrasting category. They controlled the timing of the category change. Feedback was given only for the correct response with the illumination of a light inside a smoked plexiglass box. Each training block consisted of 50 category changes. A trial test was given to evaluate their ability to discriminate the sound contrasts. If no improvement was evidenced, then the training would continue for a maximum of six blocks, after which the training would be terminated. The trainees took the posttest after a month (30-40 days). During the posttest, the predetermined criterion was re-used to determine if a listener was able to discriminate the contrast or not after training. For the voicing contrast pair /tʰa-dʱa/, 14 out of 15 listeners were able to discriminate the contrast. For the place contrast pair (/tə-θə/), only 3 out of 15 listeners reached the criterion.

Werker & Tees (1984b) later investigated if feedback with correct answers would help native English adults to better discriminate two non-English contrasts. For each trial, participants were given a pair of stimuli. They were asked to determine if the stimuli were from the same or different sound categories. The researchers provided the English listeners with a brief laboratory training session (a block of 34 trials), and a sheet indicating answers for the paired syllable
presentations. Results showed that the English listeners improved their perceptual discrimination for the Hindi voicing contrast /\textipa{h\textipa{a}-\textipa{d\textipa{a}}}/, but not for the place contrast pair /\textipa{\textipa{a}-\textipa{\textipa{a}}}/. Thus, the findings of these two studies suggest that some non-native contrasts might be more difficult to learn than others.

Jamieson & Morosan (1986) suggested three criteria for the successful training of non-native contrasts. The training method should (i) place the target stimuli in a phonetic context, (ii) use an identification task with feedback, and (iii) present sequencing training stimuli so that trainees’ attention will shift to focus on critical acoustic parameters. In their study, Jamieson & Morosan (1986) trained Canadian French listeners to perceive the voicing contrast of English dental fricatives /\textipa{\textipa{e}}/ and /\textipa{\textipa{\textipa{o}}}/, which were frequently perceived as French /\textipa{\textipa{t}}/ and /\textipa{\textipa{d}}/, respectively. The experiment was a 2-day event employing a pretest-posttest paradigm. The listeners performed both identification and discrimination tasks for the pretest and posttest. During the training phase (2 sessions totalling approximately 90 minutes), a perceptual “fading technique”\(^6\) was used to present training stimuli to listeners who were instructed to identify the target sounds. Synthetic exemplars that were easily identifiable were presented at the beginning, while those that were more difficult to distinguish were presented at the end. Feedback was also given to the listeners when an incorrect response was made by illuminating a small white light above the response button. The

\(^6\) The fading technique was introduced by Terrace (1963). It provides subjects with clearly discriminable stimuli at the beginning of the training, and then increases the difficulty at the end of the training. Jamieson & Morosan (1986, p.205, 207) believed that fading helped trainees to focus their attention on the “critically relevant acoustic cues for given contrastive segments, and reduce the sensitivity to other acoustic cues”.

25
results showed that training with synthetic stimuli through the fading technique could successfully transfer the training from synthetic to natural tokens.

In a subsequent study, Jamieson & Morosan (1989) compared the effects of two training approaches, fading technique training vs. prototype training (i.e., using single good instances for each of the target categories), on Canadian French listeners' identification performance. The experimental procedure was similar to the one in Jamieson & Morosan (1986). It was found that learners trained with the fading technique identified the synthetic stimuli better than those trained with the prototype training. However, the scores for natural stimuli were comparable to each other.

Pisoni and his colleagues (Lively, et al., 1993; Lively, et al., 1994; Logan, et al., 1991; Bradlow, et al., 1997) conducted a series of large-scale training studies in which they trained Japanese learners of English to perceive and produce English /r/ and /l/. In their first study (Logan et al., 1991), they aimed to develop a set of procedures that would facilitate the learning of non-native contrasts by shifting trainees' attention to critical attributes of the stimuli. They suggested the procedures should involve (i) using forced-choice identification, (ii) using several minimal pairs of the target /r-l/ contrast in different phonetic environments produced by different talkers, (iii) using natural speech tokens, and (iv) evaluating not only learners' performance before and after training, but also their perceptual abilities for novel words and talkers.

In the study, the stimuli for the pretest and posttest were identical. The training phase lasted for 3 weeks, and involved fifteen 40-minute sessions with a
large number of words in different contexts spoken by different speakers. Feedback was given during training. If the Japanese listeners gave a correct response, the next trial was presented. If the Japanese listeners gave an incorrect response, a light indicated the correct answer on the response box, followed by a second presentation of the stimulus. Two generalization tests included novel words produced by both a novel speaker and an old speaker (one whose speech samples were already used during the training). Results showed that the training procedures were effective in assisting the Japanese learners to improve their perception of English /r-l/ contrasts, an L2 distinction not present in their L1 system.

In subsequent studies, Lively et al. (1993, 1994) employed the same methodology used in Logan et al. (1991). By using high-variability procedures, Lively et al. (1993) found that Japanese learners of English were able to transfer the knowledge gained from training to novel words spoken by both novel and old speakers, if they were trained with speech samples produced by multiple talkers during training. However, if only a single talker was employed in the training sessions, their performance was worse than that of trainees who received multi-talker input. Lively et al. (1994) further found that the high-variability procedures not only improved the perceptual performance of Japanese learners of English (in Japan) in the posttest and generalization tests, but also helped learners to maintain the knowledge gained as late as 3 and 6 months after training. Thus, this study demonstrated that the high-variability training paradigm shifted
trainees' attention to the acoustic cues that promoted a long-term modification of the Japanese learners' phonetic perceptions.

Bradlow et al. (1997) used the high-variability training paradigm with slight modification to train Japanese learners of English (in Japan) to learn the contrast between /r/ and /l/. Participants' perceptions and productions of the target sounds were assessed. The training period lasted for 3-4 weeks and involved 45 sessions. Feedback was provided during training in a slightly different way. They used a chime to signal a correct response, and a buzzer to indicate an incorrect response. The results indicated that the effects of perceptual training not only facilitated trainees' perceptual identification of the /r-l/ contrasts, but also improved their accuracy in /r/ and /l/ productions as identified by native English listeners.

Francis & Nusbaum (2002) investigated how naïve listeners re-weighted the existing phonetic dimensions in their perceptual space when learning non-native contrasts. They trained naïve English listeners to identify the three classes of Korean voiceless stops with three voicing contrasts using a pretest-posttest paradigm. Trainees took part in three sessions (60, 40, and 60 minutes) on three different days. Stimuli were spoken by a native Korean speaker. Stimuli for the pretest and posttest were only bilabial stops (except for the bilabial combinations used for the pretest and posttest) followed by the vowel /a/, while the training stimuli were all the stop consonants followed by one of the vowels /a/, /o/, and /i/. The pretest and posttest were administered on Day 1 and Day 3. The identification task was provided to participants with three choices of bilabial
stops. Two training blocks were given to the participants on Day 2 and Day 3, during which the participants were given an identification task with nine choices of stops. In each trial, immediate feedback was given for both correct and incorrect responses, followed by the second presentation of the stimulus. Francis & Nusbaum found that the perceptual training improved the native English listeners' perceptions of the different Korean voiceless stops, indicating that the training assisted these trainees to shift their attention to a previously unattended dimension of phonetic contrast.  

1.4.2 Training Suprasegmental Contrasts

With respect to perceptual training at the suprasegmental level, Leather (1990) was probably the first study that examined the effect of laboratory training on learning Mandarin tones. Two groups of Dutch speakers served as trainee groups, and the stimuli were the four tones on the Mandarin syllable *yu*. Trainees in one group were given perceptual training first, and then they were tested for their tonal production. The perceptual training involved several stages: listening to the speech samples, matching the given tones with tone labels, identifying sets of tones spoken by a single speaker, and identifying tones spoken by multiple speakers. If a trainee met the predetermined criterion of at least 80% correct identification in the last stage, together with 75% correct identification for each tone in stages 2 and 3, then the training would be terminated. Feedback was given to indicate if a response was correct or incorrect, and then the trainees were given a production test. The trainees in the other group were given the production training first, and then the perception test. Throughout the training, no
audio sound files for the stimuli were given. The trainees received only visual feedback from a computer screen displaying two panels (upper and lower). For each trial, a pre-recorded model (a pitch contour) was displayed on the upper panel, and the trainees were required to produce a tone with a matching tone contour that was shown in the lower panel. The program assessed their performance. If they passed the proficiency criterion of four consecutive satisfactory productions, or had more than 22 attempts assessed as unsatisfactory, the training for the tone in question would be terminated or abandoned. The results revealed that the perceptual training assisted the trainees in learning of the Mandarin tones better than did the production training.

Wang et al. (1999) adapted the high-variability training procedures used in Logan et al. (1991) and Lively et al. (1993, 1994) to train native English learners of Mandarin to perceive the four Mandarin tones in eight 40-minute sessions over two weeks. During training, feedback was given with every response to indicate the correct answer, followed by the second presentation of a pair of tones. A posttest and generalization tests were administered after the training, and a retention test was given after 6 months. The results showed that the learners’ tonal perceptions were improved, and that they were able to transfer training to generalization tests. The findings suggested that training encouraged non-native suprasegmental perceptual modification in a way similar to segmental training.

In a subsequent study, Wang, Jongman, & Sereno (2003) further examined whether perceptual learning of tonal properties can be transferred to learners’ tonal productions after training (Wang et al., 1999). Learners’ speech
samples before and after training were given to native Mandarin speakers for identification. The results indicated that the perceptual training of lexical tones also improved trainees' tonal productions.

So (2003)\textsuperscript{7} demonstrated that the type of feedback provided during perceptual training could also affect trainees' performance in tonal perception and production. Two groups of 12 naïve listeners (consisting of native speakers of English, Cantonese, Japanese, and Korean) were trained to learn to distinguish the tones in two Mandarin tone pairs, Tone 1-Tone 4, and Tone 2-Tone 3, using a pretest-posttest paradigm. The training session lasted for 60 minutes. According to the literature, these pairs were the most confusing to learners of Mandarin (Bluhme & Burr, 1971-72; Kiriloff, 1969; Miracle, 1989; Shen, 1989). In the experiment, the effectiveness of two training approaches, Audio-Visual (AV) vs. Simple (SIM) feedback, was examined. The participants were assigned to one of the training groups, which differed in the feedback type employed during training (AV or SIM), and each group consisted of native speakers of the four languages. Training was self-paced. Throughout the study, the trainees were given identification tasks using stimuli spoken by two native Mandarin speakers (1M, 1F). Testing stimuli for the pretest and posttest were the three syllables \textit{di, da, du}\textsuperscript{8} produced with the four Mandarin tones, and training stimuli were ten other CV syllables, in which the vowel was any one of the monophthongs /i, a, u/. During training, each trial presented a tone with two

\textsuperscript{7} So (2003) was a pilot experiment for the present study that examined the effects of two training approaches on trainees' performance. However, the L1 prosodic effect was not examined.

\textsuperscript{8} These are in Mandarin Pinyin. In IPA, they are [ti, ta, tu]. These syllables were also used in Gottfried & Suiter (1997).
identification choices, and feedback was given once the trainees had made a response. The trainees in the AV group received feedback consisting of three components: audio sound files, a static pitch graph showing the pitch contours of the two tones, and a written description (or hints) about the acoustic cue differences between the tones in the pair. For example, in terms of durational cue for the Tone 2 - Tone 3 pair, "Tone 3 is longer than Tone 2" was presented. In contrast, trainees in the SIM group received feedback that only indicated "correct" or "incorrect" for their responses. The results showed that although both training approaches improved the trainees' perceptions of Mandarin paired tones, the trainees who received the AV training showed a greater perceptual improvement in distinguishing the paired tones than those who were trained with Simple feedback. In terms of the learners' tonal productions, pitch contours and vowel duration were examined. Results showed that the trainees in the AV group produced tones closer to those produced by native Mandarin speakers than did those in the SIM group. These findings suggest that an effective type of feedback can facilitate the learning process when the duration of training is limited.

Wayland & Guion (2004) conducted a tone training study focusing on the effect of the L1 background (Chinese vs. English) and the interstimulus interval (ISI of 500 vs. 1500 ms) on learners' performance. They found that Chinese listeners outperformed native English speakers in both ISI conditions, suggesting that the trainees' tonal experiences from their native systems may facilitate learning. (The details of the study were presented in Section 1.2.2.2.)
1.4.3 Summary of Previous Perceptual Training Studies

Overall, previous research has demonstrated that, by using appropriate procedures, the long-term modification of learners' perceptions seems to be possible (e.g., Jamieson & Morosan, 1986; Logan et al., 1991). There are several crucial areas related to perceptual learning. First, a high-variability training paradigm should be employed during training, because it helps trainees to form robust categories, and to transfer the knowledge gained from the training to novel stimuli in the future. Second, natural stimuli can be used during training, because natural stimuli contain complete information about acoustic cues for the non-native sounds or phones that trainees are expected to learn (Logan et al., 1991). Third, using identification tasks rather than discrimination tasks seems to be more effective in assisting trainees to “improve categorization of non-native speech contrasts” (Jamieson & Morosan, 1986, p. 207). Identification tasks also help listeners to “develop and use phonetic memory codes in short-term memory rather than rely on information in sensory memory” (Logan et al., 1991, p. 876). Lastly, the inclusion of generalization and retention tests can provide a better assessment of the effectiveness of the training methodology.

However, several issues are worth mentioning. First, the success of these training studies raises some concerns regarding the optimal duration of perceptual training. Some questions, such as how much time is needed for training non-native contrasts, and how to determine when to terminate the training phase, are needed to be considered. The training duration varied from study to study. The training phase in some studies lasted 2-4 weeks with 8-45 sessions in total (e.g., Bradlow et al., 1997; Logan et al., 1991; Wang et al.,
1999), but some other studies trained listeners to successfully perceive L2 sound segments within a relatively short period of time. For example, the training phases in Francis & Nusbaum (2002) and Jamison & Morosan (1986) were about 80-90 minutes. The duration for the training phase(s) in Pisoni et al. (1982), Jamison & Morosan (1989), and So (2003) was less than 60 minutes, and the one used in Flege & Wang (1989; see Section 1.2.2.1) and in Werker & Tees (1984b) was just a single block of stimuli that lasted for a few minutes. A possible explanation for this seems to be related to the fact that studies using longer training phases also included a larger amount of training materials (e.g., Logan et al., 1991). However, it is also uncertain if there were trainees who had been over-learned during the training phases, given the fact that the trainees' learning processes were not uniform and large variations in performance were not uncommon (Pisoni et al., 1982; Leather, 1990; Logan & Pruitt, 1995). Some trainees might not need the complete training sessions, before being able to distinguish the non-native sounds. In contrast, some other trainees might be under-learned, and thus in need of more time to learn the non-native contrasts during training. Presumably, over-learning may occur in a long-term training study (the one with many sessions), while under-learning may occur in a short-term one (see Pisoni et al., 1982). In either case, it is difficult for researchers to know the training effect on trainees' performance, and to determine the optimal length of a training phase. Over-learning may imply that the training duration is not administered effectively, while under-learning may suggest that the training time is not sufficient for the trainees. More importantly, when trainees are under-
learned, the inclusion of these learners in the posttests will increase subject variability in overall trainees' performance, and obscure the result patterns of a perceptual training study, especially the short-term one. Thus, in order to control the training duration and to determine when to terminate the training phase, perceptual training should employ a predetermined criterion to evaluate the trainee's performance during the training phase. For example, both Pisoni et al., (1982) and Leather (1990) set the passing criterion at 80-85% of correct identification scores.

Second, the training effect in studies with a short training phase appeared to be less certain, because these studies did not include generalization and retention tests. It is still uncertain to what extent trainees' perceptions of the non-native contrasts in question have been modified through perceptual training. In contrast, most studies having a longer training phase demonstrated that learners in their studies were still able to perceive novel stimuli and speakers, and to maintain the knowledge months after training. Thus, in order to determine the effectiveness of short-term training, the inclusion of generalization and retention tests seems to be desired.

Third, in order to have a better understanding of using perceptual training to modify non-native speakers' perceptual abilities, studies should be carried out using a cross-language perspective. Because the majority of the above studies used only a single homogeneous participant group, it is unclear if the same training approaches will also be helpful to other participants from different language backgrounds.
Lastly, the procedures for perceptual training may need to include the selection of a feedback type to be employed during training. Most of the above studies focused on employing procedures (e.g., using identification and natural stimuli) that helped learners to identify and establish non-native sound categories, but researchers did not pay much attention to the feedback type they employed. The feedback used in most studies was just to indicate correct-incorrect responses, with or without the addition of a playback of the audio stimuli once to the trainees after each response. It was uncertain if the trainees really knew to what they were supposed to pay attention, based on the feedback provided to them during training. It seems that the learners in these studies were supposed to gradually determine the phonetic or acoustic dimensions of tones through trial and error during the training phases. However, So (2003) has demonstrated that an AV feedback type (audio files, pitch graph, and text) emphasizing acoustic information can facilitate the learning process when the training duration is limited. Thus, employing this kind of (AV) feedback might be helpful when a short period of training is employed.

The present study will address all of the above issues, and employ AV feedback, as well as SIM feedback, during the training phase (details will be presented in a later section). In the next section, the role of feedback in perceptual training will be further discussed, since this area has received minimal attention from researchers in speech perceptual training.
1.5 The Role of Feedback in Training

Feedback plays an important role in perceptual training studies, because it can accelerate the learning process, especially when the training time is limited (Logan & Pruitt, 1995). The purpose of giving feedback in perceptual studies is to indicate to participants whether their responses are correct or not. If the response is correct, they will continue to respond in the same way as before. On the other hand, if it is incorrect, they will know they need to re-adjust their response selection.

1.5.1 Types of Feedback

There are various ways to deliver feedback to participants during perceptual training. The feedback type employed in previous perceptual training studies (e.g., Leather, 1990) simply indicated if the subject’s response was correct or incorrect (hereafter called Simple feedback). Some studies provided not only Simple feedback, but also played back the stimulus sound when an incorrect response was made (e.g., Lively et al., 1991; Wang et al., 1999). There are also other variations of feedback. Bradlow et al. (1997), for example, used a chime to signal a correct response, and a buzzer to indicate an incorrect response. Wayland & Guion (2004) made use of the idea of Simple feedback during the training phase: the correct button would blink for 5 seconds whenever a subject’s response was incorrect.

Besides Simple feedback, another alternative that has been widely used in many previous studies is Audio-Visual (AV) feedback. Typically, AV feedback provides trainees with dual channel inputs (playing an audio sound file and
showing the visual component on the same display screen) to assist with the learning of sound segments, suprasegmentals, intonation, and other aspects of pronunciation (e.g., Anderson-Hsieh, 1992; de Bot, 1983; Molholt, 1988; Murakawa & Lambacher, 1996; Ortega-Llebaria, Faulkner, & Hazan, 2001). The visual component could be in the form of laryngographs, spectrograms, pitch contours, waveforms, intensity graphs, or video clips showing the face of a talker. The most obvious advantage of using AV training is that the additional channel of input helps to raise learners' awareness of the phonological/phonetic/acoustic characteristics (e.g., vowel duration) of the non-native target segments or suprasegmentals. This in turn assists learners in establishing the sound categories or intonation patterns. In fact, it has been suggested that visual displays (e.g., graphic representations or animation) are useful in illustrating abstract concepts, and establishing abstract spatial dimensions during learning (Lim, 2001; Mayer et al., 1996; Rugelj, 2005), because they serve as visual knowledge tools that help learners understand the subject matter, and facilitate recall and retention (Rugelj, 2005).

Hardison (2003) examined the effects of AV versus Audio-only, and Visual-only training on L2 learning. She found that Japanese and Korean learners of English who received AV feedback (showing a talker's face together with the sound files) during training (i) showed improvement in their perceptual accuracy for English /r/ and /l/, and (ii) were able to transfer the gained knowledge to perceive novel stimuli and speakers, and to produce the target segments successfully.
Lambacher (2001) trained Japanese learners of English to improve their pronunciation of English vowels with a kind of AV feedback called electronic visual feedback (EVF). The EVF displayed the acoustic characteristics of a model and the students' imitations (e.g., waveforms, pitch contours, intensity graphs, and spectrograms) in a dual display (top and bottom panels). By using this kind of EVF, the Japanese learners could visualize the differences between the sound features of the teacher's samples and their own productions. In addition, by comparing their own pronunciations with a model, students were able to evaluate their pronunciation errors and monitor their own progress.

For the learning of suprasegmentals, such as intonation and lexical tones, the components of AV feedback from the previous studies (e.g., de Bot, 1983; Sanders, Beckman, & Feng, 2005) typically consisted of an audio sound file with a visual display showing the pitch contour. Learners observed the real-time information of the model patterns, and compared them with their own productions, irrespective of their actual pitch values. By using this kind of AV feedback, previous studies have found that the learning of suprasegmental features seems to be successful (Anderson, 1992, 1994; Chun, 1998). For example, de Bot (1983) compared the effects of two training approaches, AV (i.e., audio & pitch contour) vs. A-only (audio-only), on Dutch students' learning of English intonation. It was found that the students who received the AV feedback during training outperformed those who received the A-only feedback. The findings suggest that using AV feedback is better than A-only in assisting the learning of suprasegmentals, because the visual component (showing pitch
contours) provides an additional channel input for the trainees to visualize the foreign intonation patterns. Similarly, Taniguchi & Abberton (1999) also reported that AV training that showed audio sound files and pitch contours helped Japanese trainees to improve their learning of English intonation. Hardison (2004) employed a real-time computerized pitch display over a period of three weeks while training native English learners to learn French prosody. Results indicated that students not only improved their productions, but also transferred the perceptual training to other French intonation patterns.

AV feedback has also been used as a tool for teaching lexical tones. Bluhme & Burr (1971-1972) employed AV training that provided both the audio sound file and the pitch contour simultaneously to beginner English students learning Mandarin tones. These students were encouraged to produce and then to compare their own productions to those model's productions on the monitor. The authors found that this kind of training helped students to produce the tones more accurately.

In contrast, Sanders, Beckman, & Feng (2005) observed some limitations in the use of AV feedback (audio sound files and a pitch tracker) to assist their Japanese learners to learn Mandarin tones. They found that Japanese learners had difficulty in judging the accuracy of their tonal productions. For example, Japanese learners produced both Tone 2 (mid-rising) and Tone 3 (falling-rising) with a similar dipping contour. In addition, the authors reported that Japanese learners frequently produced tones that did not resemble any one of the four Mandarin tones. Thus, the findings of the above two studies suggest that
providing an audio sound file and a pitch contour simultaneously for the form of AV feedback do not guarantee a facilitative effect on the learning of Mandarin tones.

So (2003) demonstrated that Mandarin tones could be trained successfully in a training approach that employed novel AV feedback. The components of the AV feedback included a combination of *audio files*, the *static pitch contours of a tone pair* that show not only the tone contours, but also the relative spatial relationship between the tones, and a *concise text* that explicitly directed the trainees’ attention to some crucial acoustic cues for the tones in each pair (see Section 1.4.2 for details). The combination of these components provided tri-channel input to assist learners in distinguishing the lexical tones in each pair in a short-term training period, by directly drawing their attention to the crucial acoustic information of the tones (e.g., showing tone contours and spatial relationship and durational differences of the tones while listening to the audio files). However, a disadvantage for this kind of AV feedback is that the sound files cannot be played back simultaneously with the pitch contours, which in turn may increase the time spent in training phases, because the trainees may need more time to understand the context during training.

1.5.2 Concerns about Using AV Feedback in Training

Previous studies have shown that training with AV feedback helps language learning. However, there are still some concerns about using it during perceptual training.
First, AV feedback relies heavily on the learners' own judgments in determining the differences between their own productions and the models (see Chun, 1988). The question of how well learners interpret the displays is still an unknown issue. For example, reading spectrograms or waveforms could be difficult for learners who do not have a sufficient background in phonetics or in linguistics. Therefore, it may not be very helpful in assisting learners to realize the fine phonetic differences between the segments.

Second, the combination of an audio file and facial information seems to be helpful for the learning of segments, because this kind of AV feedback shows the configuration of the speech organs (e.g., the tongue position, the shape of the lips, etc.) that are required for the production of each segment. However, this kind of feedback is probably less applicable to the learning of suprasegmentals (e.g., lexical tones). Previous studies have consistently reported that listeners of tone languages (including Japanese, Cantonese, and Thai) do not rely much on visual information from lips and facial movements. Instead, they depend more on auditory-only or auditory together with visual (i.e., AV) information (Burnham, Ciocca, Lauw, Lau, & Stokes, 2000; Burnham & Lau, 1998; Burnham, Lau, Tam & Schoknecht, 2001; Sekiyama, 1996; Sekiyama & Tohkura, 1993). The findings suggest that a visual component showing facial information alone is not a sufficient cue for lexical tones for native listeners of tone languages. Therefore, using video clips that present both audio sound files and visual-facial information may not be helpful in assisting trainees to learn lexical tones better and faster.
Third, AV feedback that consists of audio files and visual pitch contour components may also cause some problems for learners when learning intonation and/or lexical tones. As pointed out by Chun (1988) and Weltens & de Bot (1984), the pitch contour of intonation could be disrupted by voiceless segments. Since voiceless segments are not associated with pitch values, the displayed target pitch contour may appear as a series of broken pitch contours. These broken pitch contours will misrepresent the intonation patterns to learners.

