EFFECTS OF PRODUCTION TRAINING AND PERCEPTION TRAINING ON LEXICAL TONE PERCEPTION – ARE THE EFFECTS DOMAIN GENERAL OR DOMAIN SPECIFIC?

By

SHUANG LU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013
To my family and friends
ACKNOWLEDGMENTS

The journey of pursuing my PhD in the United States is truly unforgettable. I cannot imagine having accomplished this journey without the great support and help from the people listed below.

First of all, I would like to express my deepest gratitude to my advisor Dr. Edith Kaan, who taught me how to do research, how to write, how to apply for grants and everything I know about academia. I am thankful to Dr. Kaan for all her patience, intelligence, caring and guidance during the past few years. Working with her has been my greatest pleasure! Likewise, I am thankful to the co-chair of my committee Dr. Ratree Wayland for her unlimited assistance and support. Dr. Wayland’s enlightening insights have enriched the analyses of my dissertation and have always encouraged me to think beyond the obvious.

Dr. Wind Cowles and Dr. Brenda Smith are the other two members of my committee, and both have provided invaluable support and help for me. Dr. Cowles introduced me into the fascinating world of psycholinguistics. It has been a great pleasure to work as her teaching assistant for the last four years. Many thanks are directed to Dr. Smith for her insightful suggestions on my dissertation and for helping me recruit musicians for my dissertation research.

Besides, I would like to thank all of my professors in the Department of Linguistics at the University of Florida for teaching me all the wonderful classes and providing me great help and advice. Special thanks go to Dr. Paula Golombek, Dr. Caroline Wiltshire, Dr. Jason Rothman and Dr. Brent Henderson. Appreciation is also given to our office manager Kelli Granade for taking care of the grant money for my dissertation research and for all her helps throughout the past few years.
I would also like to thank all my fellow graduate students in the Linguistic Department and all my wonderful lab mates in the Brain and Language Laboratory for making graduate school an incredible memorable experience for me. I am thankful for all the research assistants who helped me with the preparation and conduction of my dissertation research, including Joe Kirkham, Eric Holgate, Carlie Overfelt and Kayla Johnson. I am also thankful to all the participants who participated in my experiments.

My family and friends have always been the strongest supporters for me. I am very grateful for all my friends in China and in the U.S., who always bring me true happiness and strength. I would also like to express my greatest appreciation to my parents, PP and my other family members. This journey would not have been possible without their love, support and encouragement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgments</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>9</td>
</tr>
<tr>
<td>List of Figures</td>
<td>10</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>13</td>
</tr>
<tr>
<td>Abstract</td>
<td>15</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>1 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>Project Motivation</td>
<td>17</td>
</tr>
<tr>
<td>Background</td>
<td>18</td>
</tr>
<tr>
<td>Views of the Relation between Production and Perception</td>
<td>18</td>
</tr>
<tr>
<td>Previous Training Studies</td>
<td>20</td>
</tr>
<tr>
<td>Perception training</td>
<td>20</td>
</tr>
<tr>
<td>Production training</td>
<td>21</td>
</tr>
<tr>
<td>Comparison between perception training and production training</td>
<td>23</td>
</tr>
<tr>
<td>Tones in Music and Language</td>
<td>25</td>
</tr>
<tr>
<td>Tone Perception and Production by Non-native Speakers</td>
<td>26</td>
</tr>
<tr>
<td>Effect of Music Background on Lexical Tone Perception</td>
<td>27</td>
</tr>
<tr>
<td>Effects of Laboratory Training on Lexical Tone Perception</td>
<td>29</td>
</tr>
<tr>
<td>Perception training</td>
<td>29</td>
</tr>
<tr>
<td>Production training</td>
<td>30</td>
</tr>
<tr>
<td>Interim Summary</td>
<td>30</td>
</tr>
<tr>
<td>Event-related Potentials (ERPs)</td>
<td>31</td>
</tr>
<tr>
<td>ERP Components that are Relevant to Auditory Perception</td>
<td>32</td>
</tr>
<tr>
<td>The Current Study</td>
<td>33</td>
</tr>
<tr>
<td>Research Questions</td>
<td>33</td>
</tr>
<tr>
<td>Structure of the Dissertation</td>
<td>33</td>
</tr>
<tr>
<td>Intellectual Merit</td>
<td>36</td>
</tr>
<tr>
<td>Broader Implications</td>
<td>36</td>
</tr>
<tr>
<td>2 Experiment 1: Effects of Long-term Musical Perception and Production Training on Lexical Tone Perception and Production</td>
<td>38</td>
</tr>
<tr>
<td>Methods</td>
<td>38</td>
</tr>
<tr>
<td>Participants</td>
<td>38</td>
</tr>
<tr>
<td>Stimuli</td>
<td>40</td>
</tr>
<tr>
<td>Procedure</td>
<td>41</td>
</tr>
</tbody>
</table>
3 EXPERIMENT 2: EFFECTS OF LABORATORY PERCEPTION AND PRODUCTION TRAINING ON TONE DISCRIMINATION – A BEHAVIORAL AND ERP STUDY

Methods .......................................................................................................................... 57
Participants ....................................................................................................................... 57
Stimuli .............................................................................................................................. 58
    Behavioral task ........................................................................................................... 58
    ERP ............................................................................................................................. 60
Procedure ......................................................................................................................... 61
    Day 1 ............................................................................................................................ 62
    Day 2 ............................................................................................................................ 64
    Day 3 ............................................................................................................................ 65
Data Analysis .................................................................................................................. 65
    Behavioral task ......................................................................................................... 65
    ERP ............................................................................................................................ 66
Predictions ....................................................................................................................... 68
Results ............................................................................................................................. 69
Behavioral Task .............................................................................................................. 69
    Accuracy ..................................................................................................................... 69
    Reaction times ............................................................................................................ 72
    Percentage accuracies of different tone-pairs ............................................................ 74
    Summary of behavioral results .................................................................................. 75
ERP ................................................................................................................................. 76
    100 – 200ms window .................................................................................................. 77
    MMN ............................................................................................................................ 78
    Late negativity ............................................................................................................. 79
    Summary of ERP results ............................................................................................ 80

4 EXPERIMENT 3: COMPARISON OF PERCEPTION TRAINING AND PRODUCTION TRAINING ON TONE IDENTIFICATION

Methods ............................................................................................................................ 97
Participants ....................................................................................................................... 97
Stimuli ............................................................................................................................... 98
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>99</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>101</td>
</tr>
<tr>
<td>Results</td>
<td>103</td>
</tr>
<tr>
<td>Accuracy</td>
<td>103</td>
</tr>
<tr>
<td>Reaction Times</td>
<td>105</td>
</tr>
<tr>
<td>Percentages of Correct Identification for Each Tone</td>
<td>107</td>
</tr>
<tr>
<td>Summary</td>
<td>109</td>
</tr>
<tr>
<td>5 CONCLUSIONS</td>
<td>117</td>
</tr>
<tr>
<td>Summary of Results</td>
<td>118</td>
</tr>
<tr>
<td>Discussion</td>
<td>120</td>
</tr>
<tr>
<td>Effects of Vocal Experience and Instrumental Experience on Lexical Tone</td>
<td>120</td>
</tr>
<tr>
<td>Perception and Production</td>
<td></td>
</tr>
<tr>
<td>Effects of Laboratory Perception training and Production Training on</td>
<td>120</td>
</tr>
<tr>
<td>Tone</td>
<td></td>
</tr>
<tr>
<td>Perception</td>
<td>124</td>
</tr>
<tr>
<td>Discrimination training paradigm</td>
<td>124</td>
</tr>
<tr>
<td>Identification training paradigm</td>
<td>128</td>
</tr>
<tr>
<td>Final remarks</td>
<td>129</td>
</tr>
<tr>
<td>APPENDIX</td>
<td></td>
</tr>
<tr>
<td>A MUSICAL BACKGROUNDS FOR THE ENGLISH MUSICIANS IN EXPERIMENT1</td>
<td>131</td>
</tr>
<tr>
<td>B LANGUAGE AND EDUCATION QUESTIONS</td>
<td>132</td>
</tr>
<tr>
<td>C PERSONAL HISTORY FORM FOR LANGUAGE EXPERIMENTS WITH EVENT-RELATED</td>
<td>134</td>
</tr>
<tr>
<td>POTENTIALS OR FMRI</td>
<td></td>
</tr>
<tr>
<td>D HANDEDNESS QUESTIONNAIRE</td>
<td>136</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>137</td>
</tr>
<tr>
<td>BIOGRAPHICAL SKETCH</td>
<td>144</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1.</td>
<td>Mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians</td>
<td>53</td>
</tr>
<tr>
<td>2-2.</td>
<td>Mean percentages of participants’ correct productions that were agreed upon by all 3 judges and 2 judges respectively for each tone type and group</td>
<td>53</td>
</tr>
<tr>
<td>3-1.</td>
<td>Mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the perception-only and perception-plus-production groups</td>
<td>82</td>
</tr>
<tr>
<td>3-2.</td>
<td>The numbers and the percentages of trials that were included in data analysis after artifact rejection in the pre- and post-training tasks for the perception-only and perception-plus-production groups respectively</td>
<td>82</td>
</tr>
<tr>
<td>4-1.</td>
<td>Mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the identification-only and identification-plus-imitation groups</td>
<td>111</td>
</tr>
<tr>
<td>A-1.</td>
<td>Musical backgrounds for the American English musicians in Experiment1</td>
<td>131</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1.</td>
<td>Waveforms and spectrograms with the pitch contours of the Mandarin Tone1, Tone2, Tone3 and Tone4 associated with the syllable [ma]</td>
<td>53</td>
</tr>
<tr>
<td>2-2.</td>
<td>D-Prime Scores with the standard errors for Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively</td>
<td>54</td>
</tr>
<tr>
<td>2-3.</td>
<td>Mean reaction times with the standard errors of all the correct responses for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively</td>
<td>54</td>
</tr>
<tr>
<td>2-4.</td>
<td>The mean accuracy percentages with the standard errors for the same pairs in the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively</td>
<td>55</td>
</tr>
<tr>
<td>2-5.</td>
<td>Mean accuracy percentages with the standard errors for each of the different pairs in the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively</td>
<td>55</td>
</tr>
<tr>
<td>2-6.</td>
<td>Mean percentages and the standard errors of the correctly identified productions for each tone in the Chinese non-musicians, the English vocalists, instrumentalists and non-musicians respectively</td>
<td>56</td>
</tr>
<tr>
<td>2-7.</td>
<td>Distribution of the participants’ identified production of T3</td>
<td>56</td>
</tr>
<tr>
<td>3-1.</td>
<td>Spectrograms of Tone1, Tone2 and Tone4 for the syllable [tʰu]</td>
<td>82</td>
</tr>
<tr>
<td>3-2.</td>
<td>Step illustration of the perception-only training and the perception-plus-production training</td>
<td>83</td>
</tr>
<tr>
<td>3-3.</td>
<td>Mean d-prime scores with the standard errors of the pre-training task, training session and the post-training task for the perception-only and the perception-plus-production group respectively</td>
<td>83</td>
</tr>
<tr>
<td>3-4.</td>
<td>Mean d-prime scores with the standard errors for the trained and untrained stimuli in the pre- and post-training tasks for the perception-only and the perception-plus-production group respectively</td>
<td>84</td>
</tr>
<tr>
<td>3-5.</td>
<td>Mean d-prime scores with the standard errors for the tested and untested stimuli in the training session for the perception-only and the perception-plus-production group respectively</td>
<td>84</td>
</tr>
</tbody>
</table>
3-6. Mean reaction times with the standard errors of all the correct responses in the pre-training task, the training session and the post-training task for the perception-only and production-plus-perception groups respectively .......................... 85

3-7. Mean reaction times with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups respectively ............... 85

3-8. Mean reaction times with the standard errors of the tested and untested stimuli in the training session for the perception-only and the perception-plus-production groups................................................................................................................................. 86

3-9. Mean percentage accuracies with standard errors of each tone pair in the pre- and post-training tasks for the perception-only and perception-plus-production groups ........................................................................................................................................ 86

3-10. Mean percentage accuracies with standard errors of each tone pair in the training session for the perception-only and the perception-plus-production groups ........................................................................................................................................ 87

3-11. The ERPs of the frontal electrodes (F3, Fz and F4), the central electrodes (C3, Cz and C4) and the parietal electrodes (P3, Pz and P4) collapsed over all six conditions ........................................................................................................................................ 88

3-12. The ERPs of the Fz electrode in the pre- and post-training tasks for the perception-only and perception-plus-production groups respectively .......... 92

3-13. Deviant minus standard difference waves of the Fz electrode for the six conditions in the pre- and post-training tasks for the two groups ................. 95

4-1. Step illustration of the identification-only training and the identification-plus-imitation training .............................................................................................................. 111

4-2. Mean d-prime scores with the standard errors for the identification-only and the identification-plus-imitation groups in the pre-training task, the training session and the post-training task respectively ......................................................... 112

4-3. Mean d-prime scores with the standard errors for the trained and untrained stimuli in the pre- and post-training tasks for the identification-only and the identification-plus-imitation group respectively ................................................................. 112

4-4. Mean d-prime scores with the standard errors of the tested and untested stimuli in the training session for the identification-only and the identification-plus-imitation groups respectively ................................................................. 113

4-5. Mean reaction times with the standard errors of all the correct responses in the pre-training task, the training session and the post-training for the identification-only and the identification-plus-imitation group......................... 113
4-6. Mean reaction times with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups respectively .......... 114

4-7. Mean response times with the standard errors of the tested and untested stimuli in the training session for the identification-only and identification-plus-imitation groups respectively ................................................................. 114

4-8. Mean percentages of correct identification with the standard errors for each tone type in the pre- and post-training tasks for the identification-only and the identification-plus-imitation groups .................................................................................. 115

4-9. Mean percentages of correct identification with standard errors for each tone type in the training session for the identification-only and the identification-plus-imitation groups respectively ................................................................. 115

4-10. Error distribution for the identification-only and the identification-plus-imitation groups ............................................................................................................................ 116
LIST OF ABBREVIATIONS

EEG       Electroencephalogram
ERP       Event-related potentials
F0        Fundamental frequency
MMN       Mismatch Negativity
PAIR11    Tone pair Tone1-Tone1
PAIR22    Tone pair Tone2-Tone2
PAIR33    Tone pair Tone3-Tone3
PAIR44    Tone pair Tone4-Tone4
PAIR12    Tone pairs Tone1-Tone2 & Tone2-Tone1
PAIR13    Tone pairs Tone1-Tone3 & Tone3-Tone1
PAIR14    Tone pairs Tone1-Tone4 & Tone4-Tone1
PAIR23    Tone pairs Tone2-Tone3 & Tone3-Tone2
PAIR24    Tone pairs Tone2-Tone4 & Tone4-Tone2
PAIR34    Tone pairs Tone3-Tone4 & Tone4-Tone3
SD        Standard deviation
T1        Tone1
T2        Tone2
T3        Tone3
T4        Tone4
T12       One condition in the ERP data: deviant T1 in T2 standards
T14       One condition in the ERP data: deviant T1 in T4 standards
T21       One condition in the ERP data: deviant T2 in T1 standards
T24       One condition in the ERP data: deviant T2 in T4 standards
T41       One condition in the ERP data: deviant T4 in T1 standards
One condition in the ERP data: deviant T4 in T2 standards
The relationship between speech perception and production has been debated for a long time. The Motor Theory of speech perception (Liberman et al., 1989) claims that perceiving speech is identifying the intended articulatory gestures rather than perceiving the sound patterns. It seems to suggest that speech production precedes speech perception, i.e., only when people can accurately ‘feel’ their vocal tract gestures, can they successfully perceive speech. One prediction that falls out from the Motor Theory is that training people to produce sounds should be more effective for improving the perceptual ability than training people to only perceive sounds. My dissertation aims to test this prediction by comparing the effects of laboratory perception training and production training as well as the effects of different musical backgrounds (i.e., vocalists and instrumentalists) on the perception of lexical tones.