For lexical tones, although the visual component of the AV feedback (i.e., pitch contour) shows the overall pitch contours of individual lexical tones (e.g., a rising pattern or a falling pattern), learners may still have difficulties in judging their own productions (see Sanders et al., 2005).

Taken together, it seems that studies that involve training for lexical tones should employ AV feedback that presents sound signals and pitch contours rather than video clips showing facial information. In addition, to help learners to discern the differences between target tones, including a new component (e.g., text that describes some hints for differentiating the tones) in the AV feedback may assist their learning of lexical tones. In fact, the findings of So (2003) indicated that the effect of AV feedback on trainees' performance could be reinforced by providing explicit acoustic information to shift their attention to the acoustic cues of tones in order to assist with their (perceptual) judgments of the target tones.
1.6 The Four Languages

As mentioned earlier, Mandarin is the target language in this study, and Cantonese, Japanese and English are the native languages of the naïve listeners. The following section (Section 1.6.1) will provide basic definition for tone languages. Then, the individual prosodic system of the languages involved in the present study will be described (Section 1.6.2).

1.6.1 Tone Languages

The majority of the world's languages are tone languages (Gandour, 1994; Yip, 2002). According to Pike (1948, p. 3), a tone language is "a language having lexically significant, contrastive, but relative pitch on each syllable". By using different lexical tones (i.e., pitch patterns), words from the same root-word syllable (i.e., the same sequence of segments) will have different lexical meanings, and therefore tones are phonologically distinct (Fox, 2000; Ladefoged, 1993; McQueen and Cutler, 1997; Yip, 2002).

Lexical tone is a suprasegmental feature that is superimposed on segments (loup & Tansomboon, 1987; Lehiste, 1996). For example, in Cantonese, when the syllable (or root-word) [jì] is produced with a high level tone (Tone 1), it means "doctor". The syllable means "child" if it is produced with a low falling tone (Tone 4). However, it should be mentioned that tone is different from intonation. Tones have phonemic status and function distinctively at word level, while intonation is freely applied at the sentential level or on longer stretches of speech (loup & Tansomboon, 1987; Lehiste, 1970, 1996).
There are two types of tone languages, lexical tone and accentual languages. Lexical tone languages are known for using pitch variations to convey different lexical meanings for the same sequence of segments that form a word. Lexical tone languages can be found in most East and South-East Asian languages, such as Thai, Vietnamese, and Chinese (e.g., Mandarin, and Cantonese). Accentual languages, a subtype of tone language, do have lexical tones in their phonological systems. However, the number of contrasting tones is limited to one or two, and the distribution of tones within words is restricted to specific syllables (Yip, 2002, p. 4). Typical examples are Japanese and Serbo-Croatian. In addition, there are some languages that do not use pitch to distinguish word meanings. They are called non-tone languages, such as French and English.

1.6.2 Prosodic Systems

1.6.2.1 Mandarin

Mandarin is a typical "lexical tone language" (Yip, 2002, p. 2) that has four lexical tones in its system: Tone 1 (high level), Tone 2 (mid rising), Tone 3 (falling rising), and Tone 4 (high falling). In Chinese phonology (e.g., Bauer & Benedict, 1997; Duanmu, 2004; Hashimoto, 1972; Yip, 2002), tone contours are typically described in terms of Chao's (1930) tone letters (i.e., digits) -- the system provides "simplified time-pitch graphs of the voice" (1930, p.24). Tones are represented by 2 or 3 letter values (ranged from 1 to 5 corresponding to the low and high pitch levels of a speaker's pitch range) to indicate the initial and the final, or the initial, the medial and the final position of the tones. The four
Mandarin tones are described as /55/, /35/, /214/, and /51/, respectively. For example, /ta55/ 'carry', /ta35/ 'reach', /ta214/ 'hit', and /ta51/ 'large' (see Figure 1-1 for the tone patterns).

![Tone Patterns](image)

**Figure 1-1.** Mandarin tonal patterns on the syllable /ta/ spoken by a female native Mandarin speaker (top) and a male native Mandarin speaker (bottom).³

Lexical tones, similar to vowels, are perceived via *normalization*, an auditory skill that native listeners use to extract invariant tonal patterns and to ignore a speaker's absolute physical pitch values (i.e., Fundamental frequency, $F0$) (Chan, 1984; Fourcin, 1972; Fox & Qi, 1990; Leather, 1983; Moore &

³ The speakers were two of the speakers for the present study. Speech samples were measured with *Praat* (version 4.4, available from Paul Boersma and David Weenink, http://www.fon.hum.uva.nl/praat/)
As long as the perceived tonal patterns conform to those in their mental representations, native listeners will be able to perceive the tonal patterns and to identify the tones (Ching, 1984, p.323).

Among the perceptual cues for Mandarin tones, F0 cue is the primary perceptual cue for the lexical tones\(^\text{10}\) (Gandour, 1994; Gåording et al., 1986; Halle et al., 2004; Yip, 2002). In particular, both F0 height and F0 contour are the two essential dimensions of the F0 cue for the identification of tones by native Mandarin speakers (Gandour, 1984; Massaro, et al., 1985; Moore and Jongman 1997). In addition, the F0 turning point has been identified to be an important cue for distinguishing between Tone 2 and Tone 3 (Shen & Lin, 1991; Moore & Jongman, 1997). Besides F0 cues, durational and amplitude cues may exist as concomitants, but are not crucial to the perception of Mandarin tones (Gandour, 1994). Previous studies (Ho, 1976; Howie, 1976; Liu & Samuel, 2004) have suggested that Mandarin tones appear to have some intrinsic durational differences: Tone 3 tends to be the longest; and Tone 4 is the shortest. Tone 2 is generally shorter than Tone 3, but longer than Tone 1. However, Tseng (1990) found that duration was not a crucial acoustic parameter of Mandarin tonal production. Consistent with Tseng’s findings, Moore & Jongman (1997) reported that duration alone was not sufficient to distinguish Tone 2 and Tone 3 in Mandarin. It became an important cue when combined with another cue (change in the F0 direction) from onset of the tones. In terms of amplitude cue, Whalen and Xu (1992) indicated that the amplitude contour also plays a role in tonal

\(^{10}\) In fact, this is also the primary perceptual cue for the lexical tones of other tone languages, such as Cantonese (Ching, 1984; Vance, 1977) and Thai (Abramson, 1975, 1978).
identification, but this cue alone is not sufficient for distinguishing all the four Mandarin tones. For example, native Mandarin listeners in the study were able to identify Tones 2, 3, and 4 but not Tone 1, when the tones were on the ba syllable. It should also be mentioned that Klatt (1973) found that the F0 contours of synthetic Mandarin tones alone with constant duration and amplitude also resulted in remarkable identification of tones (90% of correct) by native speakers, when the F0 range \((F_{0\text{max}} - F_{0\text{min}})\) was reduced to 4 Hz. Thus, taken the findings of the previous studies into consideration, F0 patterns (including F0 contours and F0 height) are primarily used to differentiate phonemic tones (see Gandour, 1994, 2006; Jongman, Yue, Moore, & Sereno, 2006; Leather, 1987). Tonal perception by non-native listeners was substantially different from native Mandarin speakers. Gandour (1983, 1984) identified that pitch height and pitch direction are the two important dimensions of tone perception for Chinese listeners, and that pitch height is more important for English listeners. Since English has no lexical tones, English listeners directed their attention to “the level of fundamental frequency characteristics” of the stimuli (Gandour, 1983, p. 171). Lee, Vakoch, & Wurm (1996) found that Mandarin listeners perceived the four Mandarin tones better than did Cantonese listeners who performed better than did the English listeners. The results imply that L1 tonal experience exerts a facilitative effect on non-native listeners' tonal perceptions. In contrast, Huang (2001) demonstrated that native Mandarin listeners' perceptions of the Tone 2 and Tone 3 were affected by Mandarin tone sandhi rule: Tone 3 becomes Tone 2 when followed by another Tone 3 (Chao, 1968). However, native English
listeners in her study were not affected by this phonological influence.

Some studies (Bluhme & Burr, 1971-1972; Kirilloff, 1969; Miracle, 1989; Shen, 1989) have also reported that the similarities in the contours of some tones increase production and perceptual difficulties for non-native learners. For instance, the pitch contours for Tone 1 and Tone 4 start at a high level pitch. Tone 2 and Tone 3 both have a rising portion of the pitch contours, but they also exhibit different degrees of falling-rising patterns. According to Fon & Chiang (1999), Taiwanese Mandarin rising tone (i.e., Tone 2) shows a dipping at the initial portion and then a moderately rising pattern in the final portion. The dipping pattern can also be seen in the productions of Mandarin speakers in some previous studies (e.g., Gandour, 1983; Howie, 1976; Moore & Jongman, 1997). In fact, some native Mandarin speakers from Beijing in this study also produced the rising tone (Tone 2) with a "falling and rising" pattern.

1.6.2.2 Cantonese

Cantonese is another typical lexical tone language. Its tonal system consists of six phonemic tones (see Figure 1-2): three level tones (high: Tone 1 [55], mid: Tone 3 [33], and low: Tone 6 [22]), two rising tones (high rising: Tone 2 [35] and low rising: Tone 5 [23]), and one low falling tone (Tone 4 [21]). For example, /si55/ 'poem', /si35/ 'history', /si33/ 'attempt', /si21/ 'time', /si23/ 'city', /si22/ 'trained person'. At the phonetic level, Cantonese Tone 1 [55] has two allotones, a high falling tone [53] and a high level tone [55] (Bauer & Benedict 1997; Hashimoto 1972; Yip, 2002). The high falling tone is also well documented in a tone sandhi environment (Hashimoto, 1972; Bauer & Benedict, 1997). A high
falling [53] becomes a high level [55] before a high falling or a high level. However, unlike Guangzhou Cantonese speakers, most Hong Kong Cantonese speakers “have lost the high falling tone, or use it in certain syntactic environments, or use it in free variation with high level” (Bauer & Benedict, 1997, p.167).

Figure 1-2. Cantonese (HK) tonal patterns on the syllable /si/ spoken by a female native Cantonese speaker (top) and a male native Cantonese speaker (bottom).\textsuperscript{11}

\textsuperscript{11} The speakers were two of the participants in the present study. Speech samples were measured with \textit{Praat}.
When comparing the tonal systems between Cantonese and Mandarin, some similarities and differences can be identified. With respect to their similarities, Cantonese Tone 1 and Tone 2 are similar to those of Mandarin Tone 1 and Tone 2, and their pitch contours are compatible to one another. Both Cantonese Tone 1 and Mandarin Tone 1 are high level tones, and have the same tone letter value \([55]\) in the literature (e.g., Mandarin: Duanmu, 2000; Howie, 1976; and Cantonese: Bauer & Benedict, 1997; Hashimoto, 1972; Yip, 2002). Similarly, Cantonese Tone 2 and Mandarin Tone 2 also resemble each other. Although Cantonese Tone 2 is labelled as a high rising tone, and Mandarin Tone 2 is labelled as a mid rising tone, both tones have the same tone letter value \([35]\) in previous studies (Mandarin: Duanmu, 2000; Howie, 1976; and Cantonese: Bauer & Benedict, 1997; Hashimoto, 1972; Yip, 2002). Further, the pitch contours of Cantonese rising tones (high and low) are typically produced with a dipping at the initial portion and then a rising pattern at the final portion (Bauer & Benedict, 1997; So, 1999; also see Figure 1-2). This kind of dipping pattern is also found in the pitch contour of Mandarin Tone 2 (Fon & Chiang, 1999; Howie, 1976; also see Figure 1-1).

In term of differences, Cantonese tonal system does not have a counterpart for Mandarin Tone 3. However, as noted earlier, Cantonese rising tones (Tone 2 and Tone 5) do exhibit a dipping portion in their pitch patterns (see Figure 1-2) that is similar to Mandarin Tone 3. In addition, Cantonese does not have a high falling tone at the phonemic level, but Cantonese Tone1 does have a high falling allotone \([53]\). This is similar to Mandarin Tone 4 \([51]\).
1.6.2.3 Japanese

Japanese, a pitch-accent language, is a “subtype of tone language” with a few contrasting tones in its prosodic system (Yip 2002, p.4). In standard Japanese, pitch accent patterns are predictable\(^\text{12}\) (McCawley, 1978; Tsujimura, 1996). Native speakers of Japanese use a different accentuation (pitch) to differentiate lexical items (Haraguchi, 1999; Ito, Speer, & Beckman, 2003; McCawley, 1978; Tsujimura, 1996). For example, the Japanese disyllable, ame, illustrates a minimal pair that is dependent upon different pitch accent patterns (see Figure 1-3 for their F0 contours). The word means “candy” when it is accented on the second syllable (i.e., amé, LH), but it means “rain” when it is accented on the first syllable (i.e., âme, HL) (Tsujimura, 1996, p. 76).

Although the basic prosodic units for Japanese pitch-accents and Mandarin tones are different (mora vs. syllable or rhyme; see McCawley, 1978; Chao, 1968), the two Japanese pitch-accents (HL and LH) are similar to Chinese lexical tones in a number of ways. First, both Chinese lexical tones and Japanese pitch accents have phonemic status in their own languages, because they are contrastive, and used with lexical /grammatical functions at the word level. Second, in terms of phonetic realizations, the F0 patterns for Japanese LH and HL accents are similar to those for Mandarin Tone 2 and Tone 4 (see Figures 1-1 and 1-3). In fact, some phonologists (Duanmu, 2004; also see Woo, 1969; Yip, 1980) describe Chinese contour tones as a sequence of level tones. For example, rising and falling tones are described as LH and HL, respectively.

\(^\text{12}\) All moras that precede the accent will be assigned with high tones, and moras follow that accent will be assigned with low tones. The first mora of the word is assigned with a low tone when it is unaccented, but it will receive a high tone when it is accented.
Lastly, in terms of syllable weight, both Japanese (C)V, (C)VC, and (C)VCV sequences (e.g., *ame*) and Mandarin CV(C) syllables are represented with two moras (Cutler & Otake, 1999; Duanmu, 1993, 2005; McCawley, 1978; Sugito, 2003; Yip, 2002).

![Pitch-accent patterns](image)

*Figure 1-3. Japanese pitch-accent patterns on the disyllable *ame* spoken by a female native Japanese speaker (top) and a male native Japanese speaker (bottom).*  

13 The speakers were two of the participants in the present study. Speech samples were measured with Praat. Japanese pitch-accent patterns shown in the figure conform to those in Sugito (2003).
1.6.2.4 English

English is a non-tone language because it does not have a lexical tone system. Beckman (1986) distinguished English accents from Japanese accents in terms of their phonetic manifestations. Beckman identified English as a “stress-accent language” and Japanese as a “non-stress-accent language” (i.e., a pitch-accent language). English accents involve a combination of a number of phonetic features -- pitch, duration, intensity and others (e.g., vowel quality), while Japanese accents are realized mainly in pitch. In English, stressed syllables “recur at regular intervals of time, regardless of the number of intervening unstressed syllables” (Crystal, 1997, p. 365; also Fox, 2000).

The use of pitch at the word level is very restricted. Even for lexical stress, for example, pitch is just one of the acoustic components (other than loudness and duration) to indicate stress in English homophous pairs, such as SUBJECT (noun) and subJECT (verb) (Fear, Cutler, & Butterfield, 1995; Pennington & Ellis, 2004). However, these words are typically produced with vowel quality differences (Beckman, 1986, Cutler & Otake, 1999; Fox, 2000). Figure 1-4 illustrates the waveforms and the pitch patterns for the noun and verb forms of the word subject spoken by two English listeners (1F and 1M) who participated in the present study. Both of them did exhibit different degree of vowel reductions for the unstressed syllables, and the pitch patterns apparently were produced with different degree of variations.
Figure 1-4. English lexical stress patterns on the word subject spoken by a female native English speaker (top) and a male native English speaker (bottom). 

English uses prosody (pitch) primarily at the sentential or phrasal level (i.e., intonation). English declarative sentences are mostly uttered with a falling pitch pattern, while questions are typically produced with a rising pitch pattern. However, the same sequence of words could be used with a rising or a falling pattern and convey different meanings depending on the context. For example, Ladd (1996, p. 9) uses the phrase five pounds to illustrate two possible pitch patterns. In answering to a question: How much does it cost? One may simply reply to the question with the words five pounds with a falling tune. However, the

---

14 The native English speakers were two of the participants in the present study. Speech samples were measured with Praat.
words, *five pounds*, could be produced with a rising tune to ask for confirmation that the speaker has just heard: (Did you say) *five pounds*?

1.7 The Present Study

1.7.1 Research Questions

As mentioned in Section 1.1, the present study is motivated by the problems non-native learners experience when learning Mandarin lexical tones. Although there is a considerable amount of literature that examines the effects of L1 phonological systems and perceptual training on L2 learning (see Sections 1.2 & 1.4), little attention has been paid to the effect of an L1 tonal system on trainees' performance while learning non-native tones. To date, there is a paucity of training research in lexical tones that investigates trainees' difficulties in perceiving and/or producing L2 tonemes that are related to their existing L1 prosodic systems. Moreover, previous studies (e.g., Logan et al., 1991; Jamieson & Morosan, 1986, 1989) have emphasized that using appropriate procedures can shift a learner's attention to specific acoustic cues in order to modify the perception of non-native contrasts. However, their procedures do not include the effect of the feedback type given during training. Other previous studies have shown that training approaches using more channels of input (e.g., AV feedback) can facilitate the L2 learning better than the one using single channel of input (A-only and/or V-only) (see De Bot, 1983; Hardison, 2003). One may expect that using more channels of input may also speed up the learning process. Given that So (2003) has demonstrated that using AV feedback (audio sound files, pitch contours, and text) can improve the listeners' performance of L2 tonal perception
and production immediately after training, it might be fruitful to further investigate the effectiveness of employing this kind of AV feedback during training on the long-term modification of learners' L2 tonal perception.

Therefore, the aim of the present study is two-fold. By training three groups of naïve listeners (Cantonese, Japanese, and English) to learn the tonal system of Mandarin, this study focuses on both theoretical and methodological issues related to the perceptual learning of lexical tones. Theoretically, it examines the effects of naïve listeners' L1 prosodic backgrounds on their learning of Mandarin tones. It aims to determine to what extent an L1 system may affect the learning of a new tonal system. The findings will be discussed within the framework of the Perceptual Assimilation Model (PAM).

Methodologically, this study utilizes a novel audio-visual (AV) training approach\textsuperscript{15} that implements the acoustic cues of lexical tones to draw naïve learners' attention to the acoustic difference (e.g., tone contour) of Mandarin tones during training. Its effectiveness in modifying trainees' perception of L2 tones will be evaluated.

The reason for selecting the three languages was that native speakers of these languages use tonality differently: Cantonese (a tone language), Japanese (a pitch-accent language) and English (a stress-accent language). Cantonese and Japanese are tone languages but they differ in terms of (i) the number of tones in their native systems (6 vs. 2) and (ii) linguistic experience of using tones (or pitch patterns) at the word level. Each Cantonese syllable (i.e., a word)

\textsuperscript{15} The AV feedback for the present study is modified from the one used in So (2003). It has three components -- audio sound files, animated pitch contours, and text concisely and explicitly indicating perceptual cues. Details will be presented in Chapter 2.
carries a tone, while each Japanese syllable does not necessarily bear a tone. In fact, each Japanese word (could be more than one syllable) carries tone(s). Cantonese listeners have more experience in using pitch (patterns) than do Japanese listeners. In contrast, English is a non-tone language that does not use pitch (patterns) alone to distinguish lexical items at the word level. The linguistic experience of using pitch is thus considered to be the least. Thus, Japanese will be intermediate between the two groups. Moreover, the inclusion of Cantonese listeners allows the investigation of the effects of specific characteristics of Cantonese tones (i.e., their phonemic status and tonal contours) on their learning performance.

The present study attempts to address a number of questions and issues. With respect to the L1 prosodic effect, will the prosodic background of naïve listeners (Cantonese, Japanese, and English) affect the learning of a new tonal system? If so, to what extent does their prosodic background affect the learning process? Will the linguistic experience of using tones always facilitate the learning of L2 tone as suggested by Wayland and Guion (2004)? If so, will the learning performance of trainees from different backgrounds be different as a function of the use of tonality in their native languages? If not, what aspects (characteristics/features) of their prosodic systems constitute the learning differences? More importantly, how these aspects affect their learning of L2 tone categories at the initial stage?

With respect to the AV training approach, will the use of this novel audio-visual approach, which provides tri-channel input for learners as learning intake,
successfully draw trainees' attention to the acoustic differences of the Mandarin tones in a short-term training study? Will the tri-channel feedback always lead to better performance than the single channel feedback? In other words, will the use of this AV training approach during perceptual training be more effective and faster than the use of the Simple (SIM) training approach (feedback indicating correct-incorrect) in assisting listeners to learn a new tonal system? If so, do they provide a reliable modification of listeners' L2 tone perceptions?

Finally, since this study examines the effects of two factors, L1 prosodic backgrounds and an AV training approach, it is not unreasonable to assume that there might be some kind of interaction effect between the two factors. Specifically, this study attempts to answer the question of which language group will benefit most from the AV training approach.

1.7.2 The Outline of the Present Study

The present study is a short-term longitudinal study that employs a pretest-posttest paradigm to evaluate the training effect. There are four phases (Phases 1-4) conducted on four separate days: Pretest, Training, Posttest-Day1 (PTDay1), and Posttest-Day2 (PTDay2). The intervals between the four phases are carefully controlled in order to minimize the practice effect carrying over from the previous experimental phase. The intervals between the first three phases (Pretest - Training and Training - PTDay1) are set to 1-2 days.\footnote{The interval for a few participants was 3 days, due to scheduling conflicts.} The interval
between the PTDay1 and the PTDay2 phases is set to a month (Tees & Werker, 1984). The purpose for each phase is briefly summarized below.¹⁷

The identification test given in the Pretest phase is employed to evaluate the listeners' abilities to identify the stimuli of Mandarin tones, and to serve as a baseline for the comparison of performance in the two posttest phases: PTDay1 and PTDay2:

The purpose of the Training phase is to train listeners to perceive 288 CV words, presented in multiple blocks and sessions as six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, and T3-T4).¹⁸

In the last two phases (PTDay1 and PTDay2), four identification tests are employed. The tests on both days are identical and are used to assess the trainees' performance before and after training at different times. The results of PTDay1 will indicate the short-term or immediate effect of training, while those of PTDay2 will provide evidence for the training effect on the long-term retention of Mandarin tones.

In terms of the four identification tests used at each posttest time, PTDay1 and PTDay2, each phase consists of a posttest that is identical to the Pretest, and three generalization tests (Gen1, Gen2, and Gen3). These generalization tests are used to test how well listeners can transfer the knowledge gained from the training to novel stimuli (Gen1), novel speakers (Gen2), and nonspeech or filtered speech stimuli (Gen3).

¹⁷ Detailed descriptions will be provided in Chapter 2.
¹⁸ Throughout this thesis, the capitalized letter "T" and a number refer to a specific tone in Mandarin. For example, T1 refers to Tone 1, and T4 refers to Tone 4.
Before discussing the hypotheses for the present study, it is worth mentioning that this study indeed differs from the previous perceptual training studies in several respects.

1. Most previous studies (e.g., Leather, 1990, Logan et al., 1991; Jamieson & Morosan, 1986, 1989) employed homogenous participant groups. However, this study will employ trainees from three different language backgrounds, and examine their prosodic background differences in relation to their learning performance.

2. Previous studies typically employed L2 learners of the target language (e.g., Logan et al, 1991; Wang et al, 1999), implying that the trainees had prior knowledge of the target language before training. Thus, their performance results after training may be confounded by their existing L2 experience. To eliminate the effect of L2 experience on trainees' performance, the current study employs only naïve listeners as participants.

3. In order to examine the L1 prosodic effect, this study also ensures that naïve listeners do not have any formal musical training or Mandarin lessons before their participation in the experiments. Studies have clearly shown that listeners who had musical backgrounds outperformed those who had no musical experience when perceiving lexical tones (e.g., Burnham & Brooker, 2002; Burnham, Francis, Webster, Luksaneeyanawin, Atta-paiboon, Lacerda, & Keller, 1996; Gottfried & Riester, 2000; Gottfried, Staby, & Ziemer, 2004).
4. Wayland & Guion (2004) attempted to examine the L1 background effect in relation to tone learning by training learners to perceive a specific pair of Thai tones. The present study extends this line of research by examining a complete tonal system in order to have a better understanding of the influence of an L1 prosodic system on L2 tone learning.

5. Previous studies typically did not control training duration and learner variability when learning L2 sound contrasts. This study will follow the procedure used by Pisoni et al. (1982) and Leather (1990), and employ a predetermined criterion (i) to control the training duration and minimize the situations of over- and under-learning, and (ii) to reduce learner (subject) variability and increase response consistency in perceptual categorization during the learning of L2 tones.

6. In terms of generalization tests, the present study examines trainees' abilities to transfer the training knowledge not only to novel stimuli and speakers (e.g., Logan et al., 1991), but also to nonspeech stimuli - filtered speech (as in Burnham et al., 1996; Ching, 1984; Stevens & Keller, 2001), in order to determine how well the trainees can perceive the Mandarin tonal patterns. For filtered stimuli, the segmental information of the stimuli is removed and only the tonal information is left. Lee et al. (1996) reported that native listeners of tone languages (Mandarin and Cantonese) benefited from lexical information or word advantage (the experience of similar word of their native languages) in tonal perception tasks using both Mandarin and Cantonese stimuli. Thus, the inclusion of filtered speech
stimuli in Gen3 will remove the effect of lexical information on the performance of the three listener groups.

7. Finally, previous tone training studies focused on the learners' improvement after training, and their abilities to make generalizations and/or the issue of long-term retention of the target tones. None has attempted to relate trainees' performance to current perception models. The present study will employ Best's PAM (1995) as a theoretical framework to account for the trainees' performance in relation to their L1 prosodic backgrounds before and after training.

1.7.3 Hypotheses

In the present study, three hypotheses will be tested. The first and the second hypotheses focus on the effects of two factors – L1 prosodic background and training approach. The last one is about the interaction of the two factors. They will be described in the following.

1.7.3.1 H1: The Effect of L1 Prosodic Background

The first hypothesis is about the effect of L1 prosodic background on trainees' performance. Given the differences in the prosodic systems of the three languages (see Section 1.6), it is hypothesized that listeners' L1 prosodic (phonology) systems affect the L2 learning. However, the main influence is not due to the amount of experience in tonality use in their native languages, but the interference of some lexical tones (in terms of their phonemic status, and F0 contours) in their prosodic systems during learning.
Specifically, it is hypothesized that owing to the influence of their native tonal system, Cantonese listeners will have greater difficulty in distinguishing several pairs of Mandarin tones, Tone 1-Tone 4 and Tone 2-Tone 3. For the former tone pair, Cantonese speakers may perceive them as one single toneme (Mandarin Tone 1), since the high falling tone is an allotone of the high level tone (Tone 1) in Cantonese. For the latter pair, Cantonese listeners may tend to perceive the tones as Mandarin Tone 2, because Cantonese Tone 2 [35] is similar to Mandarin Tone 2 [35], and Mandarin Tone 2 and Tone 3 share a high degree of similarity between their pitch contours. In contrast, both Japanese and English listeners will not experience such interference caused by their native prosodic systems. Further, because Japanese has the rising and the falling pitch-accent patterns that are similar to Mandarin Tone 2 and Tone 4, the Japanese listeners might find it easier to master the Mandarin tones than the English listeners.

If the results indicate that Cantonese listeners perform poorer than do the other group listeners, the result pattern will imply that learners’ phonological systems play an important role in the learning of Mandarin at the initial stage. However, if the results indicate that Cantonese listeners perform better than do the other two group of listeners due to their general linguistic experience of using tones, the result pattern will imply that learners’ phonological systems do not affect the learning of Mandarin at the initial stage, but that their linguistic experience in general plays a more important role in the learning.
The results will reveal if listener groups' performance differences can simply be attributable to linguistic experience in tonality use in general (as suggested by previous studies) or to their prosodic systems (e.g., the influences of the phonemic status of tones, and the role of contrastive pitch pattern use (for lexical items) by languages with different phonological systems. It may indicate (i) how listeners' native prosodic systems can affect their initial learning of a new tonal system (Mandarin, in this study), and (ii) whether their tonal systems will facilitate, hinder, or exert no effect on learning of L2 tones.

1.7.3.2 H2: The Effect of Training Approaches

The second hypothesis relates to about the training approaches. It is hypothesized that feedback consisting of multiple components (audio, animated graph, and text) would be more effective and helpful than feedback consisting of a single component in enhancing input (which becomes learners' intake of Mandarin tones), and in drawing learners' attention to the acoustic cues (e.g., tone contour) of Mandarin tones. In addition, it is expected that the feedback providing tri-channel input will be an ideal training procedure for a short-term training environment, and will promote long-term modification of tonal properties. It is because the combination of the three components helps illustrate abstract concepts and establish abstract spatial dimensions (Lim, 2001; Mayer et al., 1996; Rugelj, 2005), which in turn helps learners understand the subject matter, and facilitate recall and retention (Rugelj, 2005).

In particular, in this study, it is hypothesized that the (tri-channel) AV feedback will be more effective than will the (single-channel) SIM feedback in
assisting trainees to learn the Mandarin tones, because the AV feedback provides three components\textsuperscript{19} to shift their attention to the perceptual properties of lexical tones. Further, it is expected that the AV feedback will provide better long-term modifications to the trainees’ non-native tonal perception than will the SIM feedback. If this is true, then the trainees who are trained with the AV feedback should outperform those who are trained with the SIM feedback in all the test conditions in the posttest phases (PTDay1 and PTDay2). They should also spend less time on training than those in the SIM groups. These results will imply that the AV feedback is more useful and desirable for helping trainees to establish L2 categories (Mandarin tones in this study) during short-term training. Thus, using feedback that provides with multi-channel input should be considered as a useful procedure (similar to those employed in previous studies) that (i) shifts learners’ attention to the perceptual properties of tones and (ii) modifies their perceptual knowledge of the L2 tones. Consequently, it helps trainees to establish the new L2 categories.

However, if the results indicate that the performance of the trainees in both the AV and the SIM groups is compatible or (in an unlikely case) the trainees in the SIM groups outperform those in the AV groups, this implies that alternating feedback type (by adding more channels of input) during lexical tone training is not necessary. Thus, optimal training procedures for perceptual training do not need to include this AV approach.

\textsuperscript{19} See Section 2.2.2 for details.
1.7.3.3 H3: Cantonese Listeners and AV Training

Finally, since this study examines both the effects of the L1 prosodic systems and of the AV training approach, it is possible that AV training will be more helpful to a specific language group than to the other groups. It is expected that the group with the poorest performance after training will benefit the most from the AV training approach. In other words, based on the predictions in the first hypothesis, if the posttest results indicate that the Cantonese group exhibits the poorest performance among the three language groups, then the Cantonese listeners might benefit the most from the AV training approach, even though their tonal system may increase difficulties in learning Mandarin tones. If the opposite result pattern (i.e., ENG is the poorest) is found, the English group should benefit the most from the AV training, even though the English trainees have limited experience in using pitch at the word level.
CHAPTER 2: THE EXPERIMENT

This chapter describes the experimental methodology for the present study. It provides information about the participants (Section 2.1), the materials including the stimuli and the feedback types employed for the identification tests and the training sessions (Section 2.2), the procedure (Section 2.3), and the domains of analyses (Section 2.4).

2.1 Participants

Thirty adults\textsuperscript{20} were paid to participate in this 4-day short-term longitudinal experiment. They were recruited based on their native languages: Hong Kong Cantonese (CAN), Japanese (JPN), and Canadian English (ENG). The Cantonese listeners \((n=10)\) ranged in age from 18–26 years \((M=21.7\text{ years})\), and the Japanese listeners \((n=10)\) ranged in age from 18-36 years \((M=23.8\text{ years})\). All native speakers of Cantonese and of Japanese were born and raised in their home countries, and came to Canada after the age of 15 years. The Canadian English listeners \((n=10)\) ranged in age from 18–35 years \((M=21.7\text{ years})\), and were all born and raised in Canada. All of the participants were undergraduate students either at Simon Fraser University or at colleges in Vancouver, British Columbia. None of them was a linguistics major, and none had taken more than

\textsuperscript{20} The limited number of participants was due to the difficulty in recruiting naive participants who met the stringent criteria: no prior knowledge of Mandarin and no musical training experience. Originally, around 300 potential participants were recruited but only the 30 participants met the above criteria.
three lower level courses in linguistics or in phonetics. They all passed a pure-tone hearing screening (250, 500, 1000, 2000, 4000, and 8000 Hz at 25 dB HL) prior to the experiment.

The selection of all the participants was based on two criteria: they had neither learned the target language (Mandarin) nor received formal musical training prior to or during the time of the study. The former criterion ensured that experience with the target L2 tones was not present to affect the trainees' performance. The latter criterion was also crucial, because previous studies have shown that listeners with musical training outperformed those without in both production and perception tasks (e.g., Alexander, Wong, & Bradlow, 2005; Burnham & Brooker, 2002; Gottfried & Riester, 2000).

According to the participants’ language backgrounds (CAN, ENG, and JPN) and the training approach (AV and SIM) that was administered during the training phase, these 30 participants were randomly assigned to six experimental groups (n=5 per group): CAN(AV), ENG(AV), JPN(AV), CAN(SIM), ENG(SIM), and JPN(SIM). Listeners in the AV groups received Audio-Visual feedback while those in the SIM groups received Simple feedback during training (these training approaches will be discussed in later sections).

2.2 Materials

The stimuli used in the present study were produced by four native Mandarin speakers (2 females and 2 males). These speakers, aged from 22–25 years (M=23.5 years), were born and raised in Beijing, China.
Recordings of five instances each of 18 Mandarin consonant-vowel (CV) syllables,\(^{21}\) comprising a consonant followed by one of the vowels /i/, /u/ or /a/ and using each of the four Mandarin tones, were recorded from each of the speakers. The target word was placed in the final position of a carrier sentence in Chinese [我 说 X ("I say X")]. Individual recordings were made in a sound-treated room in the Phonetics Laboratory at Simon Fraser University using a high quality microphone (Sennheiser MD46) connected to a CD-Recorder (Marantz CDR300). All digital files from the CDs (44.1 kHz) were extracted using a speech editing program (Goldwave, v.5.07), and saved as audio sound files in WAV format on a PC laptop computer (Dell Inspiron 600m). All target words were excised from the sentence frame, and normalized to peak intensity. They were used as stimuli for the Pretest, the Training, and the two posttest (PTDay1 and PTDay2) phases.

A total of 720 tokens used in the experiment were judged by two additional native Mandarin speakers (22 and 27 years old, respectively) to evaluate the intelligibility of the stimuli. All stimuli were correctly identified in a forced-choice identification task. Of the 720 tokens, 432 tokens were used for the perceptual tests in the Pretest, and in the PTDay1 and the PTDay2 phases, and 288 tokens were used in the Training phase. The details of the stimuli used in each phase are presented as follows.

The measurements of the stimuli presented later in this section were performed using Praat. Duration measurements (in ms) were made according to

\(^{21}\) These syllables are described later in this section.
standard segmentation practices (Peterson & Lehiste, 1960; Pollock, Brammer, & Hageman, 1993). Judgements of consonant and vowel boundaries (i.e., the onset and the offset of the vowel) were based on visual examination of the first and last identifiable periodic cycle in the waveform combined with the spectrogram. F0 measurements (in Hz) were made for individual F0 contours of the stimuli. As in Abramson’s Thai tone study (1976), the vocalic duration of each F0 contour was normalized: F0 values were sampled at five locations (0%, 25%, 50%, 75%, and 100%) along the vocalic portions of syllables.

In the Pretest phase, the identification test used three syllables (di, da, and du)\textsuperscript{22}, spoken by two of the four native Mandarin speakers (a female and a male). Each syllable can be with one of the four lexical tones. Altogether, 144 tokens were used for the Pretest (2 speakers x 3 syllables x 4 tones x 2 samples per tone x 3 repetitions). Table 2-1 shows the mean vowel durations and F0 values that were measured at five points along the vowel portion of the syllables.

In the Training phase, another 12 CV syllables (bi, pi, ci, shi, ma, ba, fa, zha, fu, gu, zhu, and shu)\textsuperscript{23} were used to train the listeners to learn the four Mandarin tones. All stimuli were spoken by the same female and male native Mandarin speakers whose speech samples were used as the stimuli in the Pretest. The training stimuli were 288 tokens (2 speakers x 12 syllables x 4 tones x 3 repetitions) which were evenly distributed among six sets of exercises for six pairs of tones (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, and T3-T4). The mean vowel duration

\textsuperscript{22} In IPA: [ti, ta, tu]. These three syllables were also used in Gottfried & Suiter (1997).
\textsuperscript{23} In IPA: [pi, pʰi, tsʰi, gi, ma, pa, fa, tʂa, fu, ku, tʂu, ʂu]
durations and F0 values measured at five points (0%, 25%, 50%, 75%, and 100%) along the vowel portion are presented in Table 2-2.

Table 2-1. Mean durations (ms) and F0 values (Hz) for the stimuli in the Pretest.

<table>
<thead>
<tr>
<th>Speaker (Tone)</th>
<th>Duration (ms)</th>
<th>0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (T1)</td>
<td>371.50</td>
<td>274.97</td>
<td>265.39</td>
<td>264.87</td>
<td>257.63</td>
<td>245.24</td>
</tr>
<tr>
<td>Female (T2)</td>
<td>399.17</td>
<td>220.84</td>
<td>203.30</td>
<td>228.41</td>
<td>247.66</td>
<td>268.01</td>
</tr>
<tr>
<td>Female (T3)</td>
<td>483.33</td>
<td>207.90</td>
<td>176.75</td>
<td>97.25</td>
<td>183.69</td>
<td>203.03</td>
</tr>
<tr>
<td>Female (T4)</td>
<td>280.17</td>
<td>275.35</td>
<td>259.73</td>
<td>245.94</td>
<td>223.78</td>
<td>166.41</td>
</tr>
</tbody>
</table>

| Male (T1)      | 361.17        | 148.38 | 144.78 | 142.69 | 139.63 | 135.67 |
| Male (T2)      | 390.50        | 106.28 | 93.99  | 95.93  | 116.80 | 149.94 |
| Male (T3)      | 496.50        | 109.34 | 86.32  | 68.78  | 110.38 | 134.30 |
| Male (T4)      | 257.33        | 146.29 | 137.30 | 120.04 | 99.21  | 74.51  |

Table 2-2. Mean durations (ms) and F0 values (Hz) for the stimuli in the training sessions.

<table>
<thead>
<tr>
<th>Speaker (Tone)</th>
<th>Duration (ms)</th>
<th>0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (T1)</td>
<td>330.49</td>
<td>265.29</td>
<td>261.13</td>
<td>258.95</td>
<td>252.97</td>
<td>242.62</td>
</tr>
<tr>
<td>Female (T2)</td>
<td>369.69</td>
<td>222.28</td>
<td>197.80</td>
<td>211.04</td>
<td>235.37</td>
<td>265.68</td>
</tr>
<tr>
<td>Female (T3)</td>
<td>442.43</td>
<td>226.48</td>
<td>166.93</td>
<td>103.57</td>
<td>161.48</td>
<td>206.06</td>
</tr>
<tr>
<td>Female (T4)</td>
<td>220.76</td>
<td>269.31</td>
<td>257.85</td>
<td>244.04</td>
<td>221.43</td>
<td>198.51</td>
</tr>
</tbody>
</table>

| Male (T1)      | 278.60        | 141.52 | 140.71 | 139.82 | 137.65 | 134.19 |
| Male (T2)      | 320.98        | 99.35  | 95.32  | 100.93 | 118.86 | 139.94 |
| Male (T3)      | 398.92        | 123.50 | 94.88  | 63.18  | 96.52  | 129.50 |
| Male (T4)      | 188.12        | 140.74 | 133.06 | 114.30 | 94.83  | 76.95  |

For the two posttest phases, the tests given on PTDay1 and PTDay2 were identical. In each phase, the Pretest identification stimuli were re-used as a posttest (Posttest1 on PTDay1 or Posttest2 on PTDay2). Additionally, three other kinds of novel stimuli were included for the three generalization tests (Gen1,
Gen2, and Gen3). Stimuli used during the training sessions were not used in any identification tests (either in the posttests or the generalization tests) on PTDay1 or on PTDay2.

For Gen1, 144 stimuli based on the same syllables (\(di, da, du\)), were spoken by two novel Mandarin speakers, the other female and male native speakers. Table 2-3 shows the mean durations and \(F0\) values of the stimuli.

### Table 2-3. Mean durations (ms) and \(F0\) values (Hz) for the stimuli in Gen1.

<table>
<thead>
<tr>
<th>Speaker (Tone)</th>
<th>Duration (ms)</th>
<th>0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (T1)</td>
<td>388.97</td>
<td>318.03</td>
<td>310.81</td>
<td>309.27</td>
<td>303.80</td>
<td>296.16</td>
</tr>
<tr>
<td>Female (T2)</td>
<td>420.35</td>
<td>226.52</td>
<td>223.75</td>
<td>233.87</td>
<td>264.32</td>
<td>305.08</td>
</tr>
<tr>
<td>Female (T3)</td>
<td>520.49</td>
<td>220.99</td>
<td>189.42</td>
<td>127.80</td>
<td>193.23</td>
<td>224.70</td>
</tr>
<tr>
<td>Female (T4)</td>
<td>270.57</td>
<td>332.21</td>
<td>316.52</td>
<td>279.55</td>
<td>225.10</td>
<td>175.48</td>
</tr>
<tr>
<td>Male (T1)</td>
<td>290.16</td>
<td>138.73</td>
<td>134.66</td>
<td>132.92</td>
<td>130.75</td>
<td>128.05</td>
</tr>
<tr>
<td>Male (T2)</td>
<td>334.12</td>
<td>107.95</td>
<td>107.70</td>
<td>117.97</td>
<td>132.61</td>
<td>143.40</td>
</tr>
<tr>
<td>Male (T3)</td>
<td>477.91</td>
<td>107.07</td>
<td>89.62</td>
<td>71.38</td>
<td>96.81</td>
<td>109.17</td>
</tr>
<tr>
<td>Male (T4)</td>
<td>381.46</td>
<td>159.55</td>
<td>114.47</td>
<td>102.73</td>
<td>109.72</td>
<td>71.42</td>
</tr>
</tbody>
</table>

For Gen2, new stimulus syllables (\(ch\), \(ch\), \(chu\))\(^{24}\) were spoken by the old speakers (who produced the stimuli in the Pretest and the training sessions), and 144 tokens were used (2 speakers x 3 syllables x 4 tones x 2 samples per tone x 3 repetitions). Table 2-4 presents the mean durations and \(F0\) values of the stimuli.

\(^{24}\)IPA: [tʰi], [tʰa], [tʰu]
Lastly, for Gen3, non-speech stimuli (i.e., filtered speech) were created from the 144 stimuli used in the Pretest. In this generalization test, the Pretest stimuli were low-pass filtered at 350 Hz for the female stimuli, and at 200 Hz for the male stimuli (Burnham et al., 1996; Ching, 1984; Stevens & Keller, 2001).

In all, the stimuli used in the experiment were in agreement with the descriptions for both the durational relationships and the pitch patterns among the four Mandarin tones as was mentioned in Section 1.6.2.

### 2.3 Procedure

This 4-day short-term longitudinal study employed a pretest-posttest paradigm (e.g., Logan et al., 1991; Lively et. al. 1993) using identification tasks throughout the experiment. The experiment consisted of four phases: Pretest, Training, PTDay1, and PTDay2. Each phase was conducted on a different day.

---

26 Identification tasks were used throughout the experiment. It has been suggested that identification tasks are more effective than discrimination tasks in "promoting generalization to novel stimuli not presented during training" (Logan & Pruitt, 1995, p. 359). Further, identification tasks are also more useful in training listeners to perceive non-native phonetic categories, since the tasks force the trainees to establish novel phonetic categories (Jamieson & Morosan, 1986).
The first three phases were given within 1-3 days of each other, and the final phase was a month later following the 3rd phase. All experimental phases were presented to individual participants using a custom-made program (written in HyperStudio 4.5) that ran on a PC laptop. The majority of the participants completed the experiments in the Phonetics Laboratory at Simon Fraser University. Four Japanese listeners participated in the experiments in a quiet room at the Harbour Centre campus of Simon Fraser University. The details of the four phases are described in the following sections.

2.3.1 Phase 1: Pretest

Prior to the Pretest, there was a familiarization block consisting of 24 speech samples of the four Mandarin tones on the syllables (chi, cha, chu) spoken by a female and a male native Mandarin speaker. On the computer screen, 24 buttons were displayed according to tone types and speakers. Each button was linked to a speech sample (see Figure 2-1). Once the trainees clicked on a button, an audio tone sample was presented. The process was self-paced. Participants were encouraged to listen to each of the speech samples at least once. They were not allowed to spend more than two minutes in this familiarization block.

The participants then proceeded to the Pretest. The identification test consisted of 144 trials in two random blocks (a female set and a male set), and stimuli were randomized and presented individually within each block. For each trial, a stimulus button was located at the top of the computer screen, and four response buttons corresponding to the four Mandarin tones were located under
the stimulus button (see Figure 2-2). The Pretest was a four-alternative forced-choice identification task. No feedback was given during the test. Once the listeners selected a response button, the program automatically provided a prompt to the next trial.

![Figure 2-1. A presentation screen in the familiarization block](image)

![Figure 2-2. A presentation screen in the Pretest](image)

2.3.2 Phase 2: Training

2.3.2.1 Training Procedures

The Training phase, a two-alternative forced-choice identification task, commenced between one and three days after the Pretest phase. Two sets of training programs were designed for providing the trainees with two different
types of feedback: Simple (SIM) and Audio-visual (AV) feedback. The stimuli in the programs were identical, but the type of the feedback used in each set was different. The listeners in the SIM groups received SIM feedback, while those in the AV groups were trained with AV feedback (detailed descriptions will be provided in the next section). They received feedback on a trial-by-trial basis, and the training process was self-paced.

Six training sets were prepared for the six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, and T3-T4), forming six tone blocks. Each tone block focused only on two contrastive tones (see Wang et al., 1999), and used the corresponding set of stimuli (48 training words) presented in four exercises. Stimuli were randomized within each exercise. The first three exercises (8 words each) individually provided three specific environments to the trainees, that is, syllables ending with one of the three vowels (/i, u, a/). Trainees were allowed to select any one of the three exercises to begin their training. The fourth exercise (24 words) was to be completed at the end of the training, and employed the syllables presented in the previous three exercises. However, the target tones in the fourth exercise were those that were not focused on in the previous exercises. A predetermined criterion was established for each of the tone blocks. If the listeners correctly identified at least 80% of the tokens in the fourth exercise, they could proceed to the next set of tone pairs. However, if they did not reach this criterion, the listeners were requested to retrain for that tone pair for no more than five repetitions. The training phase terminated once all six blocks of tone pair training were finished. In general, the listeners in the AV
groups took approximately 45 minutes to finish the training phase, while the listeners in the SIM groups spent from 1.25 to 2.5 hours to finish this phase.

2.3.2.2 Feedback Type: Audio-Visual (AV) vs. Simple (SIM)

As mentioned earlier, two types of (SIM & AV) feedback were used for training the listeners in the SIM and the AV groups. In the SIM training approach, as can be seen in Figure 2-3, each trial was presented individually as a single full-screen. A stimulus button was located at the top of the computer screen, and two response buttons, corresponding to the two Mandarin tones for a specific tone block, were located under the stimulus button. During individual training, a listener was required to click on the stimulus button, and then an audio was presented to her/him only once. After s/he selected an answer from the two button options, a text box would appear automatically to indicate if the response was correct or incorrect. Then, the trainee went to the next trial by clicking the arrow at the bottom right corner of the screen.

![Figure 2-3. Simple feedback for a correct (left) and an incorrect (right) response](image)

In the AV training approach, the trainees were asked to do the same task as those in the SIM groups. In this program, the stimuli, the way to present a trial,
and to prompt to a new trial screen, were identical to the SIM training approach. However, the feedback structure of AV was different from that of the SIM one.

As can be seen in Figure 2-4, the AV feedback consisted of three components: audio sound files (Audio), animated pitch contours (Animated Graph), and text descriptions (Text). Once listeners responded to a trial, three components would appear simultaneously on the screen: an audio button, an animated graph showing the pitch contours of the two tones, and a text box. The audio sound files for the pair of tones were presented consecutively when the button with an icon of a loudspeaker under the text panel was clicked. The animated pitch contours panel (bottom right-hand side of the screen) presented the pitch contours of the two tones one by one, according to the presentation sequence of the audio sound files. The trainees were told that the 5-point scale displayed in the pitch contour window was to help them to perceive and to understand the relationship between the tones in question. They were not required to know or to memorize the actual pitch values. The text box (bottom left-hand side of the screen) indicated if the response was correct or not. More importantly, it described the acoustic cue differences between the tones in the pair, the function of which was to direct listeners' attention to the crucial perceptual cues of the tones. For example, in terms of a durational cue for the Tone 1-Tone 4 pair, the description was “Tone 4 is shorter than Tone 1”.

26 The 5-point scale was adopted from Chao's tone letters (1930). The system provides “simplified time-pitch graphs of the voice” (1930, p.24). Each of the tones is represented by 2 or 3 letters to indicate the initial and the final or the initial, the medial and the final position of the tone. For Mandarin tones, T1 = [55], T2 = [35], T3 = [214], and T4 = [51].
27 See the descriptions of Mandarin tones in Section 1.6.2.
During the training, the trainees in the AV groups were instructed to at least once read the text, play back the sound files, and study the graph. They were requested to do so even if their responses were correct. This requirement functioned as reinforcement (see Francis & Nusbaum, 2002).

![Figure 2-4. AV feedback for a correct (left) and an incorrect (right) response](image)

It should be mentioned that the AV feedback (i.e., **Audio, Animated Graph, and Text**) used in the present study was adapted from the one used in So (2003). However, So (2003) employed static pitch contour graphs instead of the animated pitch contours employed in the current study. The major advantage of using the animated graphs is that animation, but not the static graphs, can illustrate *spatial-temporal actions* or *microsteps of processes* (Thompson & Riding, 1990; Tversky, Morrison, & Betrancourt, 2002). Furthermore, previous studies have demonstrated that animated graphs provide better facilitative effects than do static graphs in depicting the fine-grained actions during learning (e.g., McCuistion, 1991; Park & Gittelman, 1992). Thus, the animated graphs showing tonal contours in this study would have a greater facilitative effect on the trainees' learning than the static graphs employed in So (2003). Through illustrating the

---

28 They were permitted to replay the components several more times, if desired.
spatial-temporal actions of lexical tones, the animated contours would help the trainees establish mental patterns of tones, thus facilitating the learning process.