The present dissertation encompasses three experiments. The first experiment compared native English non-musicians, instrumentalists and vocalists on the perception and production of Mandarin tones. The results supported previous findings, demonstrating the advantage of musical training on lexical tone perception. However,
no additional benefit was found for musicians with vocal experience in addition to instrumental experience on either lexical tone perception or production. The second experiment recorded both behavioral and electrophysiological data to examine the effectiveness of a perception-only training and a perception-plus-production training on lexical tone perception. The data showed that after training participants in both groups improved on tone discrimination. Moreover, the participants in both groups did not differ in tone processing at the intentional or unintentional level after training. The third experiment, using identification training rather than the discrimination training, demonstrated that the identification-plus-imitation group identified the tones more quickly in the post-training task than in the pre-training task while the identification-only group did not show any improvement in terms of reaction time. These results suggested that the employment of the motor system does not specifically benefit the tone perceptual skills.
CHAPTER 1
INTRODUCTION

Project Motivation

The relationship between speech perception and production has been debated for a long time. The Motor Theory of speech perception (e.g. Liberman et al., 1967; 1989) asserts that speech perception is accomplished by identifying the intended articulatory gestures rather than perceiving the sound patterns. The Direct Realist Theory (e.g. Best, 1995; Fowler, 1986) agrees with the Motor Theory in that the perceptual primitives are the vocal tract gestures, but claims that the primitives are the actual vocal tract gestures rather than intended gestures mediated by an innate neural module. Despite the subtle differences between the Motor Theory and the Direct Realist Theory, both of them imply that speech production precedes speech perception, i.e., only when people can accurately ‘feel’ their vocal tract gestures can they successfully perceive speech. Moreover, several studies have shown that the accurate production of sounds in the second language precedes their accurate perception (e.g. Flege & Eefting, 1987; Sheldon & Strange, 1982; Goto, 1971), indicating that speech perception might rely on speech production. Thus, according to the Motor Theory, the Direct Realist Theory and previous studies, training people to produce sounds (or to ‘feel’ the articulatory gestures) is expected to be more effective in improving the ability of comprehension than training people to perceive sounds. The current dissertation aims to test this prediction by comparing the effects of laboratory perception training and production training as well as the effects of different musical backgrounds (i.e., vocal training and instrumental training) on the perception of lexical tones. The findings of this research will contribute to the current models of speech perception and will illuminate
the relationship between the auditory system and the articulatory system as well as the relationship between music and language.

Background

Views of the Relation between Production and Perception

The Motor Theory was first proposed by Liberman and colleagues in the 1950s. Since its first appearance, the theory has undergone several revisions. The most recent version of the Motor Theory claims that (1) speech is perceived by recognizing the intended vocal tract gestures; and (2) speech perception and speech articulation are linked through an innate mechanism (Liberman & Mattingly, 1989). Results from several behavioral and neuropsychological studies have supported these two assertions. For example, Adank, Haggort & Bekkering (2010) found that no matter whether people received auditory feedback (of their own voice) or not, vocal imitation significantly improved the comprehension of unfamiliar accent. Ojemann (1983) demonstrated that speech perception and speech production share many cortical areas, such as the left inferior frontal cortex and the superior temporal lobes. The discovery of mirror neurons (Di Pellegrino et al., 1992) was also argued to be in support of the Motor Theory (e.g. Galantucci, Fowler & Tuvey, 2006; but see Lotto, Hickok & Holt, 2009). Mirror neurons are neurons that are activated both during action execution and during action observation. It suggested that perception may be channeled through the production pathway.

On the other hand, the Motor Theory has been criticized on theoretical and empirical grounds (e.g. Lotto et al., 2009; Massaro & Chen, 2008). For instance, some research on Broca’s aphasics showed that impairment in the motor systems that control speech production was not associated with deficiency in speech comprehension (e.g.
Moineau, Dronkers & Bates, 2005). However, these studies only focused on the single-word level and most patients in the studies were over 40 years old. Thus, the retained comprehensive ability could probably be attributed to that the kinesthesia of the vocal tract gestures had been stored in the long-term memory through years of articulatory practice. Patients may have used this stored information to aid in their comprehension, resulting in the gradual deterioration of perceptual ability.

Hickok and Poeppel propose a dual-stream model of speech processing and claim that the networks between speech perception and production are partially overlapping and partially distinct (e.g. Hickok, 2001; Hickok & Poeppel, 2004 & 2007). They agree with the Motor Theory that there is a link between speech perception and production, but deny that speech production is mandatory for speech perception. Other researchers even take a position opposite from the Motor Theory, arguing that speech production relies on speech perception (e.g. the Directions Into Velocities of Articulators Model (DIVA), Guenther, Ghosh & Tourville, 2006; and the Speech Learning Model, Flege, 1995a). This assertion has been supported by a magnetoencephalography (MEG) study by Levelt et al. (1998), who showed that the initiation of articulatory processes was slightly slower than the auditory cortical activations during speech production.

In sum, previous literature demonstrates mixed results and accounts regarding the relationship between speech production and perception. Further research is needed to directly compare the effect of production learning on speech perception and effect of perceptual learning on speech production, in order to uncover the nature of relationship between speech perception and speech production. In the following sections, I will
briefly review studies regarding perception training and production training of novel phonological categories.

**Previous Training Studies**

Previous research has shown that participants’ ability to discriminate/identify novel phonemic contrasts can be improved through relatively short-term laboratory perception training or production training. Moreover, some studies also demonstrated that participants can transfer the improvement gained through the perception training to the production domain and vice versa. These results suggest that speech perception and production are linked to each other.

**Perception training**

Jamieson & Morosan (1986) trained the Canadian francophone adults to identify and discriminate the English voiced and unvoiced interdental fricatives /θ/ and /ð/. The results showed that after 90 minutes of identification training with synthesized CV tokens participants’ identification performance and inter-category discrimination improved significantly compared with the control group. Moreover, the improvement with the synthetic CV tokens was transferred to the natural CV tokens that were not included in the training. Logan, Lively & Pisoni (1991) improved the training method used in Jamieson & Morosan (1986) by utilizing a large variety of speakers and phonetic contexts (e.g. word-initial, word-final and intervocalic) to train native speakers of Japanese to distinguish the English /r/ and /l/. The participants in this study received fifteen 40-minute training sessions in which they were asked to identify the stimulus from a minimal pair contrasting in /r/ and /l/ and received feedback. Logan et al. (1991) reported that the Japanese listeners’ identification accuracy improved after training, but the improvements were more prominent in the word-final and intervocalic conditions.
than in the word-initial condition. In addition, Logan et al. (1991) performed two generalization tests and found that the participants identified /r/ and /l/ more accurately when novel words were produced by an old talker than when novel words were produced by a new talker that they had never heard in the training sessions.

Bradlow, Pisoni, Akahane-Yamada & Tohkura (1997) followed Logan et al. (1991) but extended their training sessions to a longer period of time to examine the Japanese participants’ improvements on the identification of English /r/ and /l/. The results demonstrated that the Japanese participants improved significantly on the identification of English /r/ and /l/ and performed equally well on the two generalization tests no matter whether the novel words were produced by an old talker or a new talker. Furthermore, they tested the participants’ productions of the English /r/ and /l/ and found that the participants’ productions improved significantly in the post-test compared with the pre-test, even though the training sessions did not include any overt production component. The perception and production improvements were even retained three month after training (Bradlow et al., 1999). These results provide evidence for the link and transfer between speech perception and production.

Production training

As reviewed in the previous section, several studies have investigated the effectiveness of perception training on the identification/discrimination of novel phonemes. However, only a small body of research has directly examined the effect of production training on the perception of novel sound contrasts and these studies demonstrated controversial results.

Hirata (2004) examined the production and perception of Japanese pitch and duration contrasts by native English speakers. The participants in this study received a
production training in which they first saw the fundamental frequency (F0) contours of
the model utterance produced by native Japanese speakers, and then produced the
utterance by themselves. After the participant’s production the computer program
overlaid the F0 contour of the model utterance onto the participant’s F0 contour. The
participants were asked to reproduce the utterance until they thought their F0 contour
matched the model. Moreover, they could listen to the model utterance whenever they
wanted during their reproduction. The participants received ten of 30-minute training
sessions during three and a half weeks. Hirata (2004) found that the participants
improved significantly on both production and perception of Japanese pitch and duration
contrasts after training. However, since the participants in Hirata (2004) could listen to
the model utterances as many times as they wanted during the training sessions, it is
unclear whether the improvements were due to the production training alone or the
production training plus the perceptual exposure.

Hattori (2009) obtained different result from Hirata (2004) through examining the
perception and production of the English /r/ and /l/ by native Japanese speakers.
Participants in this study received 10 sessions of 30-40 minute production training,
including one-on-one explicit instruction and feedback regarding the appropriate
articulatory gestures and positions of the production of /r/ and /l/ and repeated
imitations. Hattori (2009) found that the Japanese speakers improved significantly on
the production of English /r/ and /l/, and even showed native-like accuracy after the
production training. However, their perception did not improve after training. Hattori
(2009) took these results as evidence to assume that speech perception and production
might have separated mental representations, and learning in one domain might not transfer to the other domain.

**Comparison between perception training and production training**

Some recent studies also compared the effectiveness of perception training and production training on the perception of L2 phonemic contrasts. For example, Herd (2011) trained native English learners of Spanish to perceive and produce the Spanish intervocalic contrasts between the tap /ɾ/, the trill /ɾ/ and the dental stop /d/ using three training paradigms. In the perception training the participants heard one word that contained one of the sounds and saw a corresponding minimal pair. Then they were asked to choose the word they heard from the minimal pair and received feedback. In the production training the participants were first presented with the waveform and spectrogram of a native speaker’s production of the stimulus, and then were asked to produce the stimulus themselves. Next they compared the waveform and spectrogram of their own production with the model and were encouraged to reproduce until they thought the two waveforms and spectrograms matched. The participants who received the production training could never hear the native speakers’ productions. In the combination training the participants received three sessions of perception training and three sessions of production training as described above. Herd (2011) found that all three types of training improved participants’ perception and production of the /r, ɾ/ contrast in Spanish and the improvements in perception were also generalized to new speakers, new dialects and new words. However, the perception of the /r, d/ contrast did not improve in any of the training groups. Specifically, the participants who received the perception training showed equal perception ability of the /r, ɾ/ contrast in the pre- and post-tests, while the participants who received the production training and the
combination training declined significantly in their perception of this contrast after training. Moreover, participants in all the three training groups improved in the production of the trill /r/. Taken together, Herd (2011) concluded that the perception training was more effective than production training to transfer to the other domain.

Nevertheless, since the participants who received the perception training could hear the native speakers’ productions while the participants who received the production training were not allowed to hear the model productions, it was unclear whether the advantage of the perception training was due to the extra exposure or the perception training alone. In addition, the declined perception in the production training and the combination training group was probably due to that the participants heard their own inaccurate productions, which offset the training effect in perception.

As previously mentioned, one concern in this line of research is that it is just impossible to completely isolate speech production from speech perception, i.e. we perceive what we are producing inevitably. Therefore, some studies have compared perception-only training to perception plus production training. For example, Leach & Samuel (2007) demonstrated that perception training plus a production task was more powerful to enhance participants’ identification of words in noise (i.e. lexical configuration tasks) than perception training alone. However, they also found that the extra production task hindered the participants’ perceptual learning in the lexical engagement tasks (e.g. phonemic restoration).

Baese-Berk (2010) compared the effectiveness of perception-only training and perception-plus-production training on the perception and production of novel sound contrast by native English listeners. In the perception-only training participants heard
one syllable and saw a corresponding picture demonstrating the distribution of tokens on a continuum. The participants were asked to pay close attention to the picture and press a button on a button box to advance to the next trial when they were ready. The perception-plus-production training was exactly the same as the perception training except that participants were asked to imitate the stimulus they heard instead of just pressing a button on the button box. The participants in the perception-plus-production training were only allowed to hear and imitate each stimulus once. Baese-Berk (2010) found that the perception-only training was more effective to improve the participants’ perception of novel sound contrasts than the perception-plus production training. Moreover, even though the participants who received the perception-only training also improved in their production, the participants in the perception-plus-production training showed more significant improvement. Baese-Berk (2010) concluded that the perception-only training transferred more effectively than the perception-plus-production training, and the perceptual learning might be hindered by the additional production task.

Taken together, previous studies have examined the effects of perception training and production training on the perception of novel phonemic contrasts and demonstrated mixed results. Very few studies have compared the effects of perception training and production training on the perception of suprasegmental phonological contrasts. The current study investigates this question by focusing on lexical tones.

**Tones in Music and Language**

Music and language share several acoustic characteristics, such as duration, rhythm, and pitch. Among these characteristics, pitch has particularly attracted researchers’ interest. Both music and language use time-varying pitch to convey
information. In music, melody involves two types of pitch information: contour and interval code. Contour refers to the pitch direction variations between successive tones while interval code is the pitch distance between notes. In language, pitch differences may signal different levels of prosodic contrasts. 60% to 70% of the world’s languages are tone languages, which use pitch differences to distinguish lexical meaning and to carry grammatical distinction (Chao, 1948). For example, in Mandarin Chinese, the syllable [ma] produced with four different tones (Tone1 has high-level pitch, Tone has a high-rising pitch, Tone3 a low-dipping pitch, and Tone4 a high-falling pitch) will result in four different words: [ma1] ‘mother’, [ma2] ‘hemp’, [ma3] ‘horse’ and [ma4] ‘scold’ (Chao, 1948). In intonation languages, such as English, pitch variation is used to convey prominence and/or pragmatic meaning rather than encode lexical meaning. At the word level, English exhibits lexical stress, with some syllables being perceptually more prominent than others. The pitch of the stressed syllable is typically higher than the unstressed syllables. At the sentence level, pitch variation is used to distinguish different sentence types (question or statements) and to flag information as new or unpredictable.

**Tone Perception and Production by Non-native Speakers**

Previous studies have shown that non-native listeners of tone languages have difficulties in both comprehending and producing lexical tones (e.g. Gandour 1983; Wang et al. 1999; White, 1981). These difficulties have been attributed to: (1) different perceptual dimensions by non-native speakers compared with native speakers, and (2) the interference from suprasegmental features in the native language. For example, Gandour (1983) found that native English listeners tended to pay more attention to the pitch onset, offset and the average pitch, while native speakers of tone languages focus
more on the pitch contour. Neurophysiological research has also provided support for these differences. Krishnan et al. (2005) reported that the brain stem responses are a more accurate/consistent reflection of pitch contour in tone language listeners than in non-tone language listeners. Using Event Related brain Potentials (ERPs), Chandrasekaran, Krishnan & Gandour (2007) found that a larger mismatch negativity (MMN) was elicited by tone contrasts in tone language listeners than in non-tone language listeners. As for the native language interference, Shen (1989) claimed that English speakers tended to perceive the Mandarin Tone 4 as stressed and Tone 3 as unstressed. Moreover, English speakers are likely to consider Mandarin Tone 2 as the rising intonation in question sentences, and Tone 4 as the falling intonation in the final position of statement sentences.

In terms of production, Chen (1974) claimed that non-tone language speakers needed to learn to widen their pitch range in order to successfully produce a tone language. Several studies have found that native English learners of Chinese have difficulty in producing all four tones (e.g. Miracle, 1989). Shen (1989) reported that Tone 4 was the most difficult tone for non-native speakers, which was ascribed to the interference of the English intonation system. English speakers usually produced Tone 4 at a lower pitch onset and with a less steep falling slope than the native Chinese speakers, because they associated Tone 4 with the falling intonation in English, which starts at a lower pitch and falls down gradually.

**Effect of Music Background on Lexical Tone Perception**

Some evidence has suggested that long-term experience with pitch differences is domain general, i.e. capacities in language domain percolate to music domain and vice versa. For example, Pfordresher & Brown (2009) showed that tone language speakers
outperformed non-tone language speakers in both vocally imitating and perceptually discriminating musical pitch. On the other hand, music training has also been reported to facilitate speech perception (e.g., Anvari, Trainor, Woodside & Levy, 2002; Slevc & Miyake, 2006).

Several behavioral studies have demonstrated that musicians were better than non-musicians in lexical tone discrimination and identification. For example, Lee & Hung (2008) performed a behavioral identification task on English musicians and non-musicians. Participants were asked to identify four Mandarin tones associated with the syllable [sa] by 32 speakers after a brief tutorial. The identification by musicians was more accurate than that by non-musicians. Schwanhäußer (2007) also found that English musicians were better than non-musicians in identification and discrimination of Thai tones. At the level of brainstem, Wong et al. (2007) found that pitch tracking was more faithful and robust in musicians than in non-musicians. Using ERPs, Chandrasekaran, Krishnan & Gandour (2009) demonstrated that the mean amplitude of the MMN was larger in English musicians than in English non-musicians during Mandarin tone processing.

While there are comprehensive studies on trained instrumental musicians, relatively little is known about formally trained vocal musicians and the differences between the vocalists and instrumentalists. Besides auditory pitch discrimination ability (perception), vocal pitch control (production) is another essential skill for successful vocal musicians. Using non-linguistic pitch stimuli, Zwissler (1971) found that accurate singers performed significantly better on the auditory pitch discrimination task than the inaccurate singers, suggesting that there is a positive relationship between auditory
pitch discrimination abilities and accurate singing. However, it is unknown whether this advantage in vocal musicians could also carry over to the language domain. Nikjeh (2006) also used musical stimuli and examined the intentional and unintentional pitch discrimination and pitch production by English vocalists and instrumentalists. The results demonstrated that the vocalists and instrumentalists did not differ significantly in pitch perception or pitch production. Nikjeh (2006) claimed that these result was probably due to that the perception and production tasks were too easy to show different effect of music background between the vocalists and instrumentalists.

**Effects of Laboratory Training on Lexical Tone Perception**

Several behavioral studies have shown that short-term perceptual and production training is effective in improving the comprehension and production of lexical tones by non-tonal language speakers. Moreover, the effects of perception training and production training can be inter-transferable.

**Perception training**

Wang and colleagues conducted a series of tone experiments on the effectiveness of perception training. Wang et al. (1999) trained native American-English speakers for two weeks to identify Mandarin Chinese tones in 100 natural Chinese words. After the perception training, the English speakers improved significantly in identifying the tones. The improvement also generalized to untrained stimuli and was retained six months after training. Wang, Jongman & Sereno (2003a) further showed that the improvement in identification and discrimination of lexical tones gained through perceptual training transferred to production. Wang et al. (2003b) conducted a functional magnetic resonance imaging (fMRI) study on the same group of participants as in Wang et al. (1999). They found that after training the improvements observed in
the behavioral study were associated with an increased activation in Wernicke’s area. Effects of perception training have also been found for other tone languages. Wayland & Guion (2004) demonstrated that native English listeners revealed significantly improved ability in perceiving Thai lexical tone contrasts after an auditory training. Finally, using ERPs, Kaan et al. (2007) examined the effect of perceptual identification training on the unintentional processing of Thai tones by naïve speakers, as measured by the MMN. They found that English speakers showed an increased MMN only for high deviant tone stimuli, while the MMN for low deviant stimuli did not differ before and after training.

**Production training**

Leather (1990) was the first to examine the effect of production training on perception. In this study, Leather trained native Dutch speakers to produce the syllable [yu] with four Mandarin tones. The result showed that after the production training the Dutch speakers were able to perceive the differences in tones. Furthermore, Leather (1990) also observed an effect of perceptual training on production: after the Dutch listeners were trained to identify the four tones, their productions were more accurate. However, since Leather (1990) only used one syllable in both the training and the post-training tasks, it is unclear whether the improvement generalizes to other stimuli.

**Interim Summary**

After reviewing previous studies on the effects of perception training and production training on tone perception in both language and music domains, we can identify several gaps in the current literature. (1) Although some research has suggested that vocal experience might positively affect music pitch perception, it is unknown whether this advantage in vocal musicians could also carry over to the
language domain. (2) Very few studies have directly compared the effectiveness of production training and perception training on the perception of lexical tones. (3) To our knowledge, no study has investigated which stage of processing can be affected by production training: will the production and perception training have a different effect on lexical tone perception at the behavioral (intentional) level? Or will the two types of training also differentially affect unintentional processing (i.e. when the auditory stimuli are not relevant to the participants’ task)? The current study aims to provide more data in terms of these three aspects by using both behavioral and neurophysiological methods to compare the effects of production training and perception training (in both music and language domain) on the intentional and unintentional processing of lexical tones. In the next section, I will discuss the neurophysiological method used in the present study to examine the unintentional processing of lexical tones: the event-related potentials (ERPs).

**Event-related Potentials (ERPs)**

Event-related potentials (ERPs) are a good method to study the unintentional processing of lexical tones while the auditory stimuli are presented to participants. ERPs do not require participants’ attention and can provide high temporal resolution. The participants can, for instance, read a book or watch a movie while their brains’ responses to the auditory stimuli are recorded. In contrast, behavioral tasks provide data regarding participants’ intentional responses after they fully process the stimuli. Therefore, a passive ERPs task, combined with a behavioral task could give a thorough view of how people intentionally and unintentionally process lexical tones, and indicate which level of processing is affected by the type of training (i.e. perception vs. production training). Moreover, previous studies on speech-sound training have shown
that changes in neural activity may precede the behavioral improvement (e.g. Tremblay, Kraus & McGee, 1998). Recording neurophysiological measures may therefore be more informative than only recording behavioral measures, since perceptual changes after training may occur at the unintentional level but not (yet) at the intentional/behavioral level.

**ERP Components that are Relevant to Auditory Perception**

Several ERP components have been found to be relevant to auditory perception. The Mismatch Negativity (MMN) is a negative wave elicited by auditory stimuli that are infrequently presented and deviate from a frequently presented standard stimuli. The MMN usually occurs between 100-300ms after the onset of deviant stimuli (e.g. Näätänen & Alho, 1995), and has been claimed to reflect pre-attentive processing (although this component may be sensitive to intentional manipulations, e.g., Woldorff, Hackley & Hillyard, 1991). Many studies have shown that the behavioral improvement in speech perception is usually accompanied by an increased MMN (e.g. Tremblay et al., 1997). Sometimes the increased MMN is observed even before the behavioral improvement occurs (e.g. Tremblay et al., 1998). The Late Negativity is another negative wave that occurs around 350 to 600ms after the onset of deviant stimuli. The late negativity has been associated with the reorientation of attention (Shestakova et al., 2003). Some studies have shown that the late negativity became smaller after training (e.g. Kaan, et al., 2008), suggesting that after training the deviant stimuli became easier to distinguish and required less attention. Besides MMN and late negativity, the N1 and P2 components have also been reported to be sensitive to the physical features of auditory stimuli. The N1 is a negative wave that occurs between 80 and 120ms; the P2 is the second positive peak, which is normally observed at 180-250ms after the onset of
the stimulus. These two components are often referred to as the “N1-P2” complex. Previous research has shown that the amplitude of P2 increases after training (e.g. Atienza, Cantero & Dominguez-Marín, 2002). The P1 and P300 (P3a and P3b) might also be relevant to the current study. The P1 and P300 are the first and third positive peaks that often occur at about 100ms and 300ms after the stimulus onset respectively. The amplitude of P1 is modulated by selective attention (e.g. Van Voorhis & Hillyard, 1977). P3a often follows the MMN and has been argued to be an index of an automatic attention switch toward a deviant stimulus (Escera et al., 1998). If these ERP components are differentially affected by the type of training (perception-plus-production or perception-only), it suggests that the additional production component during training changes the unintentional processing of the auditory stimuli after training.

The Current Study

Research Questions

The present study aims to answer three main questions:

- Are the effects of production training and perception training domain general or domain specific? That is, can the instrumental and vocal expertise percolate to lexical tone perception and production?

- Is laboratory perception-plus-production training more effective than perception-only training in facilitating lexical tone perception at the intentional level?

- Does perception-plus-production training affect the unintentional perception of lexical tones differently from perception-only training?

Structure of the Dissertation

Three experiments investigating the relationship between perception and production are presented in the current dissertation. The first experiment examines the production and perception training in the music domain, while the second and third experiments focus on the language domain. Experiment 1 compares English vocalists
and instrumentalists to see which type of musical training is more advantageous to lexical tone perception and production. Participants were tested in (1) a same/different discrimination task, in which they decided whether the two syllables they heard had a same or different tone; and (2) a production task, in which they heard a tone and imitated immediately. Experiment 2 collected both behavioral and electrophysiological (ERPs) data to examine the effectiveness of a perception-plus-production training and a perception-only training on the intentional and unintentional processing of tones by native English speakers. Participants were trained and tested over the course of three consecutive days. On the first day, participants' brain waves (ERPs) were recorded as baseline. During recording, participants were watching a silent movie while a stream of syllables was presented. After the ERP recording, participants did a behavioral same/different discrimination task. On the second day, participants received either a perception-only or a perception-plus-production training. The trainings for both groups were exactly the same except that the perception-plus-production training required participants to imitate the stimuli, while the perception-only training had participants utter a word unrelated to the stimuli after a same/different discrimination training. On the third day, both groups did the same ERP and behavioral tasks as on the first day. If the employment of the motor system benefits the perceptual skills, the perception-plus-production group should show more improvement and an increased sensitivity to the tone contrasts as a function of training than the perception-only group. However, this experiment is biased toward the perception-plus-production group because the imitation task may require the participants in this group to pay more attention to the stimuli than the participants in the perception-only group. In other words, the advantage of the
perception-plus-production training (if there is any) is probably due to the participants’ close attention on the stimuli rather than the specific training type. In order to untangle this problem, Experiment 3 was designed to be biased for the perception-only group. If the perception-plus-production group still show more improvement than the perception-only group, it is clear that the production training is more effective than the perception training and the employment of the articulatory system benefits the perceptual skills. In Experiment 3, participants were trained and tested over the course of two consecutive days. On the first day, participants first did an identification task, in which they heard one stimulus and then identified whether this stimulus was a level, rising or falling tone. After the baseline task, the participants received either a perception-only or a perception-plus-production training. The trainings for both groups were exactly the same except that after an identification training the perception-plus-production training required participants to imitate the stimuli, while the perception-only training had participants utter the tone types of stimuli (i.e. level, rising or falling). On the second day, all participants did the same identification task as on the first day. This experiment is biased toward the perception-only group because the pre- and post-training tests asked participants to identify the tone types of the stimuli; but the tone types were only reinforced to the participants who received the perception-only training during the production process. The results of the three experiments will be presented and used to investigate the effectiveness of perception training and production training on lexical tone perception. Lastly, I will discuss the relationship between the results and other relevant literature on perception training and production training, and will suggest directions for future research in this area.
**Intellectual Merit**

The current study will contribute to psycho-/neuro-linguistic literature on speech learning and perception. First, there has been a long debate about whether production relies on perception or vice versa. Results of this study will illuminate the relationship between speech perception and production. Second, most behavioral and electrophysiological studies have thus far ignored the effect of production training on the perception of lexical tones, focusing only on the effect of perception training upon production. If our predictions are borne out, the perception-plus-production training should be more effective than the perception-only training on lexical tone perception. This result may inspire researchers to re-evaluate the importance of production training in second language learning. Third, the proposed research also introduces methodology that allows linguists to integrate production training into the study of speech perception. The training paradigm developed in the present study can be easily extended to investigate other aspects of speech.