There are two advantages to the use of this type of 3-component AV feedback. First, trainees can observe the pitch movements while the audio sound files are being played back. As a result, the animated pitch contours together with the audio sound files provide trainees with dual channel inputs to illustrate spatial-temporal actions. These components thus fulfil a similar function as in previous studies that used a real-time display to show pitch contours together with the audio sound files (Bluhme & Burr, 1971-1972; deBot, 1983; Sanders, et al., 2005). Second, the concise texts explicitly direct the trainees' attention to the crucial acoustic cues for the Mandarin tones. The trainees can understand the content of the texts more easily when they relate the texts to the animated pitch contours and the audio sound files.

2.3.3 Phases 3 and 4: Posttest-Day1 (PTDay1) and Posttest-Day2 (PTDay2)

The identification tests in the two posttest phases, PTDay1 & PTDay2, were identical, and were used to assess trainees' performance before and after training at different times. PTDay1 took place one or two days after the training sessions. PTDay2 was administered approximately 1 month after the PTDay1.

29 For the present study, though animated graphs did not synchronize the pitch contours and the corresponding sound files in the same way as the real-time displays, earlier studies were also unable to synchronize the two components in phase. Nevertheless, the results of these studies demonstrated a facilitative effect of AV feedback (e.g., Abberton & Fourcin, 1975; de Bot, 1983). Further, in a feedback delay experiment conducted by Weltens & de Bot (1984), AV feedback was still found to be effective even the feedback was given after the end of each speech signal. They concluded that "pitch visualizers" did not have to be "real-time in order to be effective" (Weltens & de Bot, 1984, p. 87).
In each posttest phase, the participants were given not only the same identification test used in the Pretest (144 trials), but also three other generalization tests: Gen1, Gen2, and Gen3. Each test consisted of 144 trials in two random blocks (a female set and a male set), and stimuli were randomized and presented individually within each block. The presentation and response task was the same as the one in the Pretest. The three generalization tests were employed to investigate how well the listeners can transfer the knowledge gained from the training to perceive or identify: (i) old stimuli used in the Pretest but produced by novel speakers (Gen1), (ii) novel stimuli spoken by old speakers (Gen2), and (iii) nonspeech stimuli created by low-pass filtering the stimuli used in the Pretest (Gen3).

2.4 Analysis

The primary focus of the present study was to examine the performance of the participants at each experimental phase: Pretest, Training, PTDay1 and PTDay2. Section 2.4.1 will describe the method used to analyze the identification tests in the Pretest, PTDay1 and PTDay2 phases. Section 2.4.2 will present the way in which the Training phase was analyzed. Finally, Section 2.4.3 will describe the methods to compare the participants' performance across time to see the effects of training, and to examine the trainees' retentions of the Mandarin tones.

30 See Section 2.2.1.
2.4.1 Analyses for Pretest and Posttest Phases

Listeners' performance on the identification tests in the Pretest, the PTDay1 and the PTDay2 phases was statistically compared in three domains. First, the listeners' percent correct scores were computed and compared. The results were expected to show how well trainees could identify the four lexical tones.

Second, because percent correct scores did show how well listeners could differentiate the target tones from their counterparts, the listeners' sensitivity to each lexical tone was examined in terms of A-Prime (A') scores. A' scores were calculated by using the formulas provided by Snodgrass, Levy-Berger, & Haydon (1985, p. 451):

If $H$ is greater than $FA$: $A' = 0.5+(((H- FA)*(1+ H - FA))/(4*H)*(1- FA)))$
If $H$ is equal to $FA$: $A' = 0.5$
If $FA$ is greater than $H$: $A' = 0.5-(((FA-H)*(1+FA-H))/(4*FA)*(1-H)))$

In the above formulas, $H$ stands for Hit rates (the percentage of hits for a particular response category), and $FA$ stands for False Alarms (the percentage of false alarms). The calculated $A'$ scores fall in the range from 0 to 1 (i.e., from poor to perfect discrimination). The greater the accuracy of the listeners' tonal identifications, the closer the $A'$ scores for the tones would be to 1. In contrast, if the listeners are less able to identify the target tones from the other tones, the $A'$ scores would be at 0.5 or lower (i.e., insensitive to tonal contrasts).

Third, in order to determine the types and the frequencies of the listeners' tonal confusions, the listeners' tonal identification errors were also examined. Individual confusion matrices were constructed according to the listeners' L1
backgrounds, and the training approach they received during training (AV vs. SIM).

2.4.2 Analysis for the Training Phase

The trainees' performance in the Training phase was also examined in terms of the amount of time the learners spent on training. The analysis compared the number of sessions taken for each tone block. This comparison investigated the performance differences among the trainee groups.

2.4.3 Analysis for Training Improvement

The identification tests (i.e., the one used for the Pretest and the posttest, and the three generalization tests) were administered at different times to examine the trainees' performance across time, the purpose of which was to examine the effect of training on the listeners' overall improvements and their retention of the Mandarin tones after a month.

To this end, only percent correct scores, but not A' scores and tonal confusions, were computed for analysis of improvement shown by trainees after training. The main reason was that previous studies has demonstrated that using percent correct scores is sufficient to show the effect of training on trainees' improvements and their retentions of the target segments or suprasegmentals (e.g., Bradlow et al., 1997; Wang et al., 1999).
CHAPTER 3: RESULTS: PRETEST

In this chapter, the performance (tonal identification) of the listeners in the AV and the SIM groups for the Pretest was analyzed in three domains: percent correct scores (%), sensitivity to the four Mandarin tones (A?), and tonal confusions (errors). The results will be presented in Sections 3.1-3.3, respectively.

3.1 Percent Correct Scores

Figure 3-1 displays the mean % scores of the listeners in the AV and the SIM group in the Pretest. It can be seen that the mean scores of the listeners in the AV groups (50-59%) were similar to those in the SIM (50-59%) groups. The pattern was consistent across the language groups (Cantonese, English, and Japanese). In addition, the scores of the Cantonese and of the Japanese listeners were slightly higher than those of the English listeners. The pattern was seen in both the AV and SIM groups.

A 2-way analysis of variance (ANOVA) with two between-subjects factors, L1 Group (CAN, ENG, and JPN) and Training Approach (AV and SIM), was carried out to examine their effects on the listeners' mean scores. The results showed that L1 Group effect was significant, $F(2,24)=3.296, \ p<.05$. Post-hoc Tukey tests further revealed that the English listeners' mean score ($\approx 50\%$) was significantly lower than those ($\approx 59\%$ each) of the Cantonese and the Japanese
listeners (ps<.05), but the Cantonese and Japanese groups did not differ from each other (p>.05). However, the effects of Training Approach and the interaction of L1 Group x Training Approach were not significant, $F(1,24)=.001$, $p>.05$, and $F(2,24)=.005$, $p>.05$, respectively. The results imply that the mean scores for the listeners in the AV and the SIM groups were similar to one another before training, and that L1 backgrounds affected the listeners' performance, as was evidenced by the Cantonese and the Japanese listeners' scores, which were significantly higher than those of the English listeners.

![Figure 3-1. Mean percent correct scores (%) in Pretest by the listener in the AV (left) and the SIM (right) groups](image)

### 3.2 A-prime Scores

A summary of the listeners' mean $A'$ scores for the four Mandarin tones is presented in Table 3-1. It can be seen that the listeners' $A'$ scores ranged from 0.72 to 0.88, except for the English listeners' $A'$ scores for Tone 4 ($A'$ scores ≈0.5).
Table 3-1. Mean A’ scores for the Mandarin tones in Pretest for the listener groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2 AV</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 1</th>
<th>Tone 2 SIM</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.8480</td>
<td>0.7492</td>
<td>0.8033</td>
<td>0.8792</td>
<td>0.8379</td>
<td>0.8094</td>
<td>0.7943</td>
<td>0.8454</td>
</tr>
<tr>
<td>ENG</td>
<td>0.7802</td>
<td>0.7307</td>
<td>0.8286</td>
<td>0.5301</td>
<td>0.8105</td>
<td>0.7457</td>
<td>0.8655</td>
<td>0.5509</td>
</tr>
<tr>
<td>JPN</td>
<td>0.8708</td>
<td>0.7210</td>
<td>0.7973</td>
<td>0.8341</td>
<td>0.8361</td>
<td>0.7545</td>
<td>0.8437</td>
<td>0.8105</td>
</tr>
</tbody>
</table>

The listeners' mean A’ scores were submitted to a 3-way mixed-design ANOVA with L1 Group (CAN, ENG, and JPN) and Training Approach (AV and SIM) as the 2 between-subjects factors, and with Tone (T1, T2, T3, and T4) as the within-subjects factor. The results revealed significant effects of L1 Group, $F(2,24)=5.54$, $p=.01$, and of Tone, $F(3,72)=6.973$, $p<.001$. However, the effect of Training Approach did not reach any statistical significance ($p>.05$), suggesting that the sensitivity of the listeners in the AV and the SIM groups were similar to each other regardless of the listeners' L1 backgrounds. In addition, the interaction of L1 Group x Tone was also significant, $F(6,72)=8.399$, $p<.001$.

To explore the interaction of L1 Group x Tone, four separate 1-way ANOVAs were carried out to investigate the effect of L1 Group on the mean A’ scores for the four tones. The analyses indicated a significant L1 Group effect for Tone 4 only, $F(2,27)=15.847$, $p<.001$, but not for the other three tones ($ps>.05$). Post-hoc Tukey tests revealed that the English listeners’ A’ scores for Tone 4 were significantly lower than those of the Cantonese and the Japanese listeners ($ps<.001$), implying that the English listeners' sensitivity to Tone 4 (A’ scores ≈0.5 for the AV and SIM groups) before training was significantly lower than that of the
Cantonese (AV = 0.8792; SIM = 0.8454) and the Japanese listeners (AV=0.8341; SIM=0.8105).

### 3.3 Tonal Confusions

The listener groups' responses in the Pretest are tabulated in Table 3-2, which consists of the six matrices for the three listener groups (Cantonese, English, and Japanese). Each language group has two matrices that show the patterns for the AV and the SIM groups (top and bottom sections, respectively). The shaded cells are the correct responses for the target tones, and the rest of the cells indicate the incorrect responses for the target tones.

#### Table 3-2: Confusion Matrices for the Responses in Pretest Performed by the Listener Groups

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td>Response</td>
<td>Response</td>
<td>Response</td>
</tr>
<tr>
<td>Target</td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
</tr>
<tr>
<td>T1</td>
<td>121 27 0 32 180</td>
<td>118 33 16 13 180</td>
<td>124 42 1 13 180</td>
</tr>
<tr>
<td>T2</td>
<td>18 98 61 3 180</td>
<td>40 91 39 10 180</td>
<td>6 103 67 4 180</td>
</tr>
<tr>
<td>T3</td>
<td>2 86 92 0 180</td>
<td>4 27 106 43 180</td>
<td>0 71 103 6 180</td>
</tr>
<tr>
<td>T4</td>
<td>58 6 1 115 180</td>
<td>87 43 5 45 180</td>
<td>49 36 0 95 180</td>
</tr>
<tr>
<td>Total</td>
<td>199 217 154 150 720</td>
<td>249 194 166 111 720</td>
<td>179 252 171 118 720</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SIM</strong></th>
<th>Response</th>
<th>Response</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
</tr>
<tr>
<td>T1</td>
<td>120 8 0 52 180</td>
<td>131 29 5 15 180</td>
<td>110 58 1 11 180</td>
</tr>
<tr>
<td>T2</td>
<td>13 115 51 1 180</td>
<td>32 97 30 21 180</td>
<td>20 101 53 6 180</td>
</tr>
<tr>
<td>T3</td>
<td>2 96 82 0 180</td>
<td>3 53 111 13 180</td>
<td>4 20 117 39 180</td>
</tr>
<tr>
<td>T4</td>
<td>71 2 0 107 180</td>
<td>120 30 6 24 180</td>
<td>44 40 1 95 180</td>
</tr>
<tr>
<td>Total</td>
<td>206 221 133 160 720</td>
<td>286 209 152 73 720</td>
<td>178 219 172 151 720</td>
</tr>
</tbody>
</table>

From Table 3-2, several confusion patterns can be noted. Some patterns are found across the language groups, while others are language-particular.
First, listeners' tonal confusions were mainly found in three pairs of tones: T1-T4, T2-T3, and T1-T2 (bi-directionally\textsuperscript{31}). For the T1-T4 pair, the listeners in all three language groups tended to misidentify target T4 as T1 more frequently than they did T1 as T4. For the T2-T3 pair, both the Cantonese and the English listeners more often misidentified target T3 as T2 than they did T2 as T3. However, the Japanese listeners frequently misidentified target T2 as T3. For the T1-T2 pair, the confusion pattern was more obviously found in the matrices for the English and the Japanese listeners than in the one for the Cantonese listeners. Second, as mentioned earlier, the Cantonese listeners' tonal confusions were more limited to the T1-T4, T2-T3, and T1-T2 pairs, and they had fewer confusions in other tone pairs. In contrast, both the English and the Japanese listeners showed confusions in almost every possible tone pair. Third, both the English and the Japanese listeners exhibited a considerable amount of confusion between the tones in the T2-T4 pairs. They frequently used Tone 2 as a default response for Tone 4. Lastly, Japanese listeners consistently selected Tone 2 as a default answer for the other tones.

The listeners' mean tonal confusions (errors) were submitted to a 3-way mixed-design ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone Pair\textsuperscript{32} (T1T2, T2T1, T1T3, T3T1, T1T4, T4T1,

\textsuperscript{31} For each tone pair, the directions of the tonal misidentifications could be in two ways. In the case of the T1-T4 pair, for example, the two possible directions are: T1 could be misidentified as T4, or vice versa.
\textsuperscript{32} Huang (2001) also used tone pair as a factor to investigate the differences in the reaction times of her listeners' perception responses.
T2T3, T3T2, T2T4, T4T2, T3T4, T4T3\(^{33}\) as the within-subjects factor. The results revealed that the effects of L1 Group, \(F(2,24)=3.121, p>.05\), and of Training Approach, \(F(1,24)=.001, p>.05\), were not significant, but that a significant effect of Tone Pair, \(F(11,264)=32.541, p<.001\), was found. The two interactions, Tone Pair x L1 Group, \(F(22,264)=5.751, p<.001\), and Tone Pair x L1 Group x Training Approach, \(F(22,264)=.001, p>.05\), were significant. However, the interactions of Tone Pair x Training Approach and of L1 Group x Training Approach did not reach any significance (\(p>.05\)).

To further explore the 3-way interaction among the factors (L1 Group Training Approach, and Tone Pair), two separate 2-way mixed-design ANOVAs were performed for the AV and the SIM groups to investigate the effects of L1 Group and of Tone Pair, as well as their interaction. For the tonal confusions made by the listeners in the AV groups, the analysis revealed that the L1 Group effect was not significant, \(F(2,12)=1.607, p>.05\), but that the effect of Tone Pair was significant, \(F(11,132)=16.676, p<.001\). The interaction of L1 Group x Tone Pair was also significant, \(F(22,132)=2.805, p<.001\), suggesting that the differences in the mean tonal errors among the Cantonese, the English, and the Japanese listeners in the AV groups were greater in some tone pairs, but not in all. Individual 1-way ANOVAs were run to further investigate the L1 Group effect for each tone pair. The results indicated that the L1 Group effect was found in the T1T2, T1T4, T2T3, and T4T3 pairs, \(Fs(2,12)=4.956, 4.366, 4.344,\) and \(5.916\), respectively, \(p<.05\). Post Hoc Tukey tests found that (i) the English listeners'
mean errors ($M$s=8, 17.4, and 8.6) for the T1T2, T1T4, and T4T3 pairs were significantly greater than for those ($M$s=1.2, 9.8, and 1.2) of the Japanese listeners ($p$s<.05); (ii) the English listeners ($M$=8.65) also had more errors than the Cantonese listeners ($M$=0) for the T4T3 pair; and (iii) the Cantonese listeners ($M$=17.2) made significantly more errors for the T2T3 pair than did the English listeners ($M$=6) ($p$.05). A summary of these patterns is presented in Figure 3-2.

![Figure 3-2](image_url)

*Figure 3-2. Tonal confusions in Pretest by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category*

In the analysis of the mean tonal confusions made by the listeners in the SIM groups, the effect of L1 Group was not significant, $F(2,12)$=1.515, $p$.05, but the effect of Tone Pair was significant, $F(11,132)$=16.470, $p$.001. Also, the interaction of L1 Group x Tone Pair was significant, $F(22,132)$=4.642, $p$.001. Individual 1-way ANOVAs were performed to examine the L1 Group effect for each tone pair. The L1 Group effect was significant for the T1T4, T2T3, T2T1,
T4T1 and T4T2 pairs, $F_{s}(2,12) = 8.572, 6.072, 16.232, 7.862, \text{ and } 6.436$, respectively, $p<.05$. Post Hoc Tukey tests further revealed that (i) the English listeners' mean errors for the T1T4 and T4T2 ($M_{s}=24$ and 4.2) pairs were significantly greater than those of the Cantonese ($M_{s}=14.2$ and 0.2) and the Japanese ($M_{s}=8.8$ and 1.2) listeners ($p<.05$); (ii) the Japanese listeners ($M=11.6$) made more errors than did the English ($M=5.8$) and the Cantonese ($M=1.6$) listeners for the T2T1 pair ($p<.05$); and (iii) the Cantonese listeners ($M=19.2$) showed significantly more errors than did the Japanese ($M=4$) and the English ($M=10$) listeners in the T2T3 and T4T1 pairs ($p<.05$). The results are presented in Figure 3-3.

*CAN-ENG ($p<.05$)*
*CAN-JPN ($p<.05$)*
*ENG-JPN ($p<.05$)*

**Figure 3-3.** Tonal confusions in Pretest by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.
In sum, the analyses of the listeners' performance in the Pretest phase that focused on the three domains (percent correct scores, A' scores, and tonal confusions) consistently showed that the performance of the listeners in the AV and the SIM groups were comparable to one another. There was no evidence of a difference between the listeners in the AV and in the SIM groups in the Pretest. In addition, the L1 effect was present in the Pretest. The English listeners' Pretest scores were significantly lower than those of the Cantonese and the Japanese listeners. The A' analysis showed that the English listeners were less sensitive to Tone 4 than the Cantonese and the Japanese listeners. The three language groups also exhibited some differences in the patterns of their tonal errors, but their tonal confusions were mainly found in three pairs of tones: T1-T2, T1-T4, and T2-T3 (further discussions will be provided in Chapter 8).
CHAPTER 4: RESULTS: TRAINING DURATION

This chapter reports the results of the analysis of the performance of the participants in the Training phase. The focus was the comparison between the means of training duration (in terms of number of sessions) by the listeners in the AV and the SIM groups.

As seen in Table 4-1, the listeners in the AV groups spent less time (mean number of sessions ranged from 1 to 1.2) on the Training phase than those in the SIM groups (mean sessions ranged from 1 to 2.8). However, all the listeners, regardless of their L1 backgrounds and the training approaches they received, needed more training sessions for two specific training sets (i.e., tone pairs), T1-T4 and T2-T3. Figures 4-1 and 4-2 display the mean number of training sessions for the listeners in the AV and the SIM groups, respectively.

Table 4-1: Mean Number of Training Sessions for the Training Sets during Training

<table>
<thead>
<tr>
<th>L1 Group (Training Approach)</th>
<th>T1-T2</th>
<th>T1-T3</th>
<th>T1-T4</th>
<th>T2-T3</th>
<th>T2-T4</th>
<th>T3-T4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN (AV)</td>
<td>1</td>
<td>1</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>6.4</td>
</tr>
<tr>
<td>ENG (AV)</td>
<td>1</td>
<td>1</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>6.4</td>
</tr>
<tr>
<td>JPN (AV)</td>
<td>1</td>
<td>1</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>6.6</td>
</tr>
<tr>
<td>CAN (SIM)</td>
<td>1</td>
<td>1</td>
<td>2.8</td>
<td>2.2</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>ENG (SIM)</td>
<td>2.2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8.2</td>
</tr>
<tr>
<td>JPN (SIM)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Note: Each session contains 48 trials.
Figure 4-1. Mean number of training sessions for the training sets spent by the listeners in the AV groups.

Figure 4-2. Mean number of training sessions for the training sets spent by the listeners in the SIM groups.
To investigate the effects of three factors on the listeners' mean number of sessions spent on the tone pair training, the data were submitted to a 3-way mixed-design ANOVA with L1 Group (CAN, ENG, and JPN) and Training Approach (AV and SIM) as the between-subjects factors, and with Training Set (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, and T3-T4) as the within-subjects factor. The analysis revealed that the effects of Training Set, $F(5, 120)=9.745, p<.001$, and of Training Approach were significant, $F(1, 24)=14.383, p=.001$. Post-hoc LSD tests for the Training Set revealed that the listeners' mean training session for T1-T4 was significantly greater than those for all the other five sets ($ps<.05$). The same result pattern was found for T2-T3 ($ps<.05$), except that the mean difference between T2-T3 and T1-T2 was not significant ($p>.05$) and that the listeners' mean training session for T2-T3 was fewer than the one for T1-T4. The L1 Group effect was not significant, $F(2, 24)=.787, p>.05$. The interactions of Training Set x L1 Group, $F(10, 120)=2.205, p<.05$, of Training Set x Training Approach, $F(5, 120)=5.327, p<.001$, and of Training Set x L1 Group x Training Approach, $F(10, 120)=1.986, p<.05$, were all significant.

To examine the interaction of Training Set x L1 Group x Training Approach, two separate 2-way mixed-design ANOVAs were carried out to test the effects of L1 Group and Training Set on the data for the AV and the SIM groups. For the listeners in the AV groups, there were no significant effects of Training Set, $F(5, 60)=2.00, p>.05$, and of L1 Group, $F(2, 27)=.519, p>.05$. Moreover, the interaction of Training Set x L1 Group was not significant, $F(10, 135)=.539, p>.05$. For the listeners in the SIM groups, the analysis indicated
a significant effect of Training Set, $F(5,60)=8.496, p<.001$, but the effect of L1 Group was non-significant, $F(10,120)=1.28, p<.05$. The interaction of Training Set x L1 Group was significant, $F(10,60)=2.432, p<.01$. Separate 1-way ANOVAs were performed to explore the L1 Group effect on the listeners' mean training sessions for each of the training sets. The results showed that a significant effect of L1 Group was only found in the T2-T3 pair, $F(2,12)=6.889, p=.01$. Post-hoc Tukey tests showed that the Cantonese listener group's mean number of training sessions ($M=2.2$) was significantly greater than those of the English ($M=1$) and the Japanese ($M=1.2$) listener groups ($p<.05$).

Overall, the analysis of the listeners' training duration indicated that (i) the listeners in the AV groups took significantly less time to complete the training sessions than those in the SIM groups, and that (ii) two training sets, T1-T4 and T2-T3, required more time to learn. In addition, while the training times spent by the listeners in the three AV groups were comparable, the times required to complete the training by the listeners in the three SIM groups were different. In particular, the Cantonese trainees in the SIM group significantly spent more time learning the T2-T3 pair than did the English and Japanese trainees.
CHAPTER 5: RESULTS: IDENTIFICATION TESTS ON POSTTEST-DAY1 (PTDAY1)

This chapter reports the analyses of the listeners’ performance on the four identification tests in the PTDay1 phase. The listeners’ performance was compared with respect to the three domains: percent correct scores, A-prime (A') scores, and tonal confusions. The statistical analyses used for each of the identification tests on PTDay1 (Posttest1, Gen1, Gen2, and Gen3) were the same as those performed for the Pretest (see Chapter 3). The results for the Posttest1 will be presented in Section 5.1, and the results for the three generalization tests (Gen1, Gen2, and Gen3) will be presented in Sections 5.2-5.4, respectively.

5.1 Posttest1 on PTDay1

5.1.1 Percent Correct Scores

Figure 5-1 illustrates the listeners’ mean percent correct scores for the identification test of Mandarin tones in Posttest1. As can be seen in the figure, the listeners in the AV groups (91-97%) consistently scored higher than did the listeners in the SIM groups: (81-88%). Their score differences ranged from 8 to 11%: CAN (AV: 91.39% and SIM: 80.83%), ENG (AV: 96.39% and SIM: 85.69%), and JPN (AV: 96.53% and SIM: 88.06%).
A 2-way ANOVA with 2 between-subjects factors was conducted to determine the effects of L1 Group and of Training Approach on the listeners’ mean percent correct scores in Posttest1 on PTDay1. The results showed a significant effect of Training Approach, \( F(1,24)=17.940, p<.001 \), implying that the mean scores of the AV groups were significantly greater than those of the SIM groups. However, both the effect of L1 Group and the interaction of L1 Group x Training Approach were not significant (\( p>.05 \)).

![Mean percent correct scores (%) in Posttest1 on PTDay1 by the listener in the AV (left) and the SIM (right) groups](image)

**Figure 5-1.** Mean percent correct scores (%) in Posttest1 on PTDay1 by the listener in the AV (left) and the SIM (right) groups.

### 5.1.2 A-prime Scores

Table 5-1 presents a summary of listeners’ mean A’ scores for the four Mandarin tones in Posttest1. In general, the A’ scores were similar to one another, ranging from 0.91 to 0.99. The A’ scores for the four tones by the listeners in the AV groups were consistently higher than were those scored by the listeners in the SIM groups.
Table 5-1. Mean A' scores for the Mandarin tones in Posttest1 on PTDay1 for the listener groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9690</td>
<td>0.9901</td>
<td>0.9828</td>
<td>0.9721</td>
<td>0.9441</td>
<td>0.9273</td>
<td>0.9141</td>
<td>0.9333</td>
</tr>
<tr>
<td>ENG</td>
<td>0.9844</td>
<td>0.9911</td>
<td>0.9861</td>
<td>0.9896</td>
<td>0.9667</td>
<td>0.9400</td>
<td>0.9372</td>
<td>0.9498</td>
</tr>
<tr>
<td>JPN</td>
<td>0.9995</td>
<td>0.9865</td>
<td>0.9739</td>
<td>0.9926</td>
<td>0.9690</td>
<td>0.9542</td>
<td>0.9434</td>
<td>0.9589</td>
</tr>
</tbody>
</table>

The mean A' scores were submitted to a 3-way ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone as the within-subjects factor. The analysis showed that only the Training Approach effect was significant, $F(1, 24) = 19.817, p < .001$. The effects of L1 Group and of Tone, and the interactions among the factors, were all non-significant ($p > .05$). These results suggest that the listeners in the AV groups had higher mean A' scores for the four tones than did those in the SIM groups.

5.1.3 Tonal Confusions

The responses of the listener groups in Posttest1 are presented in Table 5-2. It can be seen that the listeners' tonal confusions were mostly found in two pairs of tones, T1-T4 and T2-T3 (errors were bi-directional). For the T1-T4 pair, the three groups of listeners tended to misidentify target T4 as T1 more frequently than they misidentified target T1 as T4. For the T2-T3 pair, the listeners more often misidentified target T3 as T2 than they did T2 as T3. The patterns were similar to those observed in the Pretest (see Chapter 3). In addition, the T1-T2 pair confusion was more frequently found in the matrices for the listeners in the SIM groups than for those in the AV groups. Finally, the
trainees in the SIM groups made more tonal errors overall than did those in the AV groups.

Table 5-2. Confusion Matrices for the Responses in Posttest1 on PTDay1 Performed by the Listener Groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>Response</td>
<td>Response</td>
<td>Response</td>
</tr>
<tr>
<td>T1</td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
</tr>
<tr>
<td></td>
<td>164</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>1</td>
<td>178</td>
<td>1</td>
</tr>
<tr>
<td>T3</td>
<td>1</td>
<td>11</td>
<td>168</td>
</tr>
<tr>
<td>T4</td>
<td>14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>180</td>
<td>193</td>
<td>169</td>
</tr>
</tbody>
</table>

| **SIM**  |            |         |          |
| Target   | Response   | Response| Response |
| T1       | T1         | T2      | T3       | T4 | Total |
|          | 163        | 6       | 0        | 13 | 180   |
| T2       | 11         | 151     | 17       | 1  | 180   |
| T3       | 0          | 54      | 0        | 164| 180   |
| T4       | 30         | 1       | 0        | 149| 180   |
| **Total**| 208        | 206     | 143      | 163| 720   |

The trainees' mean tonal errors were submitted to a 3-way mixed design ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone Pair as the within-subjects factor. The results showed that both the effects of Training Approach, $F(1,24)=22.707, p<.001$, and of Tone Pair, $F(11, 264)=21.185, p<.001$, were significant. The interaction of Training Approach x Tone Pair was also significant, $F(11,264)=7.498, p<.001$. However, the L1 Group effect and other interactions among the factors were not significant ($ps>.05$). These findings suggest that the trainees in the AV groups made significantly fewer tonal errors than did those in the SIM groups (see Figures 5-2 and 5-3).
Figure 5-2. Tonal confusions in Posttest1 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.

Figure 5-3. Tonal confusions in Posttest1 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.
5.2 Gen 1 on PTDay1: Novel Speakers

5.2.1 Percent Correct Scores

Figure 5-4 displays the mean percent correct scores for Gen1 on PTDay1 by the trainees in the AV (95-98%) and in the SIM (83-88%) groups. The listeners in the AV groups consistently obtained higher scores than did the listeners in the SIM groups (8-12% difference): CAN (AV: 94.58% and SIM: 82.5%), ENG (AV: 95.97% and SIM: 88.06%), and JPN (AV: 97.64% and SIM: 88.47%).

![Bar chart showing mean percent correct scores for Gen1 on PTDay1 by AV and SIM groups](image)

The listeners' mean scores were submitted to a 2-way ANOVA to examine the effects of L1 Group and Training Approach. The results showed that the effect of Training Approach was significant, $F(1,24)=23.756, \ p<.001$. However, the L1 Group effect, $F(2,24)=1.874, \ p>.05$, and the interaction of L1 Group x
Training Approach, $F(2,24)=.383, p>.05,$ were not significant. The results imply that the trainees in the AV groups outperformed the listeners in the SIM groups.

5.2.2 A-prime Scores

Table 5-3 summarizes the listeners’ mean A’ scores for the four Mandarin tones in Gen1 on PTDay1. The A’ scores were high, ranging from 0.92 to 1. The A’ scores for the four tones obtained by the listeners in the AV groups (0.98–1) were consistently greater than those scored by the listeners in the SIM groups (0.92–0.97).

<table>
<thead>
<tr>
<th></th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9837</td>
<td>0.9845</td>
<td>0.9930</td>
<td>0.9890</td>
</tr>
<tr>
<td>ENG</td>
<td>0.9811</td>
<td>0.9895</td>
<td>0.9898</td>
<td>0.9842</td>
</tr>
<tr>
<td>JPN</td>
<td>1.0000</td>
<td>0.9902</td>
<td>0.9782</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

The mean A’ scores were submitted to a 3-way ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone as the within-subjects factor. The analysis showed that only the Training Approach effect was significant, $F(1,24)=27.642, p<.001.$ The effects of L1 Group and of Tone, and the interactions of L1 Group x Tone, of Training Approach x Tone, of L1 Group x Training Approach, and of L1 Group x Training Approach x Tone were all non-significant ($ps>.05$). These results indicate that the listeners in the AV groups had higher mean A’ scores for the four tones than did those in the SIM groups (see Figures 5-5 and 5-6).
5.2.3 Tonal Confusions

Table 5-4 presents the listener groups' responses for Gen1 on PTDay1, the patterns of which were comparable to those observed in Posttest1.

The trainees' mean tonal errors were submitted to a 3-way mixed-design ANOVA to examine the effect of L1 Group, Training Approach and Tone Pair. The results revealed significant effects of Training Approach, $F(1,24)=28.795$, $p<.001$, and of Tone Pair, $F(11,64)=14.514$, $p<.001$. The interaction of Training Approach x Tone Pair was also significant, $F(11,264)=6.658$, $p<.001$. However, the L1 Group effect and other interactions among the factors (L1 Group x Tone Pair, L1 Group x Training Approach, and L1 Group x Training Approach x Tone Pair) were not significant ($p>.05$). These findings indicate that the trainees in the AV groups made significantly fewer tonal errors than did those in the SIM groups.

Table 5-4. Confusion Matrices for the Responses in Gen1 on PTDay1 Performed by the Listener Groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>T1</td>
<td>T2</td>
<td>T3 - T4</td>
</tr>
<tr>
<td>T1</td>
<td>172</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>4</td>
<td>170</td>
<td>6</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>3</td>
<td>177</td>
</tr>
<tr>
<td>T4</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>181</td>
<td>173</td>
<td>183</td>
</tr>
</tbody>
</table>

| **SIM**  |           |         |          |
| Response | Total     |         |          |
| Target   | T1 | T2 | T3 - T4 | Total |
| T1       | 169| 1  | 0        | 180   |
| T2       | 3  | 150| 27       | 180   |
| T3       | 0  | 44 | 136      | 180   |
| T4       | 40 | 1  | 139      | 180   |
| **Total**| 212| 196| 163      | 720   |
Figure 5-5. Tonal confusions in Gen 1 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.

Figure 5-6. Tonal confusions in Gen 1 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.
5.3 Gen 2 on PTDay1: Novel Stimuli

5.3.1 Percent Correct Scores

Figure 5-7 shows the mean percent correct scores for Gen2 on PTDay1 by the listeners in the AV (91-97%) and in the SIM (79-85%) groups. As can be seen in the figure, the scores of the listeners in the AV groups were consistently higher than those of the listeners in the SIM groups (10-13% difference): CAN (AV: 92.08% vs. SIM: 79.44%), ENG (AV: 91.39% vs. SIM: 81.81%) and JPN (AV: 97.6% vs. SIM: 85.28%).

![Figure 5-7. Mean percent correct scores (%) in Gen2 on PTDay1 by the listener in the AV (left) and the SIM (right) groups](image)

The mean correct scores were submitted to a 2-way ANOVA to examine the effects of L1 Group and Training Approach, and their interaction. The results revealed a significant effect of L1 Group, $F(2,24)=3.560$, $p<.05$. Post-hoc Tukey tests revealed that the Japanese listeners’ mean score was significantly higher...
than that of the Cantonese trainees ($p=.05$). The effect of Training Approach was also significant, $F(1,24)=38.897$, $p<.001$, implying that the trainees in the AV groups outperformed the listeners in the SIM groups. However, the interaction of L1 Group x Training Approach was not significant, $F(2,24)=.263$, $p>.05$.

5.3.2 A-prime Scores

The mean $A'$ scores for the Mandarin tones in Gen2 on PTDay1 by the listener groups are tabulated in Table 5-5. In general, the $A'$ scores were high, ranging from 0.9 to 1. The $A'$ scores for the four tones by the listeners in the AV groups (0.96-1) were greater than those scored by the listeners in the SIM groups (0.9-0.96).

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>AV Tone 1</th>
<th>AV Tone 2</th>
<th>AV Tone 3</th>
<th>AV Tone 4</th>
<th>SIM Tone 1</th>
<th>SIM Tone 2</th>
<th>SIM Tone 3</th>
<th>SIM Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9735</td>
<td>0.9757</td>
<td>0.9786</td>
<td>0.9759</td>
<td>0.9431</td>
<td>0.9285</td>
<td>0.8982</td>
<td>0.9331</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG</td>
<td>0.9583</td>
<td>0.9774</td>
<td>0.9750</td>
<td>0.9693</td>
<td>0.9243</td>
<td>0.9475</td>
<td>0.9419</td>
<td>0.9200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPN</td>
<td>0.9986</td>
<td>0.9861</td>
<td>0.9806</td>
<td>0.9986</td>
<td>0.9561</td>
<td>0.9518</td>
<td>0.9253</td>
<td>0.9550</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Listeners' $A'$ scores were submitted to a 3-way mixed-design ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone as the within-subjects factors. Only the Training Approach effect was found to be significant, $F(1,24)=38.220$, $p<.001$. The effects of L1 Group and of Tone, and the interactions of L1 Group x Tone, of Training Approach x Tone, of L1 Group x Training Approach, and of L1 Group x Training Approach x Tone were all non-significant ($p>.05$). The results suggest that the listeners in the AV groups
obtained higher mean A' scores for the tones than their counterparts in the SIM groups.

5.3.3 Tonal Confusions

Table 5-6 presents a summary of the listener groups' responses in Gen2 test on PTDay1. Again, the patterns for the listeners' tonal errors were similar to those observed in the previous sections.

Table 5-6. Confusion Matrices for the Responses in Gen2 on PTDay1 Performed by the Listener Groups.
Figure 5-8. Tonal confusions in Gen 2 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.

Figure 5-9. Tonal confusions in Gen 2 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.
Their tonal errors were submitted to a 3-way mixed-design ANOVA to examine the effects of the three factors (L1 Group, Training Approach and Tone Pair). The results showed significant effects of L1 Group, \( F(2,24)=3.363, p=.05 \), of Training Approach, \( F(1,24)=28.795, p<.001 \), and of Tone Pair, \( F(11,264)=14.514, p<.001 \). Post-hoc Tukey tests revealed that the mean error of Cantonese listeners was significantly greater than that of the Japanese listeners \( (p=.05) \). The interaction of Training Approach x Tone Pair was also significant, \( F(11,264)=9.547, p<.001 \). However, the other interactions (L1 Group x Tone Pair, L1 Group x Training Approach, and L1 Group x Training Approach x Tone Pair) were not significant \( (ps>.05) \). These findings suggest that the tonal errors made by the trainees in the AV groups were significantly fewer than those made by the trainees in the SIM groups (see Figures 5-8 and 5-9).

5.4 Gen 3 on PTDay1: Nonspeech (Filtered Speech)

5.4.1 Percent Correct Scores

Figure 5-10 illustrates the mean percent correct scores for Gen3 (Nonspeech) on PTDay1 in the AV (97-99%) and in the SIM (86-92%) groups. The scores of the listeners in the AV groups were consistently higher than those of the listeners in the SIM groups (7-11% difference): CAN (AV: 96.53% vs. SIM: 85.56%), ENG (AV: 97.36% vs. SIM: 89.44%) and JPN (AV: 98.75% vs. SIM: 91.81%).
A 2-way ANOVA was carried out to examine the effects of L1 Group and Training Approach, and their interaction on listeners' mean percent correct scores. The results showed that the L1 Group effect was significant, $F(2,24)=3.290$, $p=.05$. Post-hoc Tukey tests revealed that the Japanese trainees' mean score was significantly higher than that of the Cantonese trainees ($p<.05$). The effect of Training Approach was significant, $F(1,24)=40.576$, $p<.001$, indicating that the trainees in the AV groups outperformed the listeners in the SIM groups. However, no significance was found in the interaction of L1 Group x Training Approach, $F(2,24)=.806$, $p>.05$.

5.4.2 A-prime Scores

The listeners' mean $A'$ scores for the four Mandarin tones for Gen3 on PTDay1 are tabulated in Table 5-7. The $A'$ scores ranged from 0.92 to 1. The $A'$
scores for the four tones by the listeners in the AV groups (0.98-1) were consistently higher than for those scored by the listeners in the SIM groups (0.92-0.98).

Table 5-7. Mean A′ scores for the Mandarin tones in Gen 3 on PTDay1 for the listener groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Mean A′ scores</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9930</td>
<td>0.9902</td>
<td>0.9870</td>
<td>0.9981</td>
<td>0.9653</td>
<td>0.9536</td>
<td>0.9193</td>
<td>0.9585</td>
<td></td>
</tr>
<tr>
<td>ENG</td>
<td>0.9828</td>
<td>0.9915</td>
<td>0.9934</td>
<td>0.9967</td>
<td>0.9606</td>
<td>0.9606</td>
<td>0.9672</td>
<td>0.9619</td>
<td></td>
</tr>
<tr>
<td>JPN</td>
<td>1.0000</td>
<td>0.9958</td>
<td>0.9889</td>
<td>0.9986</td>
<td>0.9730</td>
<td>0.9752</td>
<td>0.9623</td>
<td>0.9758</td>
<td></td>
</tr>
</tbody>
</table>

A 3-way mixed-design ANOVA was performed to examine the effects of L1 Group, Training Approach, and Tone on the listeners' mean A′ scores. The analysis showed that the effects of Training Approach, $F(1,24)=45.166, p<.001$, and of Tone, $F(3,72)=4.353, p<.01$, were significant. The interaction of L1 Group x Tone was also significant, $F(6,72)=3.012, p<.01$. The effects of L1 Group, and the interactions of Training Approach x Tone, of L1 Group x Training Approach, and of L1 Group x Training Approach x Tone, were all non-significant ($p>0.05$).

For the interaction of L1 Group x Tone, individual ANOVAs were performed for each of the tones to test the effect of L1 Group on their A′ scores. The results indicated that the L1 Group effect was not significant for each of the four individual tones ($p>0.05$). However, it is worth mentioning that the L1 Group effect on Tone 3 was marginally significant, $F(2,27)=2.865, p=0.07$. This finding may be attributed to the fact that the Cantonese listeners in the SIM group had a relatively lower A′ score for Tone 3 ($M=0.91$) than did those of the English
and the Japanese \((M=0.96)\) listeners in the SIM groups. These findings suggest that the listeners in the AV groups had consistently higher mean \(A'\) scores for the four tones than did those in the SIM groups.

5.4.3 Tonal Confusions

Table 5-8 summarizes the listener groups’ responses in Gen3 on PTDay1. The listeners’ tonal error patterns were comparable to those observed in the Posttest1, Gen1, and Gen2.

Table 5-8. Confusion Matrices for the Responses in Gen3 on PTDay1 Performed by the Listener Groups.

<table>
<thead>
<tr>
<th>Target</th>
<th>AV Response</th>
<th>English Response</th>
<th>Japanese Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>T1: 176 3 0 1 180</td>
<td>T4: 180 4 0 180</td>
<td>T4: 179 180 0 0 180</td>
</tr>
<tr>
<td>T2</td>
<td>2 177 1 0 180</td>
<td>2 180 0 0 180</td>
<td>2 180 0 0 180</td>
</tr>
<tr>
<td>T3</td>
<td>0 171 0 180</td>
<td>0 176 0 180</td>
<td>0 172 0 180</td>
</tr>
<tr>
<td>T4</td>
<td>1 0 179 180</td>
<td>0 179 0 180</td>
<td>0 179 0 180</td>
</tr>
<tr>
<td>Total</td>
<td>179 189 172 180 720</td>
<td>169 190 178 183 720</td>
<td>180 189 172 179 720</td>
</tr>
</tbody>
</table>

A 3-way mixed-design ANOVA was carried out to examine the effects of three factors (L1 Group, Training Approach and Tone Pair) on the trainees’ mean tonal errors. The results showed that both the effects of Training Approach, \(F(1,24)=52.528, p<.001\), and of Tone Pair, \(F(11, 264)=29.984, p<.001\), were significant. The interactions of L1 Group x Tone Pair, \(F(22,264)=3.018, p<.001\),
of Training Approach x Tone Pair, $F(11,264)=15.434$, $p<.001$, and of L1 Group x Training Approach x Tone Pair, $F(22,264)=2.519$, $p<.001$, were also significant. However, the effect of L1 Group, and the interaction of L1 Group x Training Approach, failed to reach significance ($ps>.05$).

To further investigate the interaction of L1 Group x Training Approach x Tone Pair, two individual mixed ANOVAs were performed to test the effects of L1 Group and of Training Approach on the mean tonal errors by the listeners in the AV and in the SIM groups. For the AV groups, the results indicated that only the effect of Tone Pair was significant, $F(11, 132)=5.721$, $p<.001$. The effects of L1 Group, and the interaction of L1 Group x Tone Pair were not significant ($ps>.05$) (see Figure 5-11).

![Figure 5-11. Tonal confusions in Gen 3 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category](https://via.placeholder.com/150)

Figure 5-11. Tonal confusions in Gen 3 on PTDay1 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.
For the SIM groups (see Figure 5-12), the results showed that L1 Group effect was not significant \((p>.05)\) The effect of Tone Pair, \(F(11,132)=26.168, p<.001\), was significant, as well as the interaction of L1 Group x Tone Pair, \(F(22,132)=3.105, p<.001\). Individual 1-way ANOVAs were carried out to examine the L1 Group effect for each tone pair. The results indicated that a significant effect of the L1 Group was found in the T2T3 pair, \(F(2,12)=18.532, p<.001\). Post-hoc Tukey tests revealed that the Cantonese listeners \((M=10.4)\) made more errors than did the Japanese \((M=4.6)\) and the English \((M=4.4)\) listeners \((p=.001\) and \(p<.001\), respectively).

![Figure 5-12. Tonal confusions in Gen 3 on PTDay1 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category](image)

Overall, the analyses of the percent correct score, the \(A'\) scores, and the tonal confusions, consistently indicated two patterns of results. First, the trainees' performance was greatly affected by the training approach (AV vs. SIM) they
received during the training. Specifically, the trainees in the AV groups consistently outperformed those in the SIM groups. Second, the effect of L1 Group was not found across the four identification tests, except that the Cantonese listeners in the SIM group significantly made more tonal errors for the T2T3 pair than did the listeners in the English and Japanese listener groups in Gen 3 on PTDay1.
CHAPTER 6: RESULTS: IDENTIFICATION TESTS ON POSTTEST-DAY2 (PTDAY2)

This chapter reports the analyses of the listeners' performance on the four identification tests in the PTDay2 phase. Their performance on each of the identification tests was compared in terms of percent correct scores, A' scores, and tonal confusions. The statistical analyses used for each of the identification tests on PTDay2 (Posttest2, Gen1, Gen2, and Gen3) were the same as those performed for the Pretest (see Chapter 3). The results for Posttest2 will be presented in Section 6.1, and the ones for the three generalization tests (Gen1, Gen2, and Gen3) will be presented in Sections 6.2-6.4, respectively.

6.1 Posttest2 on PTDay2

6.1.1 Percent Correct Scores

Figure 6-1 depicts the listeners' mean percent correct scores for the identification test in Posttest2 on PTDay2. The listeners in the AV groups (90-96%) consistently scored higher than did the listeners in the SIM groups (77-85%). Their score differences ranged from 9 to 16%: CAN (AV: 90.28% vs. SIM: 77.36%), ENG (AV: 96.25% vs. SIM: 80.56%), and JPN (AV: 94.31% vs. SIM: 85.42%).
Training Approach

As in Posttest1 on PTDay1, a parallel 2-way ANOVA on the mean percent correct scores yielded a significant effect of Training Approach, \(F(1,24)=30.711, p<.001\). This indicated that the mean scores of the AV groups were significantly higher than those of the SIM groups. The effect of L1 Group, and the interaction of L1 Group x Training Approach, were not significant (\(ps>.05\)).

6.1.2 A-prime Scores

The listeners' mean A' scores for the four Mandarin tones in Posttest2 are tabulated in Table 6-1. The range for the listeners' mean A' scores was from 0.85 to 0.99. In addition, the A' scores for the four tones by the listeners in the AV groups were higher than those by the listeners in the SIM groups.
Table 6-1. Mean A’ scores for the Mandarin tones in Posttest2 on PTDay2 for the listener groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9631</td>
<td>0.9710</td>
<td>0.9661</td>
<td>0.9728</td>
<td>0.9569</td>
<td>0.9213</td>
<td>0.8592</td>
<td>0.9245</td>
</tr>
<tr>
<td>ENG</td>
<td>0.9845</td>
<td>0.9897</td>
<td>0.9860</td>
<td>0.9892</td>
<td>0.9314</td>
<td>0.9320</td>
<td>0.9445</td>
<td>0.9049</td>
</tr>
<tr>
<td>JPN</td>
<td>0.9889</td>
<td>0.9865</td>
<td>0.9628</td>
<td>0.9949</td>
<td>0.9607</td>
<td>0.9542</td>
<td>0.9389</td>
<td>0.9404</td>
</tr>
</tbody>
</table>

A 3-way mixed-design ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone as the within-subjects factor was carried out. Significant effects of Training Approach, F(1,24)= 35.508, p<.001, and of Tone, F(3,72)=3.624, p<.05, were found, but the L1 Group effect did not reach significance (p>.05). The interactions of L1 Group x Tone, and of L1 Group x Training Approach x Tone were significant, F(6,72)=2.45, p<.05, and F(6,72)=2.866, p<.05, respectively. However, the interactions of L1 Group x Training Approach, and of Tone x Training Approach, were not significant (ps>.05).

To explore the 3-way interaction, two individual 2-way ANOVAs were performed for the AV and the SIM groups, with L1 Group as the between-subjects factor, and with Tone as the within-subjects factor. For the AV groups, no significant effects were found for the L1 Group and the Tone, nor was their interaction (L1 Group x Tone) (ps>.05), indicating that the listeners’ mean A’ scores were comparable.

For the SIM groups, the analysis showed that the L1 Group effect was not significant (p>.05), but that the effect of Tone was significant, F(3,36)= 3.295, p<.05. The interaction of L1 Group x Tone was significant, F(6,36)= 3.100, p<.05.
Individual ANOVAs were conducted to investigate the L1 Group effect on the mean A’ scores for the individual tones. It was found that that the L1 Group effect was significant only for Tone 3, $F(2,12)=5.766, p<.05$. Post-hoc Tukey tests revealed that, among the three SIM groups, the Cantonese listeners’ mean A’ score for Tone 3 ($M=0.8592$) was significantly lower than those of the English ($M=0.9445$) and the Japanese ($M=0.9389$) trainees ($ps<.05$).

**6.1.3 Tonal Confusions**

Table 6-2 presents the listener groups’ responses in Posttest2. The listeners’ tonal confusions were mainly found in three pairs of tones, T1-T4, T2-T3, and T1-T2 (bi-directional). For the T1-T4 pair, trainees in all three language groups tended to misidentify target T4 as T1 more frequently than they misidentified target T1 as T4. For the T2-T3 pair, the trainees more often misidentified target T3 as T2 than they did target T2 as T3. For the T1-T2 pair, the Cantonese trainees tended to misidentify target T2 as T1, but both the English and the Japanese trainees frequently misidentified target T1 as T2 (i.e., the opposite direction). In addition, the confusion in the T1-T2 pair was more frequently found in the matrices for the listeners in the SIM groups than for those in the AV groups. Finally, the trainees in the SIM groups made more errors than did those in the AV groups. As in the identification tests on PTDay1, this pattern was consistent among the three language groups.