**Broader Implications**

This study is primarily fundamental in nature. However results from the present and similar studies will eventually have implications for language education and language software development. With the economic development of China, there is an increasing demand for learning Chinese. Findings of the current study will help to develop more effective teaching methods for learning lexical tones, which have been claimed as an obstacle in learning Chinese by non-native speakers. Newer versions of language-learning software can also be programmed to utilize production training for different tone learning. Moreover, if both perception-only and perception-plus-production training lead to changes at the neural level, then further identification of the neural
substrates for pitch perception and vocal pitch control may have implications for treatment strategies for patients who have neurological, language and vocal impairments, such as dyslexia, aphasia, and hearing/production impairment.
CHAPTER 2
EXPERIMENT 1: EFFECTS OF LONG-TERM MUSICAL PERCEPTION AND PRODUCTION TRAINING ON LEXICAL TONE PERCEPTION AND PRODUCTION

The first experiment examines the effects of production training and perception training in the music domain on lexical tone perception and production. Previous research has demonstrated that musicians were more accurate than non-musicians in lexical tone discrimination and identification (e.g. Alexander, Wong & Bradlow, 2005; Wong et al., 2007a). This experiment further compared American English vocalists and instrumentalists to see which type of musical background was more advantageous to lexical tone perception and production. Learning to sing is very different from learning an instrument. Singers receive extensive modeling practice (i.e. they model their singing on perfect tones and remember the sensations of their productions on the basis of feedback) during vocal training, while instrumental musicians learn to match their hand position and finger placement (depending on different types of instruments) with the notes during instrumental practice. Therefore, one would expect that vocal musicians have more advantage than instrumental musicians in imitating lexical tones. Moreover, if there is indeed a strong connection between production and perception, as suggested by the motor theory (e.g. Liberman & Mattingly, 1989), the vocalists might also be superior to the instrumentalists in the perception of lexical tones. This experiment tested these predictions through a same/different discrimination task and a production imitation task.

Methods

Participants

A total of 60 participants were recruited from the University of Florida student body: 30 native speakers of American English with formal music training (15 vocalists
and 15 instrumentalists), 15 native speakers of American English without music training and 15 native speakers of Mandarin Chinese without music training. The English vocalists (1 male and 14 females) were between the ages of 19 and 30 (M = 22.53, SD = 2.95), and the English instrumentalists (4 males and 11 females) were between the ages of 18 and 23 (M = 20.2, SD = 1.42). All English vocalists had at least 3 years of continuous formal vocal training (M = 8.33, SD = 4.15), starting at or before the age of 19 (M = 13.60, SD = 4.05) and were still singing at the time of their participation. Thirteen of the vocalists had also received some instrumental training before they had formal vocal training, primarily on piano. The English instrumentalists had diverse instrumental backgrounds. All of them had at least 3 years of continuous formal instrumental training (M = 8.87, SD = 3.09), starting at or before the age of 14 (M = 9.6, SD = 2.16) and were still playing at least one instrument at the time of their participation. None of the English musicians claimed that they had perfect /absolute pitch, which refers to the ability to identify the pitch of any musical note without a reference pitch. Appendix A summarizes the details of the English musicians’ musical backgrounds. The English non-musicians (4 males and 11 females) were between the ages of 19 and 23 (M = 20.6, SD = 1.24). None of the English participants had previous exposure to any tonal languages. The Chinese non-musicians (6 males and 9 females) were between the ages of 22 and 27 (M = 23.73, SD = 1.94). All of them were originally from mainland China, and did not know any tone language other than Mandarin Chinese. All non-musicians had less than 2 years of musical training and had not performed music (played an instrument or sang) within the past 4 years. An additional 11 participants were recruited, but were excluded from data analysis either because they did not meet
the musical criteria or because of technical problems with the experimental programs. All participants received course credit or financial compensation for their participants. The experiment was approved by the University of Florida Institutional Review Board.

**Stimuli**

Stimuli consisted of six monosyllabic syllables ([tʰi], [li], [mi], [tʰo], [lo] and [mo]) associated with four Mandarin tones (Tone1: high-level; Tone2: high-rising; Tone3: low dipping; Tone4: high falling). These particular syllables were selected because American English has these segments as well. Therefore, Americans speakers could focus their attention on the lexical tones and were not distracted by unfamiliar syllables. Moreover, all the syllables with the four tones resulted in real words in Mandarin Chinese. Two female native speakers of Mandarin Chinese were recruited to record the stimuli using the Shure SM58 microphone and Marantz PMD660 digital recorder in a sound-attenuating booth. Each speaker produced each stimulus twice, yielding a total of 96 tokens (6 syllables × 4 tones × 2 speakers × 2 tokens/speaker). Only female talkers were recorded for stimuli because previous study had shown that adult female talkers were more intelligible than male adult speakers (Bradlow, Torretta & Pisoni, 1996). All the target words were produced in the phrase “Qing shuo TARGET zhege zi” (please say TARGET this word), in order to better mimic natural speech and to reduce the phonetic differences (e.g. duration and creaky voice) between the four tones. This design might be a confounding factor in the current study, because the tone of the target word could be influenced by the surrounding tones (Francis et al., 2006). Nevertheless, our pilot intelligibility task showed that if we recorded the syllables in isolation, even English speakers who had no exposure to lexical tones could easily notice the differences between the Tone3 and the other three tones according to the
longer duration and obvious turning point. Thus, we decided to exploit the frame recording in order to prevent the ceiling effect. The target syllables were segmented by using PRAAT 5.1.25 (Boersma & Weenink, 2010) and then were normalized for amplitude. Five native Mandarin-speaking informants were asked to judge the stimuli for intelligibility. Only words with a mean accuracy rate of 95% or above were included in the experiment. Figure 2-1 shows the waveforms and spectrograms with the pitch contours of the Mandarin Tone1, Tone2, Tone3 and Tone4 associated with the syllable [ma].

Procedure

Participants were first asked to read through the informed consent statement and complete a questionnaire that includes language, educational background and musical training experience (Appendix B). Informed consent was obtained from all of the participants after the purpose and procedures of the experiment were fully explained by the experimenter. Then the participants received a standard hearing test, in which they were wearing headphones and beeps of varying pitch levels (between 250 and 8k Hz at 25 dB) were presented randomly to either their left or right ear. They were asked to raise the corresponding hand to indicate in which ear they heard the sound. This task ensured that all the participants could hear the stimuli well and our results were not confounded by participants’ different hearing abilities. Participants’ auditory working memories were also tested on site through a backward digit span test. In the test, participants heard strings of numbers and were asked to immediately repeat them back in the given order or reverse order. Working memory has been shown to affect the perception of pitch differences (e.g. Payne, 2003). We therefore included this factor in data analysis to make sure that the advantages of tone discrimination were not due to
better working memory. The auditory working memory test was conducted in the participants' native languages, i.e. Chinese for the Chinese non-musicians and English for the English musicians and non-musicians. The mean accuracy percentages of the auditory working memory test were 81.19% (SD = 10.50%), 54.76% (SD = 9.12%), 58.81% (SD = 12.80%) and 60.71% (SD = 11.37%) for the Chinese non-musician, the English non-musician, the English vocalists and the English instrumentalists respectively. The accuracy percentage for each participant was submitted to a one-way ANOVA with group as the between-subject factor. The result showed that there was a main effect of group, $F(3, 56) = 17.21, p < .001$. Further pairwise comparisons yielded that the Chinese non-musicians responded more accurately in the auditory working memory test than the English non-musicians [$t(28) = 7.36, p < .001$], the English vocalists [$t(28) = 5.12, p < .001$] and the English instrumentalists [$t(28) = 5.24, p < .001$]. Importantly, no difference was found between the three English groups, $ps > 0.80$. The differences between the Chinese non-musicians and the three English groups might be due to that the pronunciations of the Chinese number words are shorter than the pronunciations of the English number words (Stigler, Lee & Stevenson, 1986).

In addition, participants’ musical aptitude was tested on site by using the Advanced Measures of Music Audiation (AMMA, Gordon, 1989). During the test, pairs of brief musical samples were presented to the participants and they were asked to indicate whether the samples in each pair were the same or different; if different, participants had to indicate whether the discrepancy is rhythmic or tonal in nature. This test provided a standardized and effective way of determining the participants’ musical abilities irrespective of music training or experience. Mean accuracy percentages of the
tonal test, the rhythmic test and the composite performance in AMMA for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians are presented in Table 2-1. The accuracy percentages were generated automatically by the AMMA program for each participant and were submitted to three one-way ANOVAs with group as the between-subject factor. The results showed that the effect of group was not significant for the tonal test \( F(3, 56) = .63, p = .60 \), the rhythmic test \( F(3, 56) = .50, p > .68 \), or the composite performance \( F(3, 56) = .79, p > .50 \), indicating that the four groups did not differ in tonal or rhythmic skills.

After the AMMA test, the participants did a behavioral discrimination task and a production imitation task on lexical tones. The order of the two tasks was counterbalanced across participants in order to prevent practice effects. In both tasks, stimuli were presented to participants via headphones (Sennheiser HD 280 PRO), using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The discrimination task employed a forced-choice AX paradigm, in which the stimuli were presented in pairs. Each trial contained two different tokens of the same syllable produced by one speaker, with the tokens being separated by inter-stimulus-interval (ISI) of 500 milliseconds. The ISI was set to 500ms because previous studies have shown that participants were sensitive to nonnative sound contrasts at the 500ms ISI (e.g. Cowan & Morse, 1986; Pisoni, 1973; Van Hessen & Schouten, 1992). The participants were asked to indicate whether the second stimulus had the same or different tone as the preceding one by pressing a button on a button box. In order to prevent the dominant-hand bias, half of the participants did the Same-Different task, in which the leftmost button on the button box indicated ‘same’ while the rightmost
denoted ‘different’; the other half of the participants had the reverse pattern. A short tutorial and 8 practice trials were presented before the experimental trials. In the tutorial, participants were presented with the syllable [ma] associated with the four Mandarin tones, which were produced by a different female native Mandarin speaker. While the participants were listening to the four tones, they were also presented with the corresponding tone figures, which helped them to understand the tone contours. The 8 practice trials consisted of different syllables from the experimental stimuli. During practice, participants received feedback after they made their decisions, and these data were dropped from analysis. Next, a total of 288 experimental trials was presented, with 144 ‘same’ and 144 ‘different’ trials. The discrimination of ‘same’ trials involves categorization skills because participants need to form categories for the pitch contours and to ignore other within-category acoustic and articulatory features. To the contrary, the discrimination of ‘different’ tone pairs might encourage participants to pay more attention to the within-category properties (e.g. Jamieson & Morosan, 1989). The participants were encouraged to respond as quickly and accurately as possible. No feedback was given in the experimental trials and participants’ responses automatically triggered the next trial after a 1500 millisecond inter-trial-interval. Each participant’s accuracy and reaction time were logged. A short break with at least 30 seconds was given halfway through the task, in order to prevent fatigue.

In the self-paced imitation task, the participants first heard one stimulus via headphones (Sennheiser HD 280 PRO) and were asked to imitate the sound immediately. After their imitation in each trial, they had to click the mouse to start the next trial. Before the experimental trials, the participants received a short tutorial on
Mandarin tone production and were given 8 practice trials which consisted of different syllables produced by the same female talker as in the tutorial. These practice data were dropped from analysis. After the practice trials, a total of 96 (6 syllables × 4 tones × 2 speakers × 2 tokens/speakers) experimental trials were presented in two blocks, with a 30-second break in between. All participants were recorded through a microphone (Shure SM58) and a digital recorder (Marantz PMD660) in a sound-attenuating booth. They were encouraged to repeat or correct whenever they feel necessary.

Lastly, the participants were asked to self-evaluate their performance and gave suggestions on both discrimination and imitation tasks. The entire experiment took about 1.5 hours per participant.

Data Analysis

Discrimination task

Data obtained in the discrimination task were converted to d-prime scores, which were calculated as the hit rates (correct responses in the ‘different’ trials) minus the false alarm rates (incorrect responses in the ‘same’ trials). Hit rates of 1 were corrected with 1-1/2N, where N equaled the total number of the ‘different’ trials. False alarm rates of 0 were replaced by 1/2N, where N equaled the total number of the ‘same’ trials. A one-way ANOVA with group as the between-subject factor was performed on the d-prime scores to see whether there were differences between the four groups in lexical tone discrimination. Furthermore, each participant’s mean reaction time (RT) was calculated only for the correct responses. For each participant the RT outliers that were beyond the mean RT of all the correct responses plus 2.5 standard deviations for that participant were replaced by this value. The RT data were then submitted to a one-way
ANOVA with group as the between-subject factor, in order to examine whether the English musicians discriminated tones more quickly than the English non-musicians. In addition, the accuracy percentages for each tone pair were also calculated for each participant. A repeated measure ANOVA was conducted on the accuracy percentages with tone pair as the within-subject factor and group as the between-subject factor. According to the literature (e.g. Roberts & Russo, 1999), all the $p$-values and the $F$-values were adjusted using the Greenhouse-Geisser correction and the post hoc paired $t$ tests were adjusted using the Bonferroni correction for multiple comparisons. The uncorrected degrees of freedom were reported in both the ANOVA tests and the $t$ tests. The result was considered statistically significant when the $p$-value was less than .05, and was considered as marginally significant when the $p$-value was between .05 and .10.

**Imitation task**

The production responses from each participant were segmented by using PRAAT 5.1.25 (Boersma and Weenink, 2010). Only the best production for each trial was selected for data analysis, resulting in a total of 96 productions per participant. For each trial the pitch contour of the best production best resembled the contour of the target stimulus, as determined by the experimenter using the PRAAT program.

Thirty-six native speakers of Mandarin Chinese (20 male, 16 female) who did not participate in the experiment were recruited as judges to identify the participants’ imitations. All judges were between the ages of 20 and 31 (M=22.28, SD=1.99) and with no reported speech or hearing impairment. Each participant’s productions were randomly assigned to 3 judges. The production was considered accurate when two of the judges had the same identification, which was consistent with the target stimulus.
The accuracy percentage for each tone type by each participant was calculated. A repeated measure ANOVA were conducted on the production accuracy data with tones as the within-participant factor and group as the between-participant factor.

Predictions

According to previous studies, we should find that the English musicians outperform the English non-musicians in both the discrimination task and the imitation task as a result of musical training. Moreover, the English vocalists should perform better than the English instrumentalists in the imitation task, since the vocalists have been trained to control their vocal pitch and to tune their sensations of production during their vocal training. If the vocal expertise benefits the perceptual skills, as suggested by the Motor theory (e.g. Liberman and Mattingly, 1989), we should also find that the English vocalists were superior to the instrumentalists in the discrimination task.

Results

Discrimination Task

Accuracy

Mean d-prime scores with the standard errors for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians are presented in Figure 2-2. The mean d-prime scores were 4.25 (SD = .72), 3.12 (SD = .40), 2.84 (SD = .63) and 2.41 (SD = .90) for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively. A one-way ANOVA was performed on the d-prime scores with group as the between-subject factor, showing that there was a main effect of group, $F (3, 56) = 19.56, p < .001$. Pairwise comparisons yielded that the Chinese non-musicians were more accurate in tone discrimination than the English vocalists ($p < .001$), the English
instrumentalists ($p < .001$) and the English non-musicians ($p < .001$). Moreover, the English vocalists performed significantly better than the English non-musicians ($p = .038$). However, the difference between the English instrumentalists and the English non-musicians ($p > .57$) and the difference between the English vocalists and the English instrumentalists ($p = 1.00$) did not reach significance.

To sum up, the English vocalists outperformed the English non-musicians in tone discrimination, but the English instrumentalists did not show such advantage.

**Reaction times**

Mean reaction times with the standard errors of all the correct responses for the four groups are illustrated in Figure 2-3. The mean reaction times were 1023ms (SD = 256.98), 1198ms (SD = 299.73), 1034 (SD = 430.17) and 1055ms (SD = 257.45) for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively. A one-way ANOVA with group as the between-subject factor showed that the effect of group was not significant, $F (3, 56) = .97, p > .41$. In sum, the four groups did not differ in the reaction times for the tone discrimination task.

**Accuracy percentage for each tone pair**

Because there are more 'different' pairs (Tone1-Tone2: pair12; Tone1-Tone3: pair13; Tone1-Tone4: pair14; Tone2-Tone3: pair23; Tone3-Tone4: pair34) than the 'same' pairs (Tone 1-Tone1: pair11; Tone2-Tone2: pair22; Tone3-Tone3: pair33; Tone4-Tone4: pair44), each 'same' pair had more experimental trials (i.e. 36 trials for each pair) than each 'different' pair (i.e.24 trials for each pair). Therefore, we decided to perform analyses on the 'same' pairs and the 'different' pairs separately. The accuracy percentages for each tone pair were calculated and were collapsed over the order of
presentation within the trials (e.g. pair12 consisted of Tone1-Tone2 and Tone2-Tone1) for each of the four groups. Figure 2-4 illustrates the accuracy percentages with the standard errors for the same pairs of the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians.

The accuracy percentages for each participant were submitted to a 4 (pair) × 4 (group) repeated measures ANOVA with pair as the within-subject factor and group as the between-subject factor. The result showed that the effect of pair \(F(3, 168) = 20.23, p < .001\), effect of group \(F(3, 56) = 8.46, p > .001\) and the interaction of pair × group \(F(9, 168) = 3.11, p = .004\) were significant. Pairwise comparisons revealed that for the ‘same’ pairs the Chinese non-musicians outperformed the English vocalists \((p = .005)\), the instrumentalists \((p < .001)\), and the English non-musicians \((p = .014)\). No difference was found between the three English groups. For the pair11, the Chinese non-musicians outperformed the English vocalists \((p = .005)\), and the instrumentalists \((p = .019)\); The English non-musicians performed equally well as the Chinese non-musicians on the discrimination of pair11 \((p = .44)\). No other difference was found in other comparisons between the four groups. For the pair22, the vocalists marginally outperformed the instrumentalists \((p = .067)\). The Chinese non-musicians were only superior to the English non-musicians \((p = .032)\) and the English instrumentalists \((p < .001)\), but not the English vocalists \((p = .58)\). Moreover, the English instrumentalists did not show an advantage on the discrimination of pair22 over the English non-musicians \((p > .97)\). For the pair33, no difference was found across all the four groups. For the pair44, the Chinese non-musicians outperformed the three English groups \((ps < .01)\), but the three English groups did not differ from each other \((ps = 1.00)\).
Figure 2-5 presents the accuracy percentages with the standard errors for each of the different pairs in the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians. A 6 (pair) × 4 (group) repeated measures ANOVA with pair as the within-subject factor and group as the between-subject factor yielded a main effect of pair \( [F(5, 280) = 14.75, p < .001] \), a main effect of group \( [F(3, 56) = 11.97, p < .001] \) and a marginally significant interaction of pair × group \( [F(15, 280) = 1.84, p = .06] \). Pairwise comparisons displayed that for the different pairs, the vocalists and the instrumentalists performed almost as accurately as the Chinese non-musicians \( (p = 1.00) \), and both musician groups outperformed the English non-musicians \( (ps < .001) \). No difference was found between the English vocalists and instrumentalists \( (p = 1.00) \). In addition, participants discriminated the pair12 and pair13 more accurately than the pair14 \( (ps < .001) \), pair23 \( (ps < .009) \), pair24 \( (ps < .04) \) and pair34 \( (ps < .001) \).

**Imitation Task**

Table 2-2 presents the mean percentages of participants’ correct productions that were agreed upon by all 3 judges and 2 judges respectively for each tone type and group. Figure 2-6 depicts the mean percentages and the standard errors of the correctly identified productions by the native Chinese judges for each tone across the four groups. The mean percentages of correct productions for all participant were submitted to a 4 (tone) × 4 (group) repeated measures ANOVA with tone as the within-subject factor and group as the between-subject factor. The result yielded a main effect of tone \( [F(3, 168) = 374.41, p < .001] \), a main effect of group \( [F(3, 56) = 6.98, p < .001] \), and a significant interaction of tone × group \( [F(9, 168) = 3.66, p = .002] \). Pairwise comparisons presented that the Chinese non-musicians outperformed the English
vocalists ($p = .024$), the English instrumentalists ($p = .004$) and the English non-musicians ($p = .001$). No difference was found between the three English groups. For Tone1 and Tone4, all four groups performed equally well ($ps > .32$). For Tone2, the Chinese non-musicians only outperformed the English non-musicians ($p = .007$). No difference was found between the three English groups. For Tone3, the Chinese non-musicians were superior to the English vocalists, the English instrumentalists and the English non-musicians ($ps < .001$).

The production accuracy for Tone3 was low even for the native Chinese speakers. This was probably due to the tone assimilation effect during the stimulus manipulation process. As described in the previous section, the experimental stimuli were segmented from the phrase “Qing shuo TARGET zhege zi” (please say TARGET this word). Because the word “zhe” which immediately follows the target word has a falling tone, the rising portion of the Tone3 might be assimilated by the following Tone4. Therefore, when the Tone3 was presented in isolation in the imitation task, the participants were more likely to perceive it as Tone4, and in turn produced it as Tone4. Figure 2-7 demonstrates the percentages of the participants’ productions of Tone3 that were identified as Tone4 by the judges.

**Summary**

This experiment compared native Chinese non-musicians, native English non-musicians, instrumentalists and vocalists on the perception and production of Mandarin tones. We expected that the English musicians should be better than the English non-musicians at both discriminating and imitating lexical tones. Moreover, the vocalists might have more advantage than the instrumentalists in imitating and discriminating lexical tones. Our results demonstrated that the English musicians were only better than
the English non-musicians at lexical tone discrimination, but not at lexical tone imitation. Specifically, the English vocalists and instrumentalists were better than the English non-musicians on the discrimination of the ‘different’ tone pairs but not the ‘same’ tone pairs. In addition, our results showed some indication that vocalists were better at discriminating lexical tones than instrumentalists (only for pair22). In terms of lexical tone production, the English musicians and non-musicians produced Tone1 (level tone) and Tone4 (falling tone) as accurately as the native Mandarin speakers, while produced Tone2 (rising tone) and Tone3 (falling-rising tone) less accurately than the native Mandarin speakers. Moreover, the English musicians did not differ from the English non-musician in the production of Tone2 and Tone3.
Table 2-1. Mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians

<table>
<thead>
<tr>
<th>Group</th>
<th>Tone</th>
<th>Rhythm</th>
<th>Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese non-musicians</td>
<td>67%</td>
<td>63%</td>
<td>66%</td>
</tr>
<tr>
<td>English vocalists</td>
<td>60%</td>
<td>64%</td>
<td>59%</td>
</tr>
<tr>
<td>English instrumentalists</td>
<td>60%</td>
<td>58%</td>
<td>58%</td>
</tr>
<tr>
<td>English non-musicians</td>
<td>64%</td>
<td>59%</td>
<td>62%</td>
</tr>
</tbody>
</table>

Table 2-2. Mean percentages of participants’ correct productions that were agreed upon by all 3 judges and 2 judges respectively for each tone type and group

<table>
<thead>
<tr>
<th>Group</th>
<th>Agreed by 3 judges</th>
<th>Agreed by 2 judges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Chinese non-musicians</td>
<td>78%</td>
<td>74%</td>
</tr>
<tr>
<td>English vocalists</td>
<td>69%</td>
<td>46%</td>
</tr>
<tr>
<td>English instrumentalists</td>
<td>61%</td>
<td>48%</td>
</tr>
<tr>
<td>English non-musicians</td>
<td>68%</td>
<td>42%</td>
</tr>
</tbody>
</table>

Figure 2-1. Waveforms and spectrograms with the pitch contours of the Mandarin Tone1, Tone2, Tone3 and Tone4 associated with the syllable [ma]. (Note: the blue lines indicate the pitch contours of the tones)
Figure 2-2. D-Prime Scores with the standard errors for Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively.

Figure 2-3. Mean reaction times with the standard errors of all the correct responses for the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively.
Figure 2-4. The mean accuracy percentages with the standard errors for the same pairs in the Chinese-non-musicians, the English vocalists, the English instrumentalists and the English non-musicians respectively.

Figure 2-5. Mean accuracy percentages with the standard errors for each of the different pairs in the Chinese non-musicians, the English vocalists, the English instrumentalists and the English non-musicians.
Figure 2-6. Mean percentages and the standard errors of the correctly identified productions for each tone in the Chinese non-musicians, the English vocalists, instrumentalists and non-musicians respectively.

Figure 2-7. Distribution of the participants’ identified production of T3. (Note: the tone before the dash refers to the target and the tone after the dash indicates the native judges’ identification of the participants’ imitation.)
CHAPTER 3
EXPERIMENT 2: EFFECTS OF LABORATORY PERCEPTION AND PRODUCTION TRAINING ON TONE DISCRIMINATION – A BEHAVIORAL AND ERP STUDY

Since the long-term perception training and production training in the music domain did not show significant difference in improving the perception and production of lexical tone, we further examined the effects of perception and production training in the language domain. The second experiment collected both behavioral and electrophysiological (ERPs) data to investigate the effectiveness of a laboratory perception-only training and a perception-plus-production training on the intentional and unintentional processing of tones by native English speakers.

Methods

Participants

A total of 22 native speakers of American English were recruited in the present study. None of the participants had participated in the Experiment 1. Participants were randomly assigned to two groups: 11 participants received only perception training (perception-only group); 11 participants received a perception plus production training (perception-plus-production group). All participants (2 males and 20 females) were between the ages of 18 and 25 years old ($M = 19.14$, $SD = 1.28$), and were undergraduate students at the University of Florida. None of the participants had previous exposure to any tone languages. All participants were right-handed according to Edinburgh Handedness inventory (Oldfield, 1971), had normal vision and had no history of speech or neurological impairment according to self-report. Participants’ hearing and auditory working memory (forward and backward digit span) were tested on site as described in Chapter 2. The mean auditory working memory accuracy percentages for the perception-only and the perception-plus-production groups were
63.31% (SD = 11.64%) and 56.49% (SD = 11.50%) respectively. There was no significant difference between the auditory memory scores for the two groups, \( t(20) = 1.38, p > .18 \). Moreover, none of the participants had received more than two years of formal musical training and had not been performing music within the past five years by the time of their participation. Participants’ musical aptitude was also tested on site by using the Advanced Measures of Music Audiation (AMMA, Gordon 1989) as described in Chapter 2. Table 3-1 presents the mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the perception-only and the perception-plus-production groups. The participants in the two groups did not differ in the musical tone perception \( [t(20) = -0.62, p > .54] \), the music rhythm perception \( [t(20) = -0.35, p > .73] \) or the composite performance \( [t(20) = -0.60, p > .55] \). An additional four participants were recruited, but were excluded from data analysis due to excessive eye or body movement artifacts or technical problems. All participants received financial compensation or course credit for their participation. The experimental procedures were approved by the University of Florida Institutional Review Board.

**Stimuli**

**Behavioral task**

Stimuli consisted of eight monosyllabic syllables ([pʰa], [pʰi], [kʰɛ], [kʰo], [tʰa], [tʰi], [tʰɛ] and [tʰo]) associated with three linear tones that resemble Mandarin Tone1 (high-level), Tone2 (high-rising), and Tone4 (high falling). Tone3 (low dipping) was not included because it has been shown to be the most confusable tone for both native Mandarin speakers and non-native speakers (e.g. Kirkham et al., 2011). The syllables [tʰa], [tʰi], [tʰɛ], [tʰo], [kʰo] and [pʰa] were used in the pre- and post-training tasks, and the syllables [pʰa], [pʰi], [kʰɛ] and [kʰo] were used in the training session. The syllables [tʰa],
[tʰi], [tʰɛ] and [tʰo] were not included in the training session because they were used to test whether participants could generalize the improvements gained through training to untrained stimuli. All of these syllables exist in American English, thus American speakers could focus their attention on the lexical tones and not be distracted by unfamiliar syllables. Two female native speakers of American English were asked to produce each syllable twice with a high pitch in a sound-attenuated booth, yielding four tokens for each syllable. Only female speakers were recorded for stimuli because adult female speakers have been rated to be more intelligible than male adult speakers (Bradlow et al., 1996). Three linear pitch contours were superimposed starting from the voiced part of each syllable, using the pitch-synchronous overlap and add (PSOLA) function in the Praat software (Boersma & Weenink, 2010). The procedures of stimulus manipulation were similar to Wong & Perrachione (2007b). The onset value of Tone1 (T1) was the mean fundamental frequency (F0) of all syllables produced by the two speakers. The offset of T1 was identical to its onset. Tone2 (T2) had the same ending point as T1, and its starting point was 26% lower than its ending point. Originally the onset of Tone4 (T4) was set to 10% higher than T1 and dropped by 82%, according to the values of Mandarin T4 reported in Shih (1988). Pilot data of seven participants, however, displayed a ceiling effect, i.e. there was little room for participants to improve through the training session. Then we decided to set the onset of Tone4 the same as Tone1 and its offset 26% lower than its onset. Therefore, the onset of T2 was identical to the offset of T4. All stimuli were then normalized for intensity. Except for F0 and intensity, all other acoustic features (e.g. duration and voice quality characteristics) were kept identical to the speakers’ original productions. As mentioned previously, each
syllable was produced twice by two speakers, resulting in four tokens for each syllable. The three tones were generated for each token, which means that (1) within each token the three tones had the same acoustic features except for F0; (2) the same tones for different tokens had the same F0, but different duration and voice quality. All stimuli were judged as perceptually natural by three native Mandarin speakers and three native American English speakers. A second pilot experiment with 8 participants was carried out to ensure that the discrimination of the tones in untrained native English speakers did not show a ceiling effect. Figure 3-1 shows the waveforms and spectrograms with the pitch contours of the Tone1, Tone2 and Tone4 of one token of the syllable [tʰu]. All the three sounds had the same duration, intensity and other acoustic features.