As was the case in Section 5.1.3, the tonal errors were submitted to a 3-way mixed-design ANOVA with the same factors. The results yielded significant effects of Training Approach, $F(1,24)=36.825, p<.001$, and of Tone Pair, $F(11,$
264)= 37.459, p<.001. However, the L1 Group effect was not significant (p>.05). The interactions of L1 Group x Tone Pair, $F(22,264)=3.849$, $p<.001$, of Training Approach x Tone Pair, $F(11, 264)=13.772$, $p<.001$, and of L1 Group x Training Approach x Tone Pair, $F(22, 264)=2.450$, $p<.001$, were significant, but the interaction of L1 Group x Training Approach did not reach significance (p>.05).

Table 6-2. Confusion Matrices for the Responses in Posttest2 on PTDay2 Performed by the Listener Groups.

<table>
<thead>
<tr>
<th></th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Target</strong></td>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
</tr>
<tr>
<td>T1</td>
<td>163</td>
<td>171</td>
<td>172</td>
</tr>
<tr>
<td>T2</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T4</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>185</td>
<td>172</td>
<td>172</td>
</tr>
<tr>
<td><strong>SIM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
<td><strong>Response</strong></td>
</tr>
<tr>
<td>T1</td>
<td>170</td>
<td>158</td>
<td>168</td>
</tr>
<tr>
<td>T2</td>
<td>7</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>32</td>
<td>148</td>
</tr>
<tr>
<td>T4</td>
<td>48</td>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>225</td>
<td>222</td>
<td>210</td>
</tr>
</tbody>
</table>

Two individual 2-way mixed-design ANOVAs were carried out to examine the effect of L1 Group and Tone Pair on the mean tonal errors made by the trainees in the AV and the SIM groups. For the three AV groups, the results indicated that the L1 Group effect was not significant (p>.05). The Tone Pair effect was significant, $F(11,132)=8.2$, $p<.001$. The interaction of L1 Group x Tone Pair was also found to be significant, $F(22,132)= 1.606$, $p=.05$. Individual 1-way ANOVAs were performed to examine the effect of L1 Group on the mean tonal
errors for the tone pairs. The L1 Group effect was found to be significant in the T1T2 pair, $F(2, 12)=3.447, p=.05$. Post hoc Tukey tests revealed that the Cantonese trainees' mean tonal error ($M=1.8$) was significantly greater than the mean errors of the English ($M=0$) and the Japanese ($M=0$) trainees ($p<.05$). The results are presented in Figure 6-2.

![Graph showing mean errors for different tone pairs](image)

**Figure 6-2.** Tonal confusions in Posttest2 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.

For the S1M groups, the L1 Group effect was not significant ($p>.05$). However, the Tone Pair effect was significant, $F(11,132)=31.760, p<.001$, as well as the interaction of L1 Group x Tone Pair, $F(22,132)=3.694, p<.001$.

Individual ANOVAs were performed to examine the effect of L1 Group on the mean tonal errors in tone pairs. The results showed that L1 Group effect was significantly found in the T2T3, T2T1, T4T1, and T4T3 pairs, $F_s(2,12)=9.180, 4.136, 3.129, \text{and } 4.571$, respectively, $p<.05$. Post hoc Tukey tests showed that
the Cantonese trainees' mean tonal error ($M=18.2$) for the T2T3 pair was significantly more than those of the English ($M=6.4$) and the Japanese ($M=6$) trainees ($p<.05$). In contrast, the Cantonese listeners' error ($M=0.2$) for the T2T1 pair was significantly fewer than those of the Japanese ($M=2.2$) and English ($M=2$) listeners. The English trainees' mean error ($M=2.4$) for the T4T1 pair was significantly more than that ($M=0.2$) of the Japanese trainees ($p<.05$). The mean error of Japanese trainees for the T4T3 ($M=0.8$) pair was significantly greater than those ($Ms=0$) of the Cantonese and the English trainees ($p<.05$). The results are illustrated in Figure 6-3.

![Figure 6-3](image_url)

*Figure 6-3. Tonal confusions in Posttest2 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.*
6.2 Gen 1 on PTDay2: Novel Speakers

6.2.1 Percent Correct Scores

Figure 6-4 shows the mean percent correct scores for Gen2 on PTDay2 in the AV (96-99%) and in the SIM (76-88%) groups. The scores of the listeners in the AV groups were consistently higher than those of their counterparts in the SIM groups (9-21% difference): CAN (AV: 95.69% vs. SIM: 75.83%), ENG (AV: 97.92% vs. SIM: 88.47%) and JPN (AV: 98.61% vs. SIM: 87.22%).

![Figure 6-4](image)

Figure 6-4. Mean percent correct scores (%) in Gen1 on PTDay2 by the listener in the AV (left) and the SIM (right) groups

As in Section 5.2.1, a 2-way ANOVA was performed using identical factors. The analysis showed a significant effect of L1 Group, $F(2,24)=5.878$, $p<.01$. Post-hoc Tukey tests revealed that both the Japanese and the English trainees' mean scores were significantly higher than the scores of Cantonese trainees ($p<.05$). The effect of Training Approach was significant, $F(1,24)=45.725$, $p<.001$, indicating that the trainees in the AV groups
outperformed the listeners in the SIM groups. However, the interaction of L1 Group x Training Approach failed to reach significance, $F(2,24)=2.541, p>.05$.

### 6.2.2 A-prime Scores

The mean $A'$ scores for the four Mandarin tones in the Gen1 on PTDay2 are tabulated in Table 6-3. In general, the $A'$ scores were high, ranging from 0.87 to 0.99. The $A'$ scores for the four tones by the listeners in the AV groups (0.98-0.99) were consistently higher than for those by the listeners in the SIM groups (0.87-0.98).

**Table 6-3. Mean A’ scores for the Mandarin tones in Gen 1 on PTDay2 for the listener groups.**

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9811</td>
<td>0.9905</td>
<td>0.9887</td>
<td>0.9795</td>
<td>0.9449</td>
<td>0.9268</td>
<td>0.8653</td>
<td>0.9136</td>
</tr>
<tr>
<td>ENG</td>
<td>0.9902</td>
<td>0.9958</td>
<td>0.9953</td>
<td>0.9907</td>
<td>0.9549</td>
<td>0.9650</td>
<td>0.9703</td>
<td>0.9473</td>
</tr>
<tr>
<td>JPN</td>
<td>0.9981</td>
<td>0.9944</td>
<td>0.9912</td>
<td>0.9977</td>
<td>0.9800</td>
<td>0.9514</td>
<td>0.9337</td>
<td>0.9494</td>
</tr>
</tbody>
</table>

The mean $A'$ scores were submitted to a 3-way ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone as the within-subjects factor. The analysis yielded significant effects of L1 Group, $F(2,24)=5.302, p=.01$, of Training Approach, $F(1,24)=40.902, p<.001$, and of Tone, $F(3,72)=4.367, p<.01$. The interactions of L1 Group x Tone, $F(6,72)=2.885, p=.01$, of Training Approach x Tone, $F(3,72)=4.618, p<.01$, and of L1 Group x Training Approach x Tone, $F(6,72)=2.734, p=.01$, were all significant. However, the interaction of L1 Group x Training Approach did not reach significance ($p>.05$).
To investigate the effect of L1 Group x Training Approach x Tone, two individual mixed-design ANOVAs were performed for the AV and the SIM groups. For the AV groups, the results indicated that the effects of L1 Group and of Tone, and the interaction between the two factors, were not significant (p > .05). In contrast, the results for the SIM groups showed significant effects of L1 Group, $F(2,12)=4.106, p < .05$, and of Tone, $F(3,36)=4.836, p < .01$, as well as the interaction L1 Group x Tone, $F(6,36)=2.947, p < .02$. Post-hoc Tukey tests indicated that the mean A’ scores of Cantonese listeners were significantly lower than those of the English listeners ($p = .05$). Individual ANOVAs were also performed to investigate the L1 Group effect on the mean A’ scores for each tone. The results showed that the L1 Group effect was significant in Tone 1 and Tone 3, $F(2,12)=6.044$ and $7.280, ps < .02$. Post-hoc Tukey tests revealed that the Japanese listeners’ A’ score ($M=0.98$) for Tone 1 were significantly greater than for those of the Cantonese ($M=0.9449$) and of the English ($M=0.9549$) listeners ($ps < .05$). Also, the Cantonese listeners’ A’ score ($M=0.8653$) for Tone 3 were significantly lower than those of the Japanese ($M=0.9337$) and the English ($M=0.9703$) listeners ($ps < .05$).

6.2.3 Tonal Confusions

The listener groups’ responses for Gen1 on PTDay2 are summarized in Table 6-4. The patterns of tonal errors for T1T4 and T2T3 were similar to those described in Posttest2. However, tonal errors for the T1T2 pair were mainly made by the English and Japanese listeners. Overall, the listeners in the SIM groups made more tonal errors than did those in the AV groups.
Table 6-4. Confusion Matrices for the Responses in Gen1 on PTDay2 Performed by the Listener Groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Response</strong></td>
<td><strong>Target</strong></td>
<td><strong>T1</strong></td>
<td><strong>T2</strong></td>
</tr>
<tr>
<td>T1</td>
<td>172</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>1</td>
<td>176</td>
<td>3</td>
</tr>
<tr>
<td>T3</td>
<td>1</td>
<td>6</td>
<td>173</td>
</tr>
<tr>
<td>T4</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>186</td>
<td>184</td>
<td>176</td>
</tr>
</tbody>
</table>

| **SIM**  |           |         |          |
| **Response** | **Target** | **T1** | **T2** | **T3** | **T4** | **Total** | **T1** | **T2** | **T3** | **T4** | **Total** | **T1** | **T2** | **T3** | **T4** | **Total** |
| T1       | 163       | 2       | 0       | 15      | 180    |           | 165     | 8       | 0       | 9       | 180    |           | 176     | 0       | 0       | 4       | 180    |
| T2       | 4         | 163     | 13      | 0       | 180    |           | 7       | 165     | 6       | 2       | 180    |           | 6       | 165     | 9       | 0       | 180    |
| T3       | 0         | 0       | 94      | 0       | 180    |           | 0       | 138     | 161     | 1       | 180    |           | 0       | 37      | 138     | 5       | 180    |
| T4       | 54        | 0       | 0       | 126     | 180    |           | 31      | 1       | 0       | 148     | 180    |           | 24      | 6       | 0       | 150     | 180    |
| Total    | 221       | 251     | 107     | 141     | 720    |           | 201     | 192     | 167     | 160     | 720    |           | 206     | 208     | 147     | 159     | 720    |

A 3-way mixed-design ANOVA was carried out to examine the effect of L1 Group, Training Approach, and Tone Pair on the trainees’ mean tonal errors. The results revealed that the effects of L1 Group, $F(2,24)=5.989$, $p<.01$, of Training Approach, $F(1,24)=45.853$, $p<.01$, and of Tone Pair $F(11, 264)=31.234$, $p<.001$, were significant. The interactions of L1 Group x Tone Pair, $F(22,264)=5.023$, $p<.001$, of Training Approach x Tone Pair, $F(11,264)=19.119$, $p<.001$, and of L1 Group x Training Approach x Tone Pair, $F(22,264)=3.638$, $p<.001$, were also significant. However, the interaction of L1 Group x Training Approach were not significant ($p>.05$).

To further examine the interaction of L1 Group x Training Approach x Tone Pair, two individual mixed-design ANOVAs were performed for the AV and the SIM groups. For the AV groups (see Figure 6-5), the analysis revealed
significant effect of Tone Pair, $F(11, 132)=4.613$, $p<.001$. However, neither the effect of L1 Group nor the interaction of L1 Group x Tone Pair was significant ($ps>.05$).

For the SIM groups (see Figure 6-6), the results showed that the effects of L1 Group, $F(2, 12)=4.692$, $p<.05$, and of Tone Pair, $F(11, 132)=27.316$, $p<.001$, were significant, as was the interaction of L1 Group x Tone Pair, $F(22, 132)=4.654$, $p<.001$. Individual 1-way ANOVAs were carried out to examine the L1 Group effect for each tone pair. The results indicated that the L1 Group effect was significant in the T2T3 and T2T1 pairs, $Fs(2, 12)=8.509$, and 6.50, $ps \leq .01$. Post-hoc Tukey tests revealed that the Cantonese listeners made greater T2T3 errors ($M=17.2$) than did the Japanese ($M=7.4$) and the English ($M=3.6$) listeners.
and that the English listeners made more T2T1 errors ($M=1.6$) than did the Cantonese ($M=0.4$) and the Japanese ($M=0$) listeners ($ps<.05$).

![Figure 6-6. Tonal confusions in Gen 1 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.]

6.3 Gen 2 on PTDay2: Novel Stimuli

6.3.1 Percent Correct Scores

Figure 6-7 depicts the mean percent correct scores for Gen2 on PTDay2 in the AV (88-95%) and in the SIM (74-83%) groups. It can readily be seen that the listeners in the AV groups consistently scored higher than did their counterparts in the SIM groups (12-14% difference): CAN (AV: 88.06% and SIM: 73.75%), ENG (AV: 92.92% and SIM: 80.97%), and JPN (AV: 94.86% and SIM: 83.33%).

The listeners' mean scores were submitted to a 2-way ANOVA with two between-subjects factors, L1 Group and Training Approach. The L1 Group effect
was found to be significant, \( F(2,24)=4.625, p=0.02 \). Post-hoc Tukey tests revealed that the Japanese trainees' mean score was significantly higher than that of the Cantonese trainees \( (p<0.02) \). The effect of Training Approach was also significant, \( F(1,24)=30.489, p<0.001 \), indicating that the trainees in the AV groups outperformed the listeners in the SIM groups. However, the interaction of L1 Group x Training Approach was not significant, \( F(2,24)=.144, p>.05 \).

![Figure 6-7. Mean percent correct scores (%) in Gen2 on PTDay2 by the listener in the AV (left) and the SIM (right) groups](image)

### 6.3.2 A'-prime Scores

Table 6-5 presents a summary of the mean A’ scores of the listeners for the four Mandarin tones for Gen2 on PTDay2. As in previous analyses, the A’ scores were high, ranging from 0.83 to 0.99. The A’ scores for the four tones by the listeners in the AV groups (0.95–0.99) were consistently higher than those by the trainees in the SIM groups (0.83-0.97).
To test the effects of L1 Group, Training Approach, and Tone on the mean A’ scores, a 3-way mixed-design ANOVA was performed. The analysis revealed significant effects of L1 Group, $F(2,24)=3.866, p<.05$, of Training Approach, $F(1,24)=26.855, p<.001$, and of Tone, $F(3,72)=13.919, p<.001$. The interactions of L1 Group x Tone, $F(6,72)=3.763, p<.01$, of Training Approach x Tone, $F(3,72)=4.953, p<.01$, and of L1 Group x Training Approach x Tone, $F(6,72)=2.741, p<.02$, were all significant. However, the interaction of L1 Group x Training Approach was not significant ($p>.05$).

To further study the effect of L1 Group x Training Approach x Tone, two individual mixed ANOVAs were performed for the AV and the SIM groups. For the AV groups, the results indicated a significant effect of L1 Group, $F(2,12)=3.613, p=.05$. Post-hoc Tukey tests further indicated that the Cantonese listeners’ mean A’ score was significantly lower than that of the Japanese listeners ($p=.05$). The Tone effect was also significant, $F(3,36)=3.285, p<.05$. However, the interaction between the two factors was not significant ($p>.05$).

For the SIM groups, the results showed that the L1 Group effect was not significant ($p>.05$). However, significant effects of Tone, $F(2,36)=11.688, p<.001$, and the interaction of L1 Group x Tone, $F(6,36)=3.849, p<.01$, were found.

### Table 6-5. Mean A’ scores for the Mandarin tones in Gen 2on PTDay2 for the listener groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
<th>Mean A’ scores</th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>0.9551</td>
<td>0.9617</td>
<td>0.9516</td>
<td>0.9631</td>
<td>0.9422</td>
<td>0.9009</td>
<td>0.8326</td>
<td>0.9245</td>
<td></td>
</tr>
<tr>
<td>ENG</td>
<td>0.9721</td>
<td>0.9769</td>
<td>0.9708</td>
<td>0.9834</td>
<td>0.9364</td>
<td>0.9460</td>
<td>0.9235</td>
<td>0.9185</td>
<td></td>
</tr>
<tr>
<td>JPN</td>
<td>0.9981</td>
<td>0.9676</td>
<td>0.9644</td>
<td>0.9972</td>
<td>0.9660</td>
<td>0.9193</td>
<td>0.9079</td>
<td>0.9548</td>
<td></td>
</tr>
</tbody>
</table>
Individual ANOVAs were performed to investigate the L1 Group effect on the mean $A'$ scores for each of the four tones. The results showed that the L1 Group effect was significant only in Tone 3, $F(2,12)=4.425, p<.05$. Post-hoc Tukey tests indicated that the Cantonese listeners’ mean $A'$ score ($M=0.8326$) was significantly lower than those of the English ($M=0.9235$) and the Japanese ($M=0.9079$) listeners ($p>.05$).

6.3.3 Tonal Confusions

Table 6-6 summarizes the listener groups’ responses in Gen2 on PTDay2. The listeners’ tonal error patterns were comparable to those observed in the Posttest2 and Gen1.

<table>
<thead>
<tr>
<th>Target</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
</tr>
<tr>
<td>T1</td>
<td>156 2 0 22 180</td>
<td>164 5 0 11 180</td>
<td>180 0 0 0 180</td>
</tr>
<tr>
<td>T2</td>
<td>3 165 12 0 180</td>
<td>3 172 5 0 180</td>
<td>2 165 13 0 180</td>
</tr>
<tr>
<td>T3</td>
<td>0 29 151 0 180</td>
<td>0 19 161 0 180</td>
<td>0 20 160 0 180</td>
</tr>
<tr>
<td>T4</td>
<td>18 0 0 162 180</td>
<td>8 0 0 172 180</td>
<td>2 0 0 178 180</td>
</tr>
<tr>
<td>Total</td>
<td>177 196 163 184 720</td>
<td>175 196 166 183 720</td>
<td>184 185 173 178 720</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
<td>T1 T2 T3 T4 Total</td>
</tr>
<tr>
<td>T1</td>
<td>161 4 1 14 180</td>
<td>157 10 0 13 180</td>
<td>166 11 2 1 180</td>
</tr>
<tr>
<td>T2</td>
<td>9 156 15 0 180</td>
<td>5 164 9 2 180</td>
<td>6 155 18 1 180</td>
</tr>
<tr>
<td>T3</td>
<td>3 97 80 0 180</td>
<td>1 47 131 1 180</td>
<td>0 49 129 2 180</td>
</tr>
<tr>
<td>T4</td>
<td>40 4 2 134 180</td>
<td>46 1 2 131 180</td>
<td>19 8 3 150 180</td>
</tr>
<tr>
<td>Total</td>
<td>213 261 98 148 720</td>
<td>209 222 142 147 720</td>
<td>191 223 152 154 720</td>
</tr>
</tbody>
</table>
The trainees' tonal errors were submitted to a 3-way ANOVA to examine the effect of L1 Group, of Training Approach and of Tone Pair. The analysis yielded significant effects of L1 Group, $F(2,24)=4.628$, $p<.05$, of Training Approach, $F(1,24)=30.496$, $p<.001$, and of Tone Pair, $F(11, 264)=67.111$, $p<.001$. In addition, the interactions of L1 Group x Tone Pair, $F(22,264)=5.231$, $p<.001$, of Training Approach x Tone Pair, $F(11,264)=16.860$, $p<.001$, and of L1 Group x Training Approach x Tone Pair, $F(22,264)=2.193$, $p<.01$, were significant. However, the interaction of L1 Group x Training Approach failed to reach significance ($p>.05$).

To explore the interaction of L1 Group x Training Approach x Tone Pair, two individual mixed ANOVAs were performed for the AV and the SIM groups. The effects of L1 Group and of Tone Pair on the listeners' tonal errors were examined. For the AV groups (see Figure 6-8), the results indicated that the effects of L1 Group, $F(2,12)=4.015$, $p<.05$, and Tone Pair, $F(11,132)=15.074$, $p<.001$, were significant. The interaction of L1 Group x Tone Pair was also significant, $F(22,132)=1.716$, $p<.05$. Individual 1-way ANOVAs were carried out to examine the L1 Group effect for each tone pair. The results indicated that the L1 Group effect was significant in the T4T1 pair, $F(2,12)=9.68$, $p=.01$. Post-hoc Tukey tests revealed that the Japanese listeners ($M=0$) made fewer T4T1 errors than did the Cantonese ($M=4.4$) and the English ($M=2.2$) listeners ($ps<.05$).
Figure 6-8. Tonal confusions in Gen 2 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.

For the SIM groups (see Figure 6-9), the results showed that the L1 Group effect was not significant \((p>.05)\). However, the effects of Tone Pair, \(F(11, 132)=56.044, p<.001\), and the interaction of L1 Group x Tone Pair, \(F(22, 132)=4.754, p<.001\), were significant. Individual 1-way ANOVAs were carried out to examine the L1 Group effect on each tone pair. The results indicated that the L1 Group effect was significant in the T1T4, T4T1, and T2T3 pairs, \(Fs(2,12)=3.967, 4.361, \text{ and } 5.950\), respectively, \(ps<.05\). Post-hoc Tukey tests revealed that the Cantonese listeners \((M=19.4)\) made more T2T3 errors than did the Japanese \((M=9.8)\) and the English \((M=9.4)\) listeners \(ps<.05\). The English and the Cantonese listeners made more T4T1 errors than did the Japanese listeners \(ps<.05\). The English listeners made more T1T4 errors than did the Japanese listeners \(p<.05\).
Figure 6-9. Tonal confusions in Gen 2 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category.

6.4 Gen 3 on PTDay2: Nonspeech (Filtered Speech)

6.4.1 Percent Correct Scores

Figure 6-10 shows the mean percent correct scores for Gen3 on PTDay2 in the AV (96-98%) and in the SIM (83-91%) groups. As can be seen in the figure, the listeners in the AV groups consistently had higher scores than did the listeners in the SIM groups (7-13% difference): CAN (AV: 96.25% and SIM: 83.06%), ENG (AV: 97.78% and SIM: 88.06%), and JPN (AV: 98.33% and SIM: 91.39%).

A 2-way ANOVA was carried out to examine the effects of L1 Group and of Training Approach on listeners’ mean percent correct scores. The results showed a significant effect of L1 Group, $F(2,24)=6.444, p<.01$. Post-hoc Tukey tests revealed that the Japanese trainees’ mean score was significantly higher
than that of the Cantonese trainees ($p < .01$). The effect of Training Approach was also significant, $F(1,24)=69.191$, $p<.001$, showing that the trainees in the AV groups outperformed the listeners in the SIM groups. However, the interaction of L1 Group x Training Approach was not significant, $F(2,24)=2.279$, $p>.05$.

![Figure 6-10. Mean percent correct scores (%) in Gen3 on PTDay2 by the listener in the AV (left) and the SIM (right) groups](image)

6.4.2 A-prime Scores

The mean $A'$ scores for the four Mandarin tones for Gen3 on PTDay2 are presented in Table 6-7. As in previous analyses, the $A'$ scores were high, ranging from 0.89 to 1. The $A'$ scores for the four tones by the listeners in the AV groups (0.98-1) were consistently greater than for those by the listeners in the SIM groups (0.89-0.98).
Table 6-7. Mean $A'$ scores for the Mandarin tones in Gen 3 on PTDay2 for the listener groups.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>AV</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tone 1</td>
<td>Tone 2</td>
<td>Tone 3</td>
<td>Tone 4</td>
<td>Tone 1</td>
<td>Tone 2</td>
<td>Tone 3</td>
</tr>
<tr>
<td>CAN</td>
<td>0.9835</td>
<td>0.9876</td>
<td>0.9856</td>
<td>0.9925</td>
<td>0.9687</td>
<td>0.9432</td>
<td>0.8923</td>
</tr>
<tr>
<td>ENG</td>
<td>0.9892</td>
<td>0.9963</td>
<td>0.9911</td>
<td>0.9934</td>
<td>0.9550</td>
<td>0.9637</td>
<td>0.9673</td>
</tr>
<tr>
<td>JPN</td>
<td>1.0000</td>
<td>0.9926</td>
<td>0.9870</td>
<td>0.9981</td>
<td>0.9741</td>
<td>0.9654</td>
<td>0.9589</td>
</tr>
</tbody>
</table>

The mean $A'$ scores were submitted to a 3-way ANOVA with L1 Group and Training Approach as the between-subjects factors, and with Tone as the within-subjects factor. The analysis revealed significant effects of L1 Group, $F(2,24)=5.653$, $p=.01$, of Training Approach, $F(1,24)=63.537$, $p<.001$, as well as Tone, $F(3,72)=7.159$, $p<.001$. The interactions of L1 Group x Tone, $F(6,72)=6.229$, $p<.001$, of Training Approach x Tone, $F(3,72)=3.564$, $p<.02$, and of L1 Group x Training Approach x Tone, $F(6,72)=6.097$, $p<.001$, were also significant. However, the interaction of L1 Group x Training Approach was not significant ($p>.05$).