ERP

In the ERP experiment, the stimuli were the [tʰu] syllable associated with the three tones (Tone1, Tone2 and Tone4). These stimuli were not included in the behavioral tasks or the training sessions. The stimulus recording and manipulation were exactly the same as described in the behavioral task section. The mean duration of all the tokens was 553 millisecond (SD = 54.66). On average, pitch contour started at 108ms (SD = 14.18) after the stimulus onset.

The experiment consisted of three types of blocks: in the first block, T1 was the standard, and both T2 and T4 were the deviants; in the second block, T2 was the standard, with T1 and T4 being the deviants; in the third block, T4 was the standard, and T1 and T2 were the deviants (Näätänen, Pakarinen & Rinne, 2004). Within each block each token of the deviants was presented 25 times, resulting in 200 deviant stimuli (25 × 2 tokens/speaker × 2 speakers × 2 tones). These deviant stimuli were 20% of the standard stimuli and were presented in a pseudo random order, with at least two
standard stimuli preceding each deviant. Within each block there were also 100 clean standards (25 × 2 tokens/speaker × 2 speakers), which were not immediately preceded or followed by any deviants. These clean standards were compared with the deviant stimuli in data analysis. The offset-to-onset inter-stimulus interval was varied between 500 and 650ms to prevent participants’ automatic anticipation of stimulus onset. Each block was split equally into two sub-blocks, with each sub-block lasting for about 12 minutes. The order of the six sub-blocks was randomized across participants, with a different order being presented before and after the training session. Before the experimental blocks, all participants also received a practice block, in which each token of the three tones was equally presented. The practice block, which was excluded from data analysis, lasted for about 10 minutes, in order to familiarize the participants with the experimental setting.

**Procedure**

The experiment took place over the course of three consecutive days. All participants were asked to do the pre- and post-training tasks on the first and third days. On the second day, participants either received a perception-only training or a perception-plus-production training on tones. Note that testing production without perception is hardly possible; the two training types were therefore similar in terms of auditory exposure to the stimuli, but differed in the production component: the production training required participants to imitate the stimuli while the perception training had participants utter a word unrelated to the stimuli. In this way, the production training and perception training could be directly compared, with participants in both groups receiving the same exposure and attention to the same stimuli.
Day 1

Baseline EEG data were recorded on the first day from all participants in both groups. The ERP experiment used the passive oddball paradigm, in which participants were watching a silent movie in a sound attenuating booth while the stimuli were presented binaurally through inserted ear buds at a constant and comfortable hearing level. Participants were asked to ignore the sounds and focus their attention on the movie. After each block, they received comprehension questions regarding to the movie. In order to control the attention differences across groups, participants were also asked to count the number of times a particular object occurs in the movie. Only participants who answered more than 60% of the questions correctly, and were able to produce the correct count (plus or minus 3) were included for further data analysis. The movies in the pre- and post-training sessions were different. The order of the two movies was counterbalanced across participants, with half of the participants watching the first movie in the pre-training session while the other half of the participants watching the first movie in the post-training session.

The ERP experiment was divided into six blocks according to the three conditions. The order of the blocks was randomized among participants. For each participant, the order of stimulus presentation in each block was the same before and after training. Participants were encouraged to take breaks after each block to maintain concentration during blocks and to prevent fatigue.

After the ERP recording, participants did a behavioral discrimination task. The discrimination task used a forced-choice AX paradigm, in which the stimuli (the syllables \([t^h\alpha], [t^h\i], [t^h\epsilon], [t^h\o], [k^h\o] \) and \([p^h\alpha]) were presented in pairs and were separated by 500ms inter-stimulus-interval. The ISI was set to 500ms because previous studies have
shown that participants were sensitive to nonnative sound contrasts at the 500ms ISI (e.g. Cowan & Morse, 1986; Pisoni, 1973; Van Hessen & Schouten, 1992). In the different pairs the two stimuli in each pair had two different tones which were generated from the same token of the same syllable produced by one speaker. In the same pairs the two stimuli in each pair had the same tones, but were generated from different tokens of the same syllable produced by one speaker. The participants were asked to focus only on the tones and ignore other acoustic features of the stimuli and indicate whether the two stimuli in each pair had the same or different tones by pressing a button on a button box. For half of the participants the leftmost button on the button box indicated ‘same’ and the rightmost button indicated ‘different’, and the vice versa for the other half of the participants. This design ensured that the results were not biased by participants’ button pressing preference. A total of 288 trials (6 syllables) were presented, with 144 ‘same’ and 144 ‘different’ trials. The order of the trials was randomized for each participant. The experimental trials were preceded by a short tutorial on the three tones and 6 practice trials which were excluded from data analysis. The participants were encouraged to respond as quickly and accurately as possible. Their accuracies and reaction times were logged by the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). No feedback was given through this task. A break of at least 30 seconds was given halfway through the task, in order to prevent fatigue. The ERP and behavioral discrimination tasks on the first day took about 4.5 hours for each participant.
Day 2

On the second day, the participants received either a perception-only or a perception-plus-production training, which took about one hour. The structure of each trial is illustrated in Figure 3-2.

Perception-only training. Participants in the perception-only group: (1) first heard one stimulus; (2) heard another stimulus after 500ms; (3) determined whether these two stimuli had the same or different tones; (4) received feedback; (5) heard the first stimulus again and saw the tone and the graphic indication of the tone simultaneously; (6) Said ‘next’; (7) heard the second stimulus again and saw the tone and graphic indication at the same time; and (8) Said ‘Next’. The next trial started three seconds after the participant’s production in (8).

Perception-plus-production training. Participants in the production group received exactly the same training as participants in the perception-only group, except that they were asked to imitate the stimulus instead of saying ‘Next’ in steps (6) and (8). They were encouraged to imitate the tones as accurately as they could, and were allowed to try only once. In order to match the production training to the perception training, the participants in the production group did not receive any feedback about their imitations. All participants’ productions were recorded by a Marantz PMD660 digital recorder for assessment of the participants' performance, but these data were not used in data analysis.

For each participant, the layout of the ‘same’ and ‘different’ buttons on the button box was identical in the pre- and post-training tasks and the training session. Regardless of the training type, a total of 192 trials (the syllables [pa], [phi], [ke] and [ko]) were presented in the training session, with 96 ‘same’ and 96 ‘different’ trials. The
order of the trials was randomized for each participant. The experimental trials were preceded by 6 practice trials which familiarized the participants with the training process and were excluded from data analysis. The participants were encouraged to respond as quickly and accurately as possible. Their accuracies and reaction times in step (3) were logged by the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). A break of at least 30 seconds was given halfway through the training session. After the 30 second break, the participants could take another break as long as they wanted and started the second half of the training session by pressing any button on the button box. Originally, we planned to conduct two sessions of training during the course of two consecutive days, with each session took about one hour. However, our pilot data on two participants showed that the participants’ discrimination abilities improved significantly even after the first day of training. Therefore, we decided to only include one training session. This decision may result in some unexpected results, which will be discussed in Chapter 5.

Day 3

On the third day, both groups did the same ERP and behavioral tasks as described under Day 1. After the participants finished all tasks, they were asked to self-evaluate their performance and give suggestions on all the tasks. The entire experiment took about 10 hours for each participant.

Data Analysis

Behavioral task

Data obtained from the pre- and post-training discrimination tasks and the training session were converted to d-prime scores, which were calculated using the same formulas as described in Chapter 2. Separate d-prime scores for the trained and
untrained stimuli in the pre- and post-training tasks and the training sessions were also calculated, in order to examine participants’ generalization abilities. Besides the d-prime scores, participants’ reaction times (RTs) in the pre- and post-training tasks and the training session were also calculated to see whether the participants responded faster in the post-training tasks than in the pre-training tasks. The RTs were computed only for the correct trials. RTs that greater than the mean RT of all the correct responses plus 2.5 standard deviations for an individual participant were replaced by this value. After the outlier replacement, the mean RTs of the total stimuli, trained stimuli (the syllables [kʰə] and [pʰa]) and untrained stimuli (the syllables [tʰa], [tʰi], [tʰɛ] and [tʰo]) for each participant in the pre- and post-training tasks and the training session were computed. Furthermore, the percentage of accurate responses for each tone pair-type was also calculated in order to explore which tone pair-type showed the greatest improvement and which type was the most resistant for improvement for each group. These behavioral data were submitted to several repeated measure ANOVAs, which will be described in detail in the result section.

**ERP**

Continuous EEG was recorded from 39 Ag/AgCl scalp electrodes (Fp1/2, F7/8, F5/6, F3/4, Fz, FT7/8, FC5/6, FC3/4, FCz, T7/8, C5/6, C3/4, Cz, TP7/8, CP5/6, CP3/4, FPz, P7/8, P5/6, P3/4, Pz, O1/2) at 512 Hz. Horizontal eye movements were recorded by electrodes on the left and right outer canthi, while the vertical eye movements were recorded by electrodes placed above and below the right eye. Blinks and eye movements were rejected off-line from averaging on the basis of the EOG. The EEG was referenced to the right mastoid, off-line band-pass filtered from 0.16 to 30 Hz, and arithmetically re-referenced to the mean of both mastoids after recording. The
continuous EEG was divided into epochs of 900ms after the stimulus onset plus a
200ms pre-stimulus baseline. Table 3-2 presents the numbers and the percentages of
trials that were included in data analysis after artifact rejection in the pre- and post-
training tasks for the perception-only and perception-plus-production groups
respectively.

Artifact-free epochs were analyzed from -200 to 900ms relative to the onset of
the stimulus, using the 200ms window preceding the onset as a baseline. The grand-
averaged difference waves were generated for each of the six conditions by subtracting
the average responses to the clean standard stimuli from average responses to the
 corresponding deviant stimuli. The six conditions were (1) deviant T1 in standards T2
(T12); (2) deviant T1 in standards T4 (T14); (3) deviant T2 in standards T1 (T21); (4)
deviant T2 in standards T4 (T24); (5) deviant T4 in standards T1 (T41); and deviant T4
in standards T2 (T42). The mean amplitude of the time windows spanning the P2, the
MMN and the late negativity of the difference waves were computed for each channel,
participant and condition in the pre- and post-training tasks. Time windows for these
components were defined based on visual inspection of the raw ERPs and the
literature. The 100 – 200ms window and the 200 – 400ms window encompass the P2
and the MMN respectively. The mean amplitude of the late negativity was computed
between 500 and 800ms after stimulus onset. All these data were submitted to several
repeated measures ANOVAs which will be described in detail in the result section.

According to the literature (e.g. Roberts & Russo, 1999), all the p-values and the
F-values were adjusted using the Greenhouse-Geisser correction and the post hoc
paired t tests were adjusted using the Bonferroni correction for multiple comparisons.
The uncorrected degrees of freedom were reported in both the ANOVA tests and the t
tests. The result was considered statistically significant when the p-value was less than
.05, and was considered as marginally significant when the p-value was between .05
and .10.

Predictions

If production indeed facilitates perception, as implied by the Motor Theory (e.g.
Liberman & Mattingly, 1989), we should find that the participants who received
perception-plus-production training outperformed the participants in the perception-only
group in the post-training tasks. For the behavioral task, the improvement from pre-
training to post-training should be more pronounced in the perception-plus-production
group than in the perception-only group.

If the additional production during the perception-plus-production training had an
effect on unintentional auditory processing of the tones, the perception-plus-production
group should demonstrate a larger MMN. Both groups may also reveal increased P2
amplitude after training, which has been found in several studies (e.g. Atienza et al.,
2002). However, P2 was expected to be smaller for the perception-only group than for
the perception-plus-production group after training. Moreover, both groups might
demonstrate a larger late negativity in the post-training task, because previous studies
have shown that late negativity increased with more exposure to the auditory stimuli
(e.g. Shestakova et al., 2003). This late component has been associated with neural
processes of auditory rule extraction (e.g. Zachau et al., 2005).

To sum up, if participants in the perception-plus-production group outperformed
participants in the perception-only group after training, it supported the claim that
perception relied on production (e.g. Liberman et al., 1989; Fowler, 1986). If the
advantage of production training only occurred in the ERPs, this indicated that: (1) the short-term production training affected the unintentional perception, but neither of the trainings was long enough to affect the behavioral measures; or (2) the production training did not additionally improve attentive perception. Follow-up studies would be needed to tease these two possibilities apart. However, if the perception-only group yielded better results after training than the perception-plus-production group, the additional production training did not enhance perception, which was contrary to the Motor Theory.

Results

Behavioral Task

Accuracy

Mean d-prime scores with the standard errors of the pre- and post-training tasks and the training session for the perception-only and the perception-plus-production groups are presented in Figure 3-3. The mean d-prime scores of the pre-training task, training session and the post-training task were 1.79 (SD = 0.94), 2.55 (SD = 1.49) and 2.52 (SD = 1.25) for the perception-only group and 2.11(SD = 1.12), 2.55 (SD = 1.38) and 2.99 (SD = 1.59) for the perception-plus-production group respectively.

Furthermore, each participant’s mean d-prime scores for the trained and untrained stimuli in the pre- and post-training tasks were calculated separately for the perception-only and the perception-plus-production groups in order to examine whether the participants could generalize the improvements gained through the trained stimuli to the untrained stimuli. The trained stimuli included the syllables [kʰo] and [pʰa], which were used in the pre-training task, the training session and the post-training task. The untrained stimuli were the syllables [tʰa], [tʰi], [tʰε] and [tʰo], which were only tested in
the pre- and post-training tasks. In addition, the untested syllables [kʰɛ] and [pʰi] were only used in the training session. Since the untrained stimuli were not included in the training session, we first focused our analysis on the mean d-prime scores for the trained and untrained stimuli in the pre- and post-training tasks. Figure 3-4 illustrates the mean d-prime scores with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups.