To explore the 3-way interaction, two individual mixed-design ANOVAs were performed for the AV and the SIM groups. For the AV groups, the results indicated that the effect of Tone was significant, $F(3,36)=2.935$, $p<.05$. However, neither the effect of L1 Group nor the interaction between the two factors was significant ($p>.05$). In contrast, the analysis for the SIM groups revealed significant effects of L1 Group, $F(2,12)=4.115$, $p<.05$, and of Tone, $F(3,36)=5.637$, $p<.01$, as well as the interaction of L1 Group x Tone Pair, $F(6,36)=6.620$, $p<.001$. Post-hoc Tukey tests revealed that the Cantonese listeners' mean $A'$ score was significantly lower than that of the Japanese listeners.
listeners (p=.05). Individual ANOVAs were also performed to investigate the L1 Group effect on the mean A' scores for each tone. The results showed that the significant effect of L1 Group was only found in Tone 3, $F(2,12)=16.920$, $p<.001$. Post-hoc Tukey tests indicated that the Cantonese listeners' mean A' score ($M=0.8923$) was significantly lower than those of the English ($M=0.9673$) and the Japanese ($M=0.9589$) listeners ($ps<.05$).

### 6.4.3 Tonal Confusions

Table 6-8 summaries the listener groups’ responses for Gen3 on PTDay2, showing that the tonal errors were compatible with those in Gen 2 on PTDay2.

<table>
<thead>
<tr>
<th>L1 Group</th>
<th>Cantonese</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
</tr>
<tr>
<td>T1</td>
<td>170</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>2</td>
<td>176</td>
<td>1</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>10</td>
<td>170</td>
</tr>
<tr>
<td>T4</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>175</td>
<td>190</td>
<td>171</td>
</tr>
</tbody>
</table>

| **SIM** | | | |
| Target | T1 | T2 | T3 | T4 | Total | T1 | T2 | T3 | T4 | Total | T1 | T2 | T3 | T4 | Total |
| T1 | 169 | 4 | 0 | 7 | 180 | 161 | 11 | 0 | 8 | 180 | 168 | 10 | 1 | 1 | 180 |
| T2 | 6 | 168 | 6 | 0 | 180 | 3 | 168 | 6 | 3 | 180 | 3 | 169 | 7 | 1 | 180 |
| T3 | 1 | 71 | 108 | 0 | 180 | 0 | 21 | 159 | 0 | 180 | 0 | 26 | 154 | 0 | 180 |
| T4 | 24 | 1 | 2 | 153 | 180 | 29 | 5 | 0 | 146 | 180 | 13 | 0 | 0 | 167 | 180 |
| **Total** | 200 | 244 | 116 | 160 | 720 | 193 | 205 | 165 | 157 | 720 | 184 | 205 | 162 | 169 | 720 |

To examine the effects of L1 Group, of Training Approach and of Tone Pair on the listeners' tonal errors, a 3-way mixed-design ANOVA was carried out.
The analysis revealed significant effects of L1 Group, $F(2,24)=6.451$, $p<.01$, of Training Approach, $F(1,24)=69.199$, $p<.001$, and of Tone Pair, $F(11,264)=43.950$, $p<.001$. The interactions of L1 Group x Tone Pair, $F(22,264)=6.145$, $p<.001$, of Training Approach x Tone Pair, $F(11,264)=21.445$, $p<.001$, and of L1 Group x Training Approach x Tone Pair, $F(22,264)=5.050$, $p<.001$, were significant. The interaction of L1 Group x Training Approach was not significant ($p>.05$).

To study the interaction of L1 Group x Training Approach x Tone Pair, two individual mixed-design ANOVAs were performed for the AV and the SIM groups. The effects of L1 Group and Tone Pair on the mean tonal errors were tested. For the AV groups (see Figure 6-11), the results indicated that only the effect of Tone Pair was significant, $F(11,132)=9.410$, $p<.001$. Neither the effects of L1 Group nor the interaction of L1 Group x Tone Pair was significant ($p>.05$).

![Figure 6-11](image)

*Figure 6-11.* Tonal confusions in Gen 3 on PTDay2 by the listeners in the AV groups. Each of the tone pairs on the x-axis represents a response-target category.
For the SIM groups (see Figure 6-12), the results showed that the effects of L1 Group, $F(2,12)=4.787, p<.05$, and of Tone Pair $F(11,132)=36.458, p<.001$, were significant, as was the interaction of L1 Group x Tone Pair, $F(22,132)=6.352, p<.001$. Individual 1-way ANOVAs were carried out to examine the L1 Group effect for each tone pair. The results indicated that the L1 Group effect was significant in the T2T3 and T2T4 pairs, $F$s$(2,12)=18.127$, and 21.000, $ps <.001$. Post-hoc Tukey tests revealed that the Cantonese listeners ($M=14.2$) made more T2T3 errors than did the Japanese ($M=5.2$) and the English ($M=4.2$) listeners ($ps<.05$). In addition, the English listeners ($M=1$) made more T2T4 errors than did the Cantonese ($M=0.2$) and the Japanese ($M=0$) listeners ($ps<.05$).

*Figure 6-12. Tonal confusions in Gen 3 on PTDay2 by the listeners in the SIM groups. Each of the tone pairs on the x-axis represents a response-target category*
In all, the analyses of the trainees' performance on the four identification tests on PTDay2 indicated two consistent patterns. First, the training approaches (AV vs. SIM) significantly altered the trainees' performance: the listeners in the AV groups outperformed those in the SIM groups. The findings were similar to those on PTDay1 (Chapter 5). Second, unlike the findings in the PTDay1 phase, significant effects of the L1 Group, along with some of its significant interactions with other factors, were found in the data for the percent correct scores, $A'$ scores, and tonal confusions. Further, the L1 Group effect was more often found in the performance of the trainees in the SIM groups than that of the listeners in the AV groups.
CHAPTER 7: OVERALL TRAINING IMPROVEMENT & GENERALIZATION

This chapter presents the results of several comparisons of the listeners’ percent correct scores for the identification tasks tested at different times. Section 7.1 reports the results of the comparison among the listeners’ performance in the Pretest, and in the two posttests (Posttest1 and Posttest2) on PTDay1 and on PTDay2, respectively. Section 7.2 describes the comparisons of the listeners’ scores in the three generalization tests (Gen1, Gen2, and Gen3) at the two posttest times, PTDay1 and PTDay2 (Sections 7.2.1 – 7.2.3, respectively).

7.1 Training Improvement: Pretest-Posttest1-Posttest2

The purpose of this comparison was to examine the listeners’ improvements and their abilities to retain their knowledge of Mandarin tones after training.

Figure 7-1 shows the mean percent correct scores obtained by the listeners in the Pretest, Posttest1, and Posttest2. For the listeners of all language groups, their performance in each of Posttest1 and Posttest2 was consistently better than their performance in the Pretest, regardless of which training approach they received during the Training phase. Also, the mean scores in Posttest2 were slightly lower than those in Posttest1. It should also be mentioned
that the listeners in the AV groups outperformed their counterparts in the SIM groups.

Figure 7-1. Mean percent correct scores (%) in the Pretest, Posttest1, and Posttest2 by the listeners in the AV (left) and SIM (right) groups

The mean percent correct scores in the three identification tests were submitted to a 3-way mixed-design ANOVA with L1 Group (CAN, ENG, and JPN) and Training Approach (AV and SIM) as the between-subjects factors, and with Test (Pretest, Posttest1 and Posttest2) as the within-subjects factor. The results showed that both the effects of Training Approach $F(1,24)=13.773, p=.001$, and of Test, $F(2,48)=315.910, p<.001$, were significant. With respect to the differences among the three identification tests, Post-hoc Tukey tests revealed that the listeners' mean score in Posttest1 was significantly greater than their scores in the Pretest ($p<.001$) and in Posttest2 ($p<.05$), respectively. In addition, the mean score in Posttest2 was significantly higher than that in the Pretest ($p<.001$). However, the L1 Group effect failed to reach significance ($p>.05$). The
interactions of Test x L1 Group, $F(4,48)= 4.572$, $p<.01$, and of Test x Training Approach, $F(2,48)=10.082$, $p<.001$, were significant, but not the interaction of L1 Group x Training Approach ($p>.05$).

For the interaction of Test x L1 Group, three separate 1-way ANOVAs were carried out to examine the L1 Group effect in each of the three tests. The results indicated that a significant effect of L1 Group was found only in the Pretest, $F(2,27)= 255.169$, $p<.05$. Post-hoc Tukey tests further revealed that the English listeners' mean score was significantly lower than that of the Cantonese and the Japanese listeners ($ps<.05$; see Section 3.1).

To further study interaction of Test x Training Approach, two separate ANOVAs were conducted for the AV and the SIM groups. For the three AV groups, the Test effect was found to be significant, $F(2,28)=162.913$, $p<.001$. Post-hoc LSD tests revealed that the listeners' scores in Posttest1 and Posttest2 were significantly greater than their scores in the Pretest ($ps<.001$), but the score difference between Posttest1 and Posttest2 was not significant ($p>.05$). For the three SIM groups, the results also indicated that the Test effect was significant, $F(2,28)=100.570$, $p<.001$. Post-hoc LSD tests revealed that the listeners' scores in Posttest1 and Posttest2 were significantly higher than their score in the Pretest ($ps<.001$). Further, their mean score in Posttest1 was significantly higher than that in Posttest2 ($p<.05$). These findings suggest that the AV training approach assisted the learners to learn the Mandarin tones better than did the Simple training approach.
7.2 Tests of Retention of Generalization

The purpose of the following comparisons was to investigate whether the listeners’ abilities to generalize their knowledge of Mandarin tones were retained a month after the Training phase. The results of the Gen1, Gen2 and Gen3 tests will be presented in Sections 7.2.1-7.2.3.

7.2.1 Gen 1: PTDay1 vs. PTDay2

Figure 7-2 shows the mean percent correct scores in Gen1 on PTDay1 and on PTDay2. As can be seen in the figure, the listeners’ performance in Gen1 on PTDay1 was similar to that on PTDay2. In addition, the listeners in the AV groups obtained higher scores than did their counterparts in the SIM groups.

![Figure 7-2. Mean percent correct scores (%) in Gen 1 (Novel speakers) on PTDay1 and PTDay2 by the listeners in the AV (left) and the SIM (right) groups](image)

The means scores in Gen1 were submitted to a 3-way mixed-design ANOVA with L1 Group (CAN, ENG, and JPN) and Training Approach (AV and...
SIM) as the between-subjects factors, and with Test Time (PTDay1 and PTDay2) as the within-subjects factor. The analysis yielded significant effect of Training Approach, $F(1,24)=41.139$, $p<.001$, suggesting that listeners in the AV groups performed better than did those in the SIM groups. In addition, the L1 Group effect was significant, $F(2,24)=4.310$, $p<.05$. Post-hoc Tukey tests revealed that the mean score of the Cantonese listeners was significantly lower than that of the Japanese listeners ($p<.05$), and that the difference in the mean score between the Cantonese and the English listeners was marginal ($p=.06$). Also, it was found that the interaction of Test Time x Training Approach was significant, $F(1,24)=5.227$, $p<.05$. Two separate t-tests were performed for the AV and the SIM groups, showing that the listeners in the AV groups had similar Gen1 scores on PTDay1 and on PTDay2, ($p>.05$), and that listeners in the SIM groups had a significantly higher Gen1 scores on PTDay1 than on PTDay2, $t(14)=2.386$, $p<.05$. However, the effect of Test Time was not significant ($p>.05$); nor were other interactions ($ps>.05$).

7.2.2 Gen 2: PTDay1 vs. PTDay2

Figure 7-3 depicts the mean percent correct scores in Gen2 on PTDay1 and PTDay2, showing that the listeners' scores in Gen2 on PTDay1 were slightly higher than their scores in Gen2 on PTDay2. The pattern was observed in the performance of the listeners in both the AV and the SIM groups. In addition, the listeners in the AV groups had higher scores than their counterparts in the SIM groups.
The data were submitted to a mixed ANOVA to examine the effects of L1 Group (CAN, ENG, and JPN), Training Approach (AV and SIM), and Test Time (PTDay1 and PTDay2) on the mean percent correct scores in Gen2. The effect of Training Approach was found to be significant, $F(1,24)=48.033, p<.001$, suggesting that the listeners in the AV groups performed better than did the listeners in the SIM groups. Moreover, the L1 Group effect was significant, $F(2,24)=5.242, p<.05$. Post-hoc Tukey tests revealed that the Cantonese listeners' mean score was significantly lower than the Japanese listeners' mean score ($p<.05$), and the score difference between the other groups were not significant ($ps>.05$). Nevertheless, the effects of Test Time, and of other interactions were not significant ($ps>.05$).
7.2.3 Gen 3: PTDay1 vs. PTDay2

Figure 7-4 illustrates the mean percent correct scores in Gen3 on PTDay1 and PTDay2. It can be clearly seen that the listeners' scores in Gen3 on PTDay1 were similar to their scores in the Gen3 on PTDay2. In addition, the scores of the listeners in the AV groups were higher than those of the listeners in the SIM groups.

A 3-way mixed ANOVA was performed to examine the effects of L1 Group (CAN, ENG, and JPN), Training Approach (AV and SIM), and with Test Time (PTDay1 and PTDay2) on the mean percent correct scores in Gen3. Only the effect of Training Approach was significant, $F(1,24)=65.609, p<.001$, indicating that the listeners in the AV groups performed better than did the listeners in the SIM groups. Also, the L1 Group effect was significant, $F(2,24)=5.728, p<.001$. 

Figure 7-4. Mean percent correct scores (%) in Gen 3 (Nonspeech) on PTDay1 and PTDay2 by the listeners in the AV (left) and the SIM (right) groups. 

![Graph showing mean percent correct scores (%) in Gen 3 (Nonspeech) on PTDay1 and PTDay2 by the listeners in the AV (left) and the SIM (right) groups.](image-url)
Post-hoc Tukey tests revealed that the Cantonese listeners' mean score was significantly lower than the mean score of the Japanese listeners ($p<.01$). However, other pairwise comparisons failed to reach significance ($p>.05$). The effects of Test Time, and of other interactions were not significant ($p>.05$).

In sum, the above comparisons of the listeners' performance in the identification tests on PTDay1 and on PTDay2 indicated several patterns. With respect to training improvements, the listeners' posttest scores were significantly higher than their Pretest scores, suggesting that the training improved their perceptual performance. The findings also indicated that the AV training approach assisted the listeners to retain the Mandarin tones knowledge better than the Simple feedback approach, because the significant score difference between the two posttests was only found in the performance of the listeners in the SIM groups. In addition, the L1 Group effect was significant in the Pretest, but not in either of the posttests.

With respect to the retention of generalization, two patterns from the listeners' mean scores for the three generalization tests (Gen1, Gen2, and Gen3) on PTDay1 and on PTDay2 were noted. First, the effect of Test Time was not significant in each of the generalization tests, suggesting that the listeners were still able to make the generalizations of the Mandarin tones a month after training. Second, the significant effects of L1 Group and Training Approach were observed across the identification tests. These results indicated that the two
factors (the L1 prosodic backgrounds and the training approach) influenced the listeners' performance in the PTDay1 and PTDay2 phases.\textsuperscript{34}

\textsuperscript{34} The effects of the L1 prosodic backgrounds and the training approach were reported in earlier chapters (Chapters 5 and 6).
CHAPTER 8: DISCUSSION AND CONCLUSION

This chapter will discuss the results of the experiments. Section 8.1 evaluates the three hypotheses stated in Chapter 1: the L1 prosodic effect, the training approach, and the issue of Cantonese listeners and AV training (8.1.1 – 8.1.3). Section 8.2 addresses several issues arising from the findings of the present cross-language study. Section 8.3 provides the contributions and the limitations of this study, and the future research direction. Finally, Section 8.4 presents the conclusion for the current study.

8.1 The Three Hypotheses of the Present Study

8.1.1 Hypothesis 1: L1 Prosodic Effect

In Chapter 1, it was hypothesized that listeners' native (L1) prosodic systems would influence their performance when learning a new L2 tonal system (i.e., Mandarin, in this study). Specifically, due to the differences between their L1 prosodic systems, it was expected that the Japanese listeners would have less problems in learning the four Mandarin tones, while the Cantonese listeners would have greater difficulty differentiating the two pairs of Mandarin tones (Tone 1-Tone 4 and Tone 2-Tone 3). The English listeners' performance was predicted to fall between the other two language groups, because English does not have a system for lexical tone and does not use pitch variations at the word level.
8.1.1.1 The L1 Prosodic Effect in the Four Phases

The results of the present study revealed that the expected pattern of the listeners' learning performance, as described in the above section, varied from phase to phase (i.e., Pretest, Training, PTDay1, and PTDay2). Before training, the analysis of the listeners' performance (in terms of percent correct scores, $A'$ scores, and tonal confusions) in the Pretest phase (Chapter 3) indicated that the listeners' performance partially showed the expected pattern for Hypothesis 1. Specifically, the Cantonese and the Japanese listeners (native speakers of tone languages) outperformed the English listeners (native speakers of a non-tone language). The results indicated that the L1 effect did affect the listeners' perceptions of non-native contrasts in the Pretest, even though the participants' variations in performance were controlled for by using stringent selection criteria during the recruitment process. Recall that all listeners were naïve listeners of Mandarin, and had not had any musical training prior to or during their participation in the experiment. Thus, the performance differences found among the listener groups in the Pretest phase indeed indicated that their linguistic experiences of tones played a role when listeners identified non-native tonal contrasts (see Lee et al., 1996). Moreover, the findings of the pretest of the present study were in agreement with those of Wayland & Guion (2004).

During the Training phase (Chapter 4), the performance differences among the three language groups were not obvious. However, among the three SIM groups, the Cantonese listeners took significantly more training sessions for the T2-T3 pair than did the English and the Japanese listeners.
In the PTDay1 phase (2 days after training), the L1 prosodic effects on the listeners’ performance were less evident on PTDay1, but some degree of L1 effect was still observed. For example, the Japanese listeners outperformed the Cantonese listeners (JPN > CAN) in Gen2 and Gen3. The limited L1 effect may be due to the fact that the listeners still had very recent short-term memories about the Mandarin tones on PTDay1. The PTDay1 phase took place only one or two days after the Training phase, and all the trainees reached a passing score of 80% or higher.

In the PTDay2 phase (a month after the PTDay1), the L1 effect on the trainees' performance was observed. The performance of the Cantonese listeners was poorer than that of the Japanese listeners. The performance of the English listeners appeared to be intermediate between that of the Cantonese and the Japanese listeners. In some instances, the English listeners, like the Japanese listeners, outperformed the Cantonese listeners (e.g., in the three generalization tests). However, they occasionally performed poorer than did the Japanese listeners. For example, the English listeners, similar to the Cantonese listeners, misidentified target T4 as T1 in Gen2 more than did the Japanese listeners. Taking these together, the expected pattern of performance of the three listener groups was found in the PTDay2 phase.

In sum, it was found that the performance of the Cantonese listeners was in general poorer than that of the Japanese and English listeners in the posttest phases upon the completion of the perceptual training. Therefore, it implies that linguistics experience is not sufficient to account for the findings of the present
study. In fact, as expected, the L1 phonological system of the trainees, in terms of phonemic status and F0 patterns, play a more important role in L2 tone learning.

8.1.1.2 PAM and Lexical Tone Assimilations

The analysis of the tonal confusion patterns indicates that listeners’ L1 prosodic effects were most apparent on several tone pairs. Consistently, all Cantonese, English, and Japanese listeners exhibited greater confusion for three tone pairs: T1-T4, T2-T3, and T1-T2. These patterns were not only found across the identification tasks that were administered at different times, but also in the training sessions. However, the question remains as to why the Cantonese trainees showed significantly more confusions in the T1-T4 and T2-T3 pairs, but fewer confusions in the T1-T2 pair, than did the English and the Japanese trainees.

The Perceptual Assimilation Model (PAM: Best, 1994, 1995, Best et al., 1988; Best et al., 2001; Best & Tyler, 2006) can provide a theoretical basis for explaining the discrepancies in the listener groups’ performance for tonal confusions. The following sections will discuss the possible tonal assimilations for each language in the framework of the PAM.

Japanese

With respect to tonal assimilations from Mandarin to Japanese, three PAM assimilation types could be observed: Two Category (TC), Uncategorized- Uncategorized (UU), and Uncategorized-Categorized (UC) Assimilations. TC

---

35 The asymmetrical patterns of tonal perception will be discussed in Section 8.2.2.1.
was seen in the cases where Mandarin Tone 2 (mid rising) and Tone 4 (high falling) assimilated to the Japanese LH and HL pitch-accent patterns, respectively. Mandarin Tone 1 and Tone 3 could be interpreted as a UU pair, because they did not assimilate to any tone or pitch-accent pattern in the Japanese prosodic system. According to the PAM, listeners' perceptions of uncategorized sounds are less influenced by their L1 systems, but this depends on how well those listeners perceive the similarities of the non-native contrasts. Perhaps, Tone 1 and Tone 3 have some phonetic properties (e.g., vowel duration and F0 patterns) that are relatively easy to perceive. Tone 1 involves high pitch with limited pitch movement; Tone 3 involves low pitch in the centre portion and is produced with longer vowel duration (e.g., Ho, 1976; Howie, 1976). Therefore, the pairs, T1-T4, T2-T3, and T1-T2, formed three UC pairs. According to the PAM prediction, listeners should be able to discriminate the non-native sounds of an UC pair quite well. The results of the present study were consistent with the prediction, because the Japanese listeners showed significantly fewer tonal errors for the tone pairs in the identification tests on PTDay2. Thus, the Japanese pitch-accent system (HL & LH) may facilitate the perception of Mandarin tones.

**English**

With respect to tonal assimilation from Mandarin to English, a stress-accent language, the issue is more complicated. According to Hallé et al., (2004), there are two possible interpretations: lexical tones could be perceived either as uncategorized speech categories or as nonspeech. On the one hand, English

---

36 The production of Tone 3 involving creaky voice quality (e.g., Belotel-Grenié & Grenié, 2004) is not always found in the speech of native speakers of Mandarin (see Gårding et al., 1986)
employs tone/pitch contours at the sentential levels (as intonation) to indicate the nature of utterances. For example, a falling pitch pattern signals a statement, and a rising pattern indicates a question. Thus, English listeners may perceive Mandarin tones as uncategorized speech categories, because English does have tone contours at the phrasal and sentential levels (see Section 1.6.2.4). On the other hand, English listeners may perceive lexical tones as "nonlinguistic melodic variations" (Hallé et al., 2004, p. 416). Lexical tones are not part of the phonological system of English, and are not perceived as phonemic categories (Hallé et al., 2004). From this perspective, tones are thus nonspeech melodies that are Non-Assimilable (NA) in the framework of the PAM.

For the present study, it is assumed that the English listeners perceive Mandarin tones as nonspeech, the same as melodies. There are three reasons for positing this assumption. First, even though English utterances are produced with intonation contours, these contours are not "basic prosodic units bearing contrastive linguistic significance" (Hallé et al., 2004, p. 417), and do not have any phonemic (or tonemic) status in English. These characteristics are different from those of lexical tones, which have their own phonemic status in the languages in which they occur. Second, since pitch contour is an important fundamental feature found in both nonspeech melodies and human speech (Plantinga & Trainor, 2005; Stevens & Kellers, 2001; Stevens, Kellers, & Tyler, 2004), the English listeners might perceive tone contours as nonspeech melodic contours. Third, recent studies on speech perception/processing employing Positron Emission Tomography (PET) or Functional Magnetic Resonance
Imaging (fMRI) have showed that native English speakers differ from native Mandarin speakers in processing Mandarin (lexical) tones (e.g., Gandour, 2006; Klein, Zatorre, Milner, & Zhao, 2001). For instance, when processing Mandarin tones, native Mandarin speakers showed more activation in the left hemisphere. For the native English speakers, however, neural activities were confined to the right hemisphere (Gandour, 2006; Klein et al, 2001). It has been documented that the right hemisphere is lateralized for musical processing (e.g., Tillmann, Janata, & Bharucha, 2003).

Since the English listeners perceived lexical tones as nonspeech melodies, tones are thus NA in the framework of the PAM. The discrimination performance will be expected to range from good to excellent, depending on how well the English listeners can perceive the tones. The performance of the English listeners in this study evidenced that they perceived the tones relatively well.\(^{37}\) In some instances, the English listeners performed as good as the Japanese listeners; however, in other instances, their performance was poorer than that of their Japanese counterparts. For example, in Gen1 on PTDay2, the Japanese listeners had a greater sensitivity (i.e., higher A′ scores) for Tone1 than did the English and the Cantonese listeners. Also, the English listeners sometimes outperformed the Cantonese listeners (e.g., in the generalization tests on PTDay2), but again they occasionally performed as poor as the Cantonese listeners relative to the Japanese listeners (e.g., misidentifying target T4 as T1 in Gen2 on PTDay2). Therefore, it may be interpreted that the English stress-

\[^{37}\text{Perhaps, some phonetic properties of certain tones may be relatively easy to perceive (see Section 8.1.1), or some English listeners might have related the Mandarin tones to some melodic patterns that helped them to distinguish the tones in question.}\]
accent system neither facilitates nor hinders the perception of non-native tonal contrasts when native English speakers are learning Mandarin tones.