The mean d-prime scores for the trained stimuli in the pre- and post-training tasks were 1.74 (SD = .91) and 2.53 (SD = 1.35) for the perception-only group, and 1.98 (SD = 1.05) and 2.88 (SD = 1.34) for the perception-plus-production group. For the untrained stimuli the mean d-prime scores in the pre- and post-training tasks were 1.84 (SD = .98) and 2.54 (SD = 1.22) for the perception-only group, and 2.16 (SD = 1.08) and 2.95 (SD = 1.55) for the perception-plus-production group.

The mean d-prime scores for each participant were submitted to a 2 (test time: pre- and post-training) × 2 (stimulus type: trained and untrained) × 2 (group) repeated measures ANOVA, with test time and stimulus type as the within subject factors and group as the between-subject factor. The result yielded a main effect of test time [$F(1, 20) = 53.51, p < .001$]. Paired-samples $t$ tests showed that both groups performed more accurately in the post-training task than in the pre-training task as a result of training [$t(10) = 6.04, p < .001$ and $t(10) = 4.60, p = .001$]. However, the test time × group interaction was not significant [$F(1, 20) = .22, p > .64$], suggesting that the two groups did not differ from each other in the extent of improvements. There was no difference between the mean d-prime scores for the trained and untrained stimuli [$F(1, 20) = 2.94$, $p > .05$].
indicating that the participants generalized the improvements gained through the trained stimuli to the untrained stimuli.

Moreover, each participant’s mean d-prime scores for the tested ([kʰo] and [pʰa]) and untested ([kʰɛ] and [pʰi]) stimuli in the training session were calculated separately for the perception-only and the perception-plus-production groups. Figure 3-5 illustrates the mean d-prime scores with the standard errors of the tested and untested stimuli in the training session for both groups. The mean d-prime scores of the tested and untested stimuli were 2.44 (SD = 1.55) and 2.60 (SD = 1.27) for the perception-only group, and 2.71 (SD = 1.38) and 2.37 (SD = 1.30) for the perception-plus-production group. The mean d-prime scores for each participant were submitted to a 2 (stimulus type: tested and untested) × 2 (group) repeated measures ANOVA, with stimulus type as the within subject factor and group as the between-subject factor. The result yielded a significant interaction of stimulus type × group, $F(1, 20) = 4.46, p = .047$. Further independent samples t tests showed that the two groups did not differ in the discrimination of the tested stimuli [$t(20) = -0.44, p > .66$] or the untested stimuli [$t(20) = .42, p > .68$]. In addition, paired-samples t tests revealed that the perception-only group discriminated the tested and untested stimuli equally well during the training session, $t(10) = -0.91, p > .39$. However, for the perception-plus-production group the difference between the tested and untested stimuli almost reached significance [$t(10) = 2.14, p = .058$], with the discrimination of the tested stimuli being more accurate than that of the untested stimuli. This was probably due to that the generalization ability was hindered by the participants’ inaccurate productions in the perception-plus-production group.
In sum, both the perception-only and the perception-plus-production groups improved on the tone discrimination abilities in the post-training task compared with the pre-training task. Moreover, both groups generalized the improvements gained through the trained stimuli to the untrained stimuli. During the training session, the perception-only group discriminated the tested and untested stimuli equally well while the perception-plus-production group discriminated the tested stimuli more accurately than the untested stimuli.

**Reaction times**

The participants’ reaction times were measured from the onset of the second stimulus in each trial and were rounded to full milliseconds. Mean reaction times with the standard errors of all the correct responses in the pre-training task, the training session and the post-training task for the perception-only and production-plus-perception group are illustrated in Figure 3-6. The mean reaction times in the pre-training task, the training session and the post-training task were 1317ms (SD = 321), 1762ms (SD = 620) and 1369ms (SD = 539) for the perception-only group, and 1539ms (SD = 357), 1904ms (SD = 395) and 1492ms (SD = 336) for the perception-plus-production group. The reaction times in the training sessions were longer than in the pre- and post-training tasks. This was probably because that the participants responded more carefully/slowly in a clear training situation in which they received feedback after they made their decisions. In addition, the training session took about one hour, which was much longer than the pre-and post-training tasks. Therefore, the participants might get tired of maintaining concentration during the training session, resulting in slower response times. In order to get rid of the this factor we performed further analysis first on the reaction times in the pre- and post-training tasks to test whether participants
discriminated tones more quickly in the post-training task than in the pre-training task as a result of training.

The mean response times for the trained and untrained stimuli in the pre- and post-training tasks were calculated separately for the perception-only and the perception-plus-production. Figure 3-7 illustrates the mean response times with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups. For the trained stimuli, the mean reaction times in the pre- and post-training tasks were 1343 (SD = 359) and 1369ms (SD = 525) for the perception-only group, and 1538ms (SD = 392) and 1530ms (SD = 362) for the perception-plus-production group. For the untrained stimuli, the mean reaction times in the pre- and post-training tasks were 1304 (SD = 305) and 1369ms (SD = 525) for the perception-only group, and 1540ms (SD = 348) and 1472ms (SD = 331) for the perception-plus-production group. The reaction times for each participant were submitted to a 2 (test time: pre- and post-training) × 2 (stimulus type: trained and untrained) × 2 (group) repeated measures ANOVA, with test time and stimulus type as the within-subject factors and group as the between-subject factor. The results showed that there was a significant interaction of test time × stimulus type × group, $F (1, 20) = 5.26, p = .033$. No other significant effect or interaction was found. Further paired samples $t$ tests on the trained and untrained stimuli did not show any significant difference between the reaction times in the pre- and post-training tasks for either group.

Additionally, the mean response times for the tested and untested stimuli in the training session were also calculated separately for the perception-only and the perception-plus-production groups. Figure 3-8 illustrates the mean response times with
the standard errors of the tested and untested stimuli in the training session for both groups. The mean reaction times of the tested and untested stimuli in the training session were 1755 ms (SD = 577) and 1770 ms (SD = 666) for the perception-only group, and 1921 ms (SD = 432) and 1883 ms (SD = 374) for the perception-plus-production group. The reaction times for each participant were submitted to a 2 (stimulus type: tested and untested) × 2 (group) repeated measures ANOVA, with stimulus type as the within-subject factors and group as the between-subject factor. The result did not yield any significant effect or interaction.

In sum, neither the perception-only training nor the perception-plus-production training made the participants discriminate tones more quickly in the post-training task than in the pre-training task. Moreover, there was no difference between the reaction times of the trained/tested stimuli and the untrained/untested stimuli for the two groups.

**Percentage accuracies of different tone-pairs**

Percentages of accurate discriminations for each tone pair (Tone1-Tone2: pair12; Tone1-Tone4: pair14; Tone2-Tone4: pair24; Tone1-Tone1: pair11; Tone2-Tone2: pair22; Tone4-Tone4: pair44) were calculated collapsed over the order of presentation within each trial for each participant. Analyses were only performed on the percentage accuracies in the pre- and post-training tasks, because (1) the training session had fewer trials than the pre- and post-training tasks, with 192 trials in the training session and 288 in the pre- and post-training tasks; (2) the syllables used in the training session were different from the syllables used in the pre- and post-training tasks. Therefore, we only focused on the pre- and post-training tasks in order to make direct comparisons between the perception-only and the perception-plus-production groups.
Figure 3-9 depicts the mean percentage accuracies of each tone pair in the pre- and post-training tasks for the perception-only and perception-plus-production groups. The percentage accuracy data were submitted to a 2 (test time: pre and post-training task) × 6 (pair) × 2 (group) repeated measures ANOVA with test time and pair as the within-subject factors and group as the between-subject factor. The result only yielded a main effect of test time, $F(1, 20) = 10.03$, $p = .005$, indicating that the participants discriminated the tone pairs more accurately in the post-training task than in the pre-training task. Even though the perception-only group shows more improvement in Figure 3-9, the repeated measures ANOVA did not yield any significant effect or interaction with group.

The percentage accuracies for each tone pair in the training session were also computed for the perception-only and the perception-plus-production groups respectively. Figure 3-10 presents the mean percentage accuracies with standard errors of each tone pair in the training session for both groups. The percentage accuracies in the training session were submitted to a 6 (pair) × 2 (group) repeated measures ANOVA with pair as the within-subject factor and group as the between-subject factor. The result yielded a marginally significant interaction of pair × group, $F(5, 100) = 2.47$, $p = .081$. Further independent samples $t$ tests only showed a marginally significant difference for the pair44 [$t (20) = -1.75$, $p = .095$], indicating that the perception-plus-production group discriminated pair44 more accurately than the perception-only group during the training session.

**Summary of behavioral results**

In the behavioral tasks, both the perception-only and the perception-plus-production groups improved on the tone discrimination abilities as a result of training.
Moreover, the participants in both groups generalized the improvements gained through the trained stimuli to the untrained stimuli after training. However, during the training session, the perception-only group discriminated the tested and untested stimuli equally well while the perception-plus-production group discriminated the tested stimuli more accurately than the untested stimuli. This was probably due to that during the perception-plus-production training the participants’ inaccurate productions hindered their generalization skills. In terms of reaction times, neither of the two groups discriminated tones more quickly in the post-training task than in the pre-training task. Moreover, there was no difference between the two groups in the reaction times of the trained/tested stimuli and the untrained/untested stimuli.

**ERP**

Figure 3-11 presents the ERPs of the frontal electrodes (F3, Fz and F4), the central electrodes (C3, Cz and C4) and the parietal electrodes (P3, Pz and P4) collapsed over all six conditions in the pre- and post-training tasks for the perception-only and perception-plus-production groups respectively. Descriptively, the ERPs for both groups showed a large P2 component, which was especially visible at the frontal and central electrodes between 100 and 200 milliseconds. Moreover, a MMN component was elicited for the deviant versus standard stimuli, with a prominent frontal distribution between 200 and 400 milliseconds. A late negativity component was also observed at the frontal and central electrodes between 500 and 800 milliseconds. Further analyses will focus on these three components. Figure 3-12 presents the ERPs of the Fz electrode for the six conditions in the pre- and post-training tasks for the perception-only and perception-plus-production groups respectively. Figure 3-13 shows
the deviant minus standard difference waves of the Fz electrode for the six conditions in
the pre- and post-training tasks for the two groups.

100 – 200ms window

According to the literature, we expected that the participants in both groups
should show an increased P2 as a result of training (e.g. Atienza et al., 2002). The
mean amplitudes of the difference waves between 100 and 200ms were computed for
four sites: right-frontal (RF) included the electrodes F4, F6, F8, FC4, FC6 and FT8; left-
frontal (LF) included the electrodes F3, F5, F7, FC3, FC5 and FT7; right-posterior (RP)
included the electrodes CP4, CP6, TP8, P4, P6 and P8; left-posterior (LP) included the
electrodes CP3, CP5, TP7, P3, P5 and P7. The mean amplitude differences were
submitted to a 2 (test time) × 2 (hemisphere: right and left) × 2 (anteriority: anterior and
posterior) × 6 (condition: T12, T14, T21, T24, T41, T42) × 2 (group) repeated measures
ANOVA with test time, hemisphere, anteriority and condition as the within-subject
factors and group as the between-subject factor. The result yielded significant
interactions of condition × group \[ F (5, 100) = 2.78, p = .048 \] and test time ×
hemisphere × anteriority × group \[ F (1, 20) = 8.53, p = .008 \]. Since the interaction of
condition × group was significant, further repeated measures ANOVAs were performed
for each condition separately, with test time, hemisphere and anteriority as the within-
subject factors and group as the between-subject factor. We only focused on the
significant intercept, main effect of group and significant interaction of test time × group.
For the \(T12, T24, \) and \(T41\) conditions neither the intercept nor the group effect was
significant. For the \(T14\) condition, there was a main effect of group, \(F (1, 20) = 6.13, p = .022\). The perception-only group showed a larger positivity for the deviants versus
standards than the perception-plus-production group. For the \(T21\) condition, the
intercept was significant \[ F (1, 20) = 4.88, p = .041 \], which suggested that the ERPs for the deviants were significantly more negative than those for the standards. Moreover, there was a significant interaction of test time \( \times \) group \[ F (1, 20) = 5.73, p = .027 \]. For the T42 condition, the effect of group was marginally significant, \( F (1, 20) = 3.40, p = .080 \). The perception-plus-production group revealed a larger positivity for the deviants versus standards than the perception-only group. However, as described previously, the pitch contours of the stimuli started approximately at 108ms after the stimulus onset; from Figure 3-13 we saw that the differences between the two groups started at or right after the stimulus onset. Therefore, these effect and interaction of group was probably due to some different acoustic features of the stimulus onset rather than the pitch difference.

**MMN**

Figure 3-12 demonstrates that the MMN component was more prominent in the T21, T24 and T41 conditions while less prominent or absent in the T12, T14 and T42 conditions. Moreover, previous literature has shown that the MMN is largest over the frontal electrodes (e.g. Näätänen & Michie, 1979; Alain, Woods & Knight, 1998). Figure 3-11 also presents that the MMN was larger at the frontal electrodes than at the parietal electrodes. Therefore, our analyses on the MMN will focus on the frontal electrodes.

The mean amplitudes of the difference waves between 200 and 400ms were computed for two sites: right-frontal (RF) included the electrodes F4, F6, F8, FC4, FC6 and FT8; left-frontal (LF) included the electrodes F3, F5, F7, FC3, FC5 and FT7. The mean amplitudes were submitted to a 2 (test time) \( \times \) 2 (hemisphere: right and left) \( \times \) 6 (condition: T12, T14, T21, T24, T41, T42) \( \times \) 2 (group) repeated measures ANOVA with test time, hemisphere and condition as the within-subject factors and group as the
between-subject factor. The results showed that there were main effects of test time \( F(1, 20) = 6.68, p = .018 \), and condition \( F(5, 100) = 5.13, p = .005 \). The MMN was \textit{smaller} in the post-training task than in the pre-training task. In addition, the intercept was also significant \( F(1, 20) = 13.66, p = .001 \), which means that the negativity for the deviants was significantly larger than that for the standards.

Six 2 (test time) \( \times \) 2 (hemisphere) \( \times \) 2 (group) repeated measures ANOVAs were performed for each condition with test time and hemisphere as the within-subject factor and group as the between-subject factor. No significant main effect or interaction was found for the \textit{T12, T14 and T24} conditions. The intercepts for the \textit{T21} \( F(1, 20) = 9.19, p = .007 \), \textit{T41} \( F(1, 20) = 36.63, p < .001 \) and \textit{T42} \( F(1, 20) = 4.33, p = .050 \) conditions were significant, indicating that the negativity for the deviants was significantly larger than that for the standards in these three conditions. No other main effect or significant interaction was found in the \textit{T21} and \textit{T42} conditions. In the \textit{T41 condition}, there was a significant interaction of hemisphere \( \times \) group \( F(1, 20) = 5.80, p = .026 \). The MMN was larger over left than right frontal sites in the perception-only group, and vice versa in the perception-plus-production group.