**Cantonese**

With respect to tonal assimilations from Mandarin to Cantonese, three types of assimilations predicted by the PAM were identified: **Single Category (SC), Category Goodness (CG), and Two Category (TC) Assimilations.** SC could be seen in the case that Mandarin Tone 1 (high level [55]) and Tone 4 (high falling [51]) assimilated to the Cantonese Tone 1 (high level), which has two allotones, high level [55] and high falling [53]. The consequence was that the Cantonese listeners frequently misidentified these Mandarin tones (Tone 4 with Tone 1).

CG was found in the case that Mandarin Tone 2 (mid rising [35]) and Tone 3 (falling rising [214]) assimilated to Cantonese Tone 2 (high rising [35]). In this case, the Mandarin Tone 2 assimilated to the Cantonese Tone 2 better than did the Mandarin Tone 3, and therefore the Cantonese listeners had a tendency to select Mandarin Tone 2 as the better match most of the time. There are two possible reasons for this tendency. First, the Cantonese high rising tone (Tone 2) is phonetically similar to Mandarin Tone 2 in terms of the F0 patterns (F0 height and shape). Both are described with the same tone letters [35] in the literature (Cantonese: Hashimoto, 1972; Yip, 2002, and Mandarin: Howie, 1976). The Cantonese Tone 2 is produced with a falling and rising pattern (Bauer & Benedict, 1997; So, 1999), and so is the Mandarin Tone 2 (Fon & Chiang, 1999; see Chapter 1, Section 1.6.2.1). Second, the Cantonese tonal system does not
have a tone that is similar to Mandarin Tone 3 [214]. When a Cantonese speaker listens to a Mandarin Tone 3, the best candidate (the closest L1 tone) will be a Cantonese Tone 2, because Mandarin Tone 2 and Tone 3 share considerable similarities in their pitch contours (e.g., the dip and the rising portions).

Finally, the TC assimilation was found in the case that Mandarin Tone 1 [55] and Tone 2 [35] assimilated to the Cantonese Tone 1 [55] and Tone 2 [35], respectively. Since the two Mandarin tones perfectly matched two separate Cantonese tones (in terms of the tone letters), as evidenced by the results of the present study, the Cantonese listeners did not experience too much difficulty in identifying Mandarin Tone 1 and Tone 2 when compared to the English and the Japanese listeners. Taken together, it can be inferred that the Cantonese tonal system hinders the learning of Mandarin tones, especially the tones in the T1-T4 and T2-T3 pairs.

Given that the Cantonese listeners exhibited three types of assimilations (SC, CG, and TC) as predicted by the PAM, these findings raise an additional question. Will the observed assimilation types found among the Cantonese listeners be consistent with other predictions of the PAM regarding the gradient of discrimination levels for the three assimilation types (TC > CG > SC)? 38 This gradient pattern in fact can be seen in the Training phase (Chapter 4, Section 4.1). The mean number of training sessions for the three tone pairs (T1-T4, T2-T3, and T1-T2) taken by the Cantonese listeners in the SIM group provided evidence for the gradient pattern. The Cantonese listeners on average took 1 session only for T1-T2 (TC), 2.2 sessions for T2-T3 (CG), and 2.8 sessions for

38 see Chapter 1, Section 1.3.2
T1-T4 (SC). These training duration patterns imply that the more difficult tone pair contrasts required the learners to spend more training time to learn to discriminate the contrasts among the tones, and to pass the predetermined criterion (80%) for each tone block during training. Thus, the results suggest that the Cantonese listeners in the SIM groups did demonstrate a gradient of tonal discrimination for the three assimilation types, TC (T1-T2) > CG (T2-T3) > SC (T1-T4). In other words, the T1-T2 pair is the least difficult. The T2T3 pair is moderately difficult and the T1-T4 pair is the most difficult for the Cantonese listeners to learn to discriminate the contrasts between the tones.

There are possible explanations for why the gradient pattern of tonal discrimination for the assimilation types, TC (T1-T2) > CG (T2-T3) > SC (T1-T4), was not found both (i) in the other Cantonese (AV) group during the training phase, and (ii) in the Cantonese listeners’ performance in the Pretest and the PTDay1 and PTDay2 phases. The Cantonese listeners in the AV group did not show the gradient pattern. This may be due to the effect of having received the AV feedback training approach, whose components provided the Cantonese listeners with tri-channel inputs to assist with their learning of Mandarin tones. Compared to the Simple feedback approach, the components of the AV feedback functioned as external tools to assist learning. Consequently, the gradient pattern could be obscured. For the Pretest, this may be partially attributable to the reason that some Cantonese listeners might have some degree of confusion when responding to the trials based on the four response choices, and that they might have made some random-guesses (judging by the comments given by
some participants). Therefore, the pattern was also not found in the performance of the Cantonese listeners in the Pretest phase. The reason for not finding the gradient pattern in the two posttest phases (PTDay1 and PTDay2) seems to be obvious. One may expect that perceptual modifications of the listeners' knowledge about non-native tonal contrasts would take place after training, because L1 effects on the perception of non-native contrasts are not "absolute or permanent" (Best, 1994, p. 173). Thus, the gradient pattern was not observed in the Cantonese listeners' performance in the identification tests (the posttest and the three generalization tests) on PTDay1 and PTDay2.

In sum, the effect of L1 prosodic backgrounds was evidenced in the present study. Its effect is more related to the constraints of the phonological systems of native languages, as proposed by the PAM. The findings clearly indicate that the performance of the group of Cantonese listeners was the poorest among the three groups after training. The results, unlike those found in Wayland & Guion (2004), did not show that linguistic experience in using tones facilitates L2 tone learning. In contrast, the results of the present study supported the first hypothesis that the Cantonese listeners' performance was constrained by their phonological system (e.g., the phonemic status and the F0 patterns of certain tones in the system). Similar L1 constraints were not observed in either the Japanese or the English listeners, as can be evidenced by their better performance on the two posttest days, especially on PTDay2.

Regarding L2 tone learning at the initial stage, the findings of the present study indicate that the listeners' L1 effects were present at every phase of the
experiment. This could be interpreted as the process during which the L1 effects on L2 tone learning changed over time. Before training (i.e., in the Pretest phase), the L1 effects were observed because the influence of training was absent. During training, the L1 influences interacted with the training effect. However, listeners’ pre-existing tone patterns might have facilitated, hindered, or had no effect on their learning of the non-native (Mandarin) tones, which in turn affected the training duration they needed for specific tone pairs (see Chapter 4). In the immediate posttest phase (PTDay1), the listeners’ L1 effect seemed to be less noticeable, and this might be due to the presence of a training effect. Yet, the difficulties of the Cantonese listeners, caused by their existing L1 (Cantonese) tones, could still be observed. Finally, in the PTDay2 phase, the L1 effects emerged again. Especially, the perceptual performance of the Cantonese listeners in the SIM groups evidenced the effects. Thus, the findings imply that L1 effects are still present after training. The greater the L2 knowledge gained from training, the less influence the L1 systems exerted on the learners’ performance.

8.1.2 Hypothesis 2: The Training Approach

The second hypothesis related to the effects of the two training approaches. It was hypothesized that the AV feedback (consisting of three components: Audio, Animated Graph, and Text) training would be more effective in assisting the listeners to learn the Mandarin tones than would the Simple (SIM) feedback training. Further, the AV training would promote better and more robust long-term modifications to non-native contrasts than would the SIM feedback training approach.
The results of the analyses support the second hypothesis. In Chapter 7, the results showed that both the AV and the SIM training approaches assisted listeners to have significantly better performance for tonal identifications in the two posttests (Posttest1 and Posttest2) than in the Pretest (see Section 7.1). More importantly, the results of the present study also evidenced that the listeners trained with the AV feedback training approach did outperform those trained with the SIM feedback training approach. This pattern was consistently found in the two posttests and the three generalization tests on PTDay1 and PTDay2 (see Chapters 5 and 6). In addition, the listeners in the three AV groups required significantly fewer training sessions than did those in the three SIM groups (overall means 6.47 vs. 8.2; see Chapters 4). Therefore, the AV feedback training approach was more effective in assisting tone learning than was the SIM feedback training approach.

With the help of this kind of tri-channel AV feedback approach during training, the learners could more successfully transfer the training knowledge to other contexts, such as to stimuli produced by novel speakers (Gen1), and to novel stimuli produced by old speakers (Gen2). The findings were comparable to those reported in previous studies (e.g., Logan et al., 1991; Wang et al., 1999). Further, the learners in the present study were also able to transfer the training knowledge to nonspeech stimuli (Gen3) that required these learners to apply the perceptual skill to "normalize" the tone patterns (e.g., Ching, 1984; Fourcin, 1972). This may also imply that the learners have established the tonal patterns in their mental representations.
Further, the AV feedback training approach used in this study also promoted longer-lasting mental representations than did the SIM feedback training approach (see Chapters 5 - 7). Particularly, two pieces of evidence can be used for the illustration purpose. First, in Chapter 7, in the comparison of pretest-posttests performance, the listeners in the AV groups did not have a significant score difference between the two posttests (Posttest1 and Posttest2). However, the same pattern was not observed for the listeners in the SIM groups. Their Posttest1 score was significantly higher than the Posttest2 score. Similarly, the same patterns were also observed for Gen1.

Taken together, there are several implications of the above findings. First, using the AV feedback type during training could successfully shift the listeners' attention to relevant tonal perceptual cues. This would help the trainees (i) to master tonal contrasts faster and better, and (ii) to visually notice the spatial relationship between the given tones. As a result, the learners' knowledge of tonal contrasts for non-native tones could be improved, which in turn would assist learners to establish a new tonal system. Second, training with the three AV components (Audio, Animated Graph, and Text) could be an effective method for assisting tone learning. These components not only shorten the training time but also help naïve listeners to establish the tonal categories (or system) at the phonemic level. Further, the effect of this kind of AV training is comparable to the methodologies employed in previous studies, such as the Fading technique used in Jamieson & Morosan (1986, 1989) and the high-variability training procedures used in Logan et al. (1991) and Lively et al. (1993). Thus, this type of AV training
with tri-channel input could be considered as one of the appropriate procedures for perceptual training, as well as for the training of other tonal systems (e.g., Cantonese and Thai tonal systems).

8.1.3 Hypothesis 3: Cantonese Listeners and AV Training

The third hypothesis concerned which language group would benefit the most from the AV training approach. It was hypothesized that the Cantonese listeners would benefit more than the other two groups, because their L1 tonal system led them to have greater difficulties when learning Mandarin tones (see Section 8.1.1).

The results of the statistical analyses in the present study did not show any significant effect for the L1 Group x Training Approach interaction, suggesting that the AV training approach did not facilitate the learning of Mandarin tones differently by one versus another language group. The non-significant interaction may be due to the small number of subjects participating in the present study.

However, it must be mentioned that the findings of the current study indicate that the AV training approach appeared to be more helpful to assist the Cantonese listeners than the English and the Japanese listeners in learning the Mandarin tones. Three pieces of evidence can be seen from the performance differences between the Cantonese listeners in the AV and those in the SIM groups compared to those differences in each of the English and the Japanese listeners in the AV and the SIM groups.
First, with reference to the language groups’ mean score differences between the AV and SIM groups (AV-SIM_{score\,diff}) in the eight identification tests on PTDay1 and PTDay2 (see Table 8-1; data was presented in Chapters 5 and 6), it can be seen that the AV-SIM_{score\,diff} of the Cantonese listeners were frequently greater than those of the Japanese listeners and the English listeners for the three generalization tests (Gen1, Gen2, and Gen3). However, in Posttest1 and Posttest2, the English listeners’ mean score differences were slightly higher than those of the Cantonese listeners, which in turn were higher than those of the Japanese listeners.

To test the L1 Group effect on the trainee groups’ mean score differences, the data was submitted to a 1-way ANOVA. The analysis revealed a significant L1 Group effect, $F(2,21)=458.455, p<.001$. Post-hoc LSD tests further indicated that the mean score difference of the Cantonese listeners was significantly larger than that of the Japanese ($p<.001$) and the English listeners ($p<.05$), and that the mean difference between the Japanese and the English listeners was not significant ($p>.05$). Thus, the results supported the observed pattern: CAN > (JPN or ENG), implying that the AV training approach had a greater facilitative effect on the Cantonese listeners’ learning of Mandarin tones.
Second, the performance of the Cantonese listeners in the AV group was always comparable to that of the Japanese and the English listeners in the AV groups (i.e., the L1 Group effect on their performance was generally not found, see the results of Chapters 5 and 6). In contrast, the Cantonese listeners in the SIM group had significantly poorer performance in the identification tests on both PTDay1 and PTDay2 than did the Japanese and/or the English listeners in the SIM groups (see Chapters 5 and 6).

Third, in terms of training duration (in Chapter 4), the mean number of training sessions for the Cantonese listeners in the AV group was comparable to those for the English and the Japanese listeners in the AV groups. In contrast, the mean number of training sessions for the Cantonese listeners in the SIM group was greater than those for the English and the Japanese listeners in the SIM groups, especially the training duration for the T2-T3 pair.

If the above facts are taken into consideration, it can be interpreted that the AV training was particularly helpful to the Cantonese learners, because their performance was noticeably poorer when they were trained with the SIM feedback approach. The findings imply that the tri-channel input AV feedback provides an effective and efficient training method to assist trainees to learn to distinguish tones. In addition, they suggest that learning difficulties caused by one’s L1 system (like Cantonese in this study) could be reduced by employing an appropriate training method.
8.2 Other Issues

The discussion in the following sections will focus on several issues: the cross-language characteristics of lexical tone perception found in the present study (Section 8.2.1), the influence of Japanese prosodic background on Japanese listeners' tonal perceptions (Section 8.2.2), and the effect of short-term training on the long-term perceptual modification of tones (Section 8.2.3).

8.2.1 The Cross-Language Characteristics of Lexical Tone Perception

8.2.1.1 Asymmetrical Patterns in Tonal Perception

Although this study examined on the effects of the listeners' L1 tonal systems and the AV training approach on the learning of Mandarin tones, the results of a series of analyses reflected that there is an asymmetrical perceptual pattern among Mandarin tones. Further, the pattern seemed to be language-independent. Specifically, some tones were more easily misidentified as other tones. As mentioned in Section 8.1.1, the tones in the pairs, T1-T4, T2-T3, and T1-T2, were easily confused with their counterparts, and generally caused the listeners to spend more time to learn them than other tone pairs during the Training phase. It is possible that the tones in each pair share considerable phonetic characteristics (e.g., pitch contours, and pitch height for the tonal onset and/or the offset). For the T1-T4 pair, both tones begin with a similar high pitch level. For the T2-T3 pair, both have a dip and a rising pattern. For the T1-T2 pair, both pitch contours end at a high pitch level. In fact, previous studies

---

39 Some studies have been suggesting that a high degree of phonetic similarities between two non-native segments could increase the perceptual difficulty for the listeners (e.g., Polka, 1991, 1992).
(Kirilloff, 1969; Miracle, 1989; Shen, 1989) have reported that non-native language learners have great difficulties in producing and perceiving different lexical tones. In particular, the tone pairs T2-T3 and T1-T4 are the most problematic. In contrast, the tones in the T1-T3, T2-T4, and T3-T4 pairs appeared to be less confusable with their counterparts. It may relate to the fact that the phonetic properties (e.g., F0 patterns including the F0 height and direction) of the tones in each pair are dissimilar. For example, the F0 patterns for T1 (high level) and T3 (falling rising) are different, at least in terms of the F0 patterns (level vs. falling rising) and duration (T3 is the longer than T1; e.g., Howie, 1976). Thus, the results suggest that the phonetic characteristics (similar vs. dissimilar) of lexical tones also exert an effect on the listeners' perception of Mandarin tones. Tones with more dissimilar features will be easier to be discerned and learned, whereas tones that share similar features are likely to cause more perceptual and learning difficulties for the trainees.

8.2.1.2 The Perception of Filtered Speech for Lexical Tones

Another cross-language pattern was observed in the perception of the nonspeech stimuli (filtered speech) in Gen3. The listeners in the present study tended to perform Gen3 slightly better than the other generalization tests (Gen1 and Gen2) which used human speech tokens (see Section 7.2). Similar findings have been reported in Burnham & Brooker (2002), in which they found that English speaking non-musicians discriminated the Thai tones in filtered speech better than in the human speech condition. Therefore, the results of Gen3 were in agreement with their findings. Further, this study found that the pattern is not
just limited to native English speakers, but it is also found in native speakers of
tone languages as well. Non-musically trained native speakers of tone languages
(Cantonese and Japanese) and a non-tone language (English) exhibited the
same pattern: lexical tones in a filtered speech condition were identified even
more accurately than in other testing conditions that used human speech tokens.
Perhaps, by minimizing the segmental information (Ching, 1984), the listeners
were able to shift their attention more easily to the suprasegmental features or
properties contained in the filtered speech stimuli.

8.2.2 The Influence of Japanese Prosodic Background on Tonal Perception

The results of the present study also provided evidence that the Japanese
prosodic background greatly influenced the Japanese listeners' perceptions of
the Mandarin tones. For example, the Japanese listeners had difficulties in
identifying Mandarin Tone 2 (mid rising) and Tone 4 (high falling) in the Pretest
phase. Their difficulties may be partly due to their existing pitch-accent patterns.
Although their pitch contours are similar to those of Tone 2 and Tone 4, it was
possible that the Japanese listeners had not yet established the mappings of the
pitch patterns between the two languages in the Pretest. In addition, the
Japanese listeners also had a tendency to respond to the stimuli with Tone 2
(mid rising) in the Pretest. This can be explained by the fact that two-mora words
in Japanese (e.g., ame) are likely to be pronounced with a rising pattern. For a
two-mora word, the word may be produced with either a rising or a falling pitch-
accent pattern (i.e., LH or HL). However, when the same word is unaccented,
Japanese speakers will also produce it with a rising pattern (see Cutler & Otake, 1999; Fujisaki, Ohno, & Tomita, 1996; Nagano-Madsen, 2003). Thus, it is not surprising that the Japanese listeners frequently selected Mandarin Tone 2 as their responses.

In addition, that the Japanese listeners could learn the Tone 1 and Tone 4 distinction better than the English and the Cantonese listeners might be related to the fact that some phonological properties in their L1 language facilitated their learning of L2 (Mandarin) tones. It is well known that native Japanese speakers use a prosodic feature, vowel length contrasts, at the phonemic level (e.g., Hirata, 2004). Indeed, they seem to have a tendency to make use of this phonemic feature when learning a L2 language (see Ingram & Park, 1997). Thus, in the present study, the Japanese trainees might have made use of their phonemic length-contrasting features, or duration cues, to help them to distinguish the Mandarin tones in the T1-T4 pair, as Tone 4 is always the shortest tone in Mandarin (see Section 1.6.2.1). This explanation seems to be verifiable by the observed pattern that the Japanese exhibited a very limited amount of tonal confusions for the tones in the T1-T4 pair.

8.2.3 The Effect of Short-term Training on Long-term Tonal Modification

As discussed in Chapter 1, a concern about the effect of short-term training (e.g., Pisoni et al., 1982) was that this kind of training might not have a long-term effect on learning. The results of the present study provide supportive evidence that short-term training, even for a single day, could be effective and generate a long-term effect on learning. This can be evidenced by the listeners’ performance
on Posttest2 and the three generalization tests on PTDay2, which took place a month after the PTDay1 phase (see Sections 6.4-6.6). Further, the trainees were able to transfer the knowledge gained from training to stimuli spoken by novel speakers (Gen1), novel stimuli spoken by old speakers (Gen2), and even nonspeech/filtered stimuli (Gen3), suggesting that the L2 tone patterns were established in the mental representations of the trainees. Thus, this study demonstrated that adults' lexical tone perceptions could be modified by short-term laboratory training.

8.3 Contributions, Limitations, and Future Research

Overall, this research has contributed to a better understanding of the influences of L1 prosodic backgrounds and of AV training as a novel training methodology on L2 tone learning. Specifically, the results enrich our knowledge of the role of L1 prosodic (tonal) backgrounds in learning non-native suprasegmentals, including new tonal systems. Since the performance differences among the three language groups can be accounted for by the PAM, this study thus contributes to the development and extension of the PAM at the suprasegmental level. In addition, because the model was originally developed to deal with non-native speech perception by naïve listeners, this study will also help to extend the PAM's principles on issues of L2 perceptual learning.

In addition, Strange (1995, p. 41) has suggested that studies that examine the "transfer of training provide important practical information about optimal training procedures". The findings of the present study contributed to a better understanding of optimal training procedures by examining (i) the effect of a
novel AV training approach for training lexical tones in a laboratory setting, and (ii) its potential long-term modification of trainees' perceptions of tones or the establishment of a new lexical system. It was found that the effect of the (tri-channel) AV feedback training method used in the current study was comparable to the methodologies used in previous studies, and that the AV training approach assisted naïve learners to learn the non-native tonal contrasts, or lexical tones faster, better, and with greater retention than the SIM training method. Further, the results also indicated that short-term training can have a long-term effect on the modification of learners' knowledge of the lexical tones. Thus, the findings suggest that the (tri-channel) AV training approach can be extended and applied to other tone language systems as well (e.g., Cantonese and Thai).

However, several limitations of the present study have to be identified. First, as already mentioned in Chapter 2, due to the difficulty in recruiting naïve participants who had to meet the stringent criteria, only 30 individuals were recruited as research participants among almost 300 perspective participants. As a result, only five participants each from the three language groups were assigned to the AV and the SIM conditions. It would have been more desirable if more participants were involved. Second, since the present study only focused on perceptual training, it would have been more complete if listeners' productions before and after training had also been analyzed. These analyses would have provided a better understanding of the effects of L1 prosodic backgrounds and of the AV training on the trainees' performance.
In the future, several areas could be considered when conducting training studies on learning lexical tones. For example, learners’ tonal productions should also be examined (e.g., measurements of some phonetic properties – F0 and vowel duration). This will enable researchers to explore the relationship between tonal perception and production, and to study the constraints of L1 tonal production on L2 tonal production. Moreover, native speakers of other tonal languages (e.g., Thai and Vietnamese) should be recruited so as to have a deeper and broader understanding of cross-language perception and production of L2 tones.

8.4 Conclusion

The present study focused on both the theoretical and the methodological issues of the perceptual learning of lexical tones by training three groups of naïve listeners (Cantonese, Japanese, and English) to learn the tonal system of Mandarin. Theoretically, it examined the effects of naïve listeners’ L1 prosodic backgrounds on their learning of Mandarin tones. Methodologically, this study examined the effectiveness of a novel tri-channel Audio-Visual (AV) training approach on naïve listeners’ L2 tone learning.

With respect to the L1 prosodic effect, the results indicated that L1 effects were evident in the Pretest and PTDay2 phases, but were less noticeable in the Training and PTDay1 phases. The L1 effect on L2 tone learning was attributable to the specific characteristics of a learner’s prosodic system (i.e., the phonemic status of tones and the F0 patterns) rather than to the amount of linguistic experience of tonality use. The analyses of percent correct scores, A’ scores,
and tonal confusions revealed that the Cantonese listeners' performance was the worst while Japanese listeners' performance was the best. The English listeners' performance appears to be intermediate between the other two groups. The differences could be attributed to the fact that their L1 prosodic backgrounds play an important role in the process of learning a new tonal system. On the basis of these findings, it can be interpreted that the Cantonese tonal system hinders the learning of Mandarin tones, while the Japanese pitch-accent system facilitates the learning process. However, the English stress-accent system neither helps nor hinders tone learning. These findings have a number of implications. First, listeners' tonal experiences affect the establishment of a new tonal system, but their L1 experiences do not necessarily facilitate their learning of new tones. The phonemic status or phonetic characteristics of their existing L1 tones can affect the learning process, or the establishment of a new tonal system. Second, human speech perceptual mapping is not restricted to segments, but is also applicable to suprasegmentals. Third, the findings provide supporting evidence for the assumption that the Perceptual Assimilation Model (PAM) can be extended to "suprasegmental tiers" (Hallé et al. 2004).

With respect to the effectiveness of AV training, the results indicated that the AV training approach assisted listeners better to improve their performance than did the SIM training approach. This tendency was found in every experimental phase except for the Pretest phase of the present study. The findings imply that AV training is helpful in assisting lexical tone learning, because it can shorten the training time, and can help trainees to better perceive
the Mandarin tonal contrasts. More importantly, it assists trainees to establish long-term tonal patterns.

Lastly, the overall findings suggest that the Cantonese listeners will benefit the most if they are trained with the AV approach. Although the Cantonese tonal system causes its native speakers' difficulties in learning Mandarin tones, the three components of the AV approach can be particularly helpful to their learning of the tonal contrasts.
APPENDIX: QUESTIONNAIRE

Language Background Questionnaire

Participant Code: ________  Age: ______  Date: ______

Q1. What is your first language? __________________________

Q2. Were you born in Canada?  YES / NO  (go to Q3)

Q3. Where were you born and raised? __________________________

Q4. When did you arrive in Canada?  _____ / _____ / _____
   (DD / MM / YY)

Q5. a) What is your mother’s native language? __________________________
    b) What is your father’s native language?

Q6. What other languages have you studied in school or university?  (please list)

<table>
<thead>
<tr>
<th>Language</th>
<th>How proficient are you? (circle one)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>very</td>
</tr>
<tr>
<td></td>
<td>very</td>
</tr>
</tbody>
</table>
REFERENCES