\textbf{Late negativity}

Figure 3-21 illustrated that the late negativity for the perception-only group increased in the post-training task compared to the pre-training tasks for the \textit{T12, T14, T21, T24 and T41} conditions. However, the late negativity for the perception-plus-production group decreased or remained the same in the post-training task compared to the pre-training tasks.

The mean amplitudes of the difference waves between 500 and 800ms were computed for the four regions: RF, LF, RP and LP. These mean amplitudes were
submitted to a 2 (test time) × 2 (hemisphere: right and left) × 2 (anteriority: anterior and posterior) × 6 (condition: T12, T14, T21, T24, T41, T42) × 2 (group) repeated measures ANOVA with test time, hemisphere, anteriority and condition as the within-subject factors and group as the between-subject factor. The result yielded a significant intercept \([F(1, 20) = 15.84, p = .001]\), a main effect of anteriority \([F(1, 20) = 10.64, p = .004]\), and significant interactions of hemisphere × group \([F(1, 20) = 6.37, p = .020]\), test time × anteriority \([F(1, 20) = 9.85, p = .005]\) and test time × hemisphere × anteriority × condition \([F(2.98, 59.60) = 2.88, p = .044]\). Moreover, the interactions of test time × hemisphere × group \([F(1, 20) = 3.71, p = .069]\), hemisphere × anteriority × group \([F(1, 20) = 4.10, p = .056]\) and hemisphere × anteriority × condition × group \([F(3.59, 71.76) = 2.33, p = .071]\) were marginally significant.

To investigate the source of the interactions, two 2 (test time) × 2 (hemisphere) × 6 (condition) × 2 (group) repeated measure ANOVAs were performed separately on the frontal electrodes (F3/4, F5/6, F7/8, FC3/4, FC5/6 and FT7/8) and posterior electrodes (P3/4, P5/6, P7/8, CP3/4, CP5/6 and TP7/8). Analysis on the frontal electrodes revealed a significant intercept \([F(1, 20) = 18.62, p = .000]\) and a significant interaction of hemisphere × group \([F(1, 20) = 6.56, p = .019]\). The negativity was larger over the left hemisphere in the perception-only group, but was larger over the right hemisphere in the perception-plus-production group. Analysis on the posterior electrodes yielded a significant intercept \([F(1, 20) = 4.88, p = .039]\) and a main effect of test time \([F(1, 20) = 4.57, p = .045]\), with an increased negativity after training.

**Summary of ERP results**

In the 100-200ms window the group effect was not significant for the T12, T21, T24, and T41 conditions. For the T14 condition, the perception-only group showed a
larger positivity for the deviants versus standards than the perception-plus-production group. For the \textit{T21 condition}, the ERPs for the deviants were significantly more negative than those for the standards. For the \textit{T42 condition}, the perception-plus-production group revealed a marginally larger positivity for the deviants versus standards than the perception-only group. However, these significant effect and interaction with group were probably due to some different acoustic features of the stimulus onset rather than the pitch differences, since the different effects started before the onset of the pitch contours.

In the 200-400ms window, the MMN was smaller in the post-training task than in the pre-training task. Moreover, for the \textit{T41 condition} the two groups differed in the lateralization of the MMN component, with the perception-only group showing a larger MMN at the left-frontal electrodes while the perception-plus-production group displaying a larger MMN at the right-frontal electrodes. Interpretations of these MMN results will be discussed in Chapter 5.

In the 500-800ms window, both groups showed an increased late negativity at the posterior electrodes in the post-training task compared to the pre-training task. This increased late negativity might suggest that both the perception-only and the perception-plus-production groups improved at the unintentional level of tone abstraction as a result of training.

In sum, analyses on the ERP components did not show any significantly different training effects between the perception-only and the perception-plus-production groups.
Table 3-1. Mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the perception-only and perception-plus-production groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Tone</th>
<th>Rhythm</th>
<th>Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception-only group</td>
<td>48%</td>
<td>43%</td>
<td>44%</td>
</tr>
<tr>
<td>Perception-plus-production group</td>
<td>54%</td>
<td>45%</td>
<td>49%</td>
</tr>
</tbody>
</table>

Table 3-2. The numbers and the percentages of trials that were included in data analysis after artifact rejection in the pre- and post-training tasks for the perception-only and perception-plus-production groups respectively

<table>
<thead>
<tr>
<th>Trials</th>
<th>Perception-only Pre-training</th>
<th>Post-training</th>
<th>Perception-plus-production Pre-training</th>
<th>Post-training</th>
</tr>
</thead>
<tbody>
<tr>
<td>T12_dv</td>
<td>58 (56%)</td>
<td>61 (59%)</td>
<td>56 (54%)</td>
<td>68 (66%)</td>
</tr>
<tr>
<td>T14_dv</td>
<td>62 (55%)</td>
<td>61 (54%)</td>
<td>64 (57%)</td>
<td>72 (64%)</td>
</tr>
<tr>
<td>T1_cst</td>
<td>69 (61%)</td>
<td>66 (58%)</td>
<td>62 (55%)</td>
<td>74 (65%)</td>
</tr>
<tr>
<td>T21_dv</td>
<td>71 (61%)</td>
<td>65 (56%)</td>
<td>65 (56%)</td>
<td>71 (62%)</td>
</tr>
<tr>
<td>T24_dv</td>
<td>62 (54%)</td>
<td>61 (53%)</td>
<td>63 (55%)</td>
<td>73 (64%)</td>
</tr>
<tr>
<td>T2_cst</td>
<td>65 (56%)</td>
<td>66 (58%)</td>
<td>60 (53%)</td>
<td>72 (63%)</td>
</tr>
<tr>
<td>T41_dv</td>
<td>63 (57%)</td>
<td>64 (57%)</td>
<td>61 (54%)</td>
<td>68 (60%)</td>
</tr>
<tr>
<td>T42_dv</td>
<td>68 (55%)</td>
<td>69 (56%)</td>
<td>65 (52%)</td>
<td>79 (64%)</td>
</tr>
<tr>
<td>T4_cst</td>
<td>61 (54%)</td>
<td>62 (54%)</td>
<td>61 (53%)</td>
<td>73 (64%)</td>
</tr>
</tbody>
</table>

Figure 3-1. Spectrograms of Tone1, Tone2 and Tone4 for the syllable [tʰu]. (Note: the mean duration of the stimuli was 553 milliseconds; the blue lines indicate the pitch contours of the tones; on average, the pitch contours started at 108 milliseconds (SD = 14.18) after the stimulus onset)
**Perception-only training:**

1. 500 ms ISI  
2. Stimulus1  
3. Stimulus2  
4. Same or Different  
5. Feedback  
6. Stimulus1 again  
7. Say ‘Next’  
8. Stimulus2 again  

   ![Graph showing steps for perception-only training]

**Perception-plus-production training:**

1. 500 ms ISI  
2. Stimulus1  
3. Stimulus2  
4. Same or Different  
5. Feedback  
6. Stimulus1 again  
7. Imitate Stimulus1  
8. Stimulus2 again  

   ![Graph showing steps for perception-plus-production training]

**Figure 3-2.** Step illustration of the perception-only training and the perception-plus-production training

**Figure 3-3.** Mean d-prime scores with the standard errors of the pre-training task, training session and the post-training task for the perception-only and the perception-plus-production group respectively
Figure 3-4. Mean d-prime scores with the standard errors for the trained and untrained stimuli in the pre- and post-training tasks for the perception-only and the perception-plus-production group respectively.

Figure 3-5. Mean d-prime scores with the standard errors for the tested and untested stimuli in the training session for the perception-only and the perception-plus-production group respectively.
Figure 3-6. Mean reaction times with the standard errors of all the correct responses in the pre-training task, the training session and the post-training task for the perception-only and production-plus-perception groups respectively.

Figure 3-7. Mean reaction times with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups respectively.
Figure 3-8. Mean reaction times with the standard errors of the tested and untested stimuli in the training session for the perception-only and the perception-plus-production groups.

Figure 3-9. Mean percentage accuracies with standard errors of each tone pair in the pre- and post-training tasks for the perception-only and perception-plus-production groups.
Figure 3-10. Mean percentage accuracies with standard errors of each tone pair in the training session for the perception-only and the perception-plus-production groups.
Figure 3-11. The ERPs of the frontal electrodes (F3, Fz and F4), the central electrodes (C3, Cz and C4) and the parietal electrodes (P3, Pz and P4) collapsed over all six conditions. A) Pre-training task for the perception-only group. B) Post-training task for the perception-only group. C) Pre-training task for the perception-plus-production group. D) Post-training task for the perception-plus-production group.
Figure 3-11. Continued
Figure 3-11. Continued
Figure 3-11. Continued
Figure 3-12. The ERPs of the Fz electrode in the pre- and post-training tasks for the perception-only and perception-plus-production groups respectively. A) T12 condition. B) T14 condition. C) T21 condition. D) T24 condition. E) T41 condition. F) T42 condition
Figure 3-12. Continued
Figure 3-12. Continued
Figure 3-13. Deviant minus standard difference waves of the Fz electrode for the six conditions in the pre- and post-training tasks for the two groups.
Figure 3-13. Continued
This experiment is a follow-up study of the second experiment. The second experiment was a little biased toward the perception-plus-production group, because the imitation task in the perception-plus-production training probably required the participants to pay more attention to the stimuli than the production task in the perception-only training. The third experiment was designed to be biased toward the perception-only group. If the perception-plus-production group could show more improvement than the perception-only group, it indicated that the perception-plus-production training was more effective than the perception-only training and the employment of the articulatory system benefits the perceptual skills. In this experiment, all participants were trained and tested on tone identification, instead of tone discrimination, over the course of two consecutive days. Participants were also divided into two groups: in the identification-only group the participants only received an identification training, while in the identification-plus-imitation group the participants received an identification training plus an imitation of the stimulus. The trainings were exactly the same except that the identification plus imitation training required participants to imitate the stimuli, while the identification only training had participants utter the tone type of the stimuli (i.e. level, rising or falling).

**Methods**

**Participants**

Twenty native speakers of American English were recruited in the current experiment, with 10 being assigned to the identification-only group and the other 10 in the identification-plus-imitation group. All participants (3 males and 17 females) were
between the ages of 19 and 23 years old (M = 20.05, SD = 1.12), and were undergraduate students at the University of Florida. None of the participants had previous exposure to any tone languages. Participants’ hearing, auditory working memory and music abilities were tested on site as described in Chapter 2. The mean accuracy percentages of the auditory working memory test were 58.57% (SD = 14.70%) and 53.57% (SD = 9.52%) for the perception group and the imitation group respectively. The two groups did not differ from each other in the auditory working memory test, \( t(18) = .90, p > .37 \). Table 4-1 presents the mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the identification-only and the identification-plus-imitation groups. The participants in the two groups did not differ in the musical tone perception \( t(18) = - 0.97, p > .34 \), the music rhythm perception \( t(18) = - 0.16, p > .87 \) or the composite performance \( t(18) = - 0.75, p > .46 \). None of the participants had participated in the first or the second experiment. Other recruitment criteria were exactly the same as described in Experiment 2. An additional 4 participants were recruited, but were dropped from the analysis either because they had displayed ceiling effect (percentage accuracy > 98%) in the pre-training task or because they did not meet the recruitment criteria. All participants received course credit for the participation. The experimental procedures were approved by the University of Florida Institutional Review Board.

**Stimuli**

Twelve monosyllabic syllables ([kʰa], [kʰɛ], [kʰi], [kʰo], [pʰa], [pʰɛ], [pʰi], [pʰo], [tʰa], [tʰi], [tʰɛ] and [tʰo]) associated with three linear tones (Tone1, Tone2 and Tone4) served as stimuli in the current experiment. Four more syllables ([kʰa], [kʰi], [pʰɛ] and [pʰo],) were included in the current experiment than in the Experiment 2 to ensure that the
training sessions had enough trials and were long enough to show the training effect. Except for these four syllables, other stimuli were also used in Experiment 2. The syllables \([k^h\varepsilon], [k^h\circ], [p^h\alpha], [p^h\iota], [t^h\alpha], [t^h\varepsilon]\) and \([t^h\circ]\) were used in the pre- and post-training tasks, and the syllables \([k^h\alpha], [k^h\varepsilon], [k^h\iota], [k^h\circ], [p^h\alpha], [p^h\varepsilon], [p^h\iota] \) and \([p^h\circ]\) were used in the training session. The syllables \([t^h\alpha], [t^h\iota], [t^h\varepsilon] \) and \([t^h\circ]\) were not included in the training session in order to test whether participants could generalize the improvements gained through the trained stimuli to the untrained stimuli. All of these syllables exist in American English, thus American speakers could focus their attention on the lexical tones and not be distracted by unfamiliar syllables. The stimulus recording and manipulation were exactly the same as described in Chapter 3.

**Procedure**

The experiment took place over the course of two consecutive days. On the first day, participants first did an identification task, in which they heard one stimulus and then identified whether this stimulus had a level, rising or falling tone by pressing the corresponding button on a button box. The button box was connected to the computer and the participants’ response automatically triggered the next trial, with a 3000ms inter-trial-interval. The order of the trials was randomized by the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). For all the participants, the leftmost button on the button box indicated the ‘level’ tone, the middle button indicated the ‘rising’ tone and the rightmost button indicated the ‘falling’ tone. Before the real experiment, participant received a short tutorial on the three tones through seeing the tone (F0) graphs associated with the three tone types. After the tutorial there were six practice trials with the syllable \([p^h\alpha]\), which ensured that the participants were clear about their task in the experiment. These practice trials were later dropped from the data analysis.
During the tutorial and the practice trials, the experimenter stayed in the experimental room, so that the participants could ask any question they had. A total of 96 (8 syllables × 3 tones × 2 speakers × 2 tokens/speaker) trials were presented in this identification task. The participants’ responses and reaction times were recorded by the software automatically. No feedback was given through this task. The identification task took about fifteen minutes for each participant.

After the identification task, the participants received either an identification-only or an identification-plus-imitation training, which took about an hour. The trainings were exactly the same except that the identification-plus-imitation training required the participants to imitate the stimuli; while the identification-only training had the participants say the tone types of stimuli (i.e. “level”, “rising” or “falling”). Figure 4-1 illustrates the structure of each trial in the identification-only and the identification-plus-imitation training respectively.

**Identification-only training.** The participants in the identification-only group: (1) first heard one stimulus; (2) identified whether the stimulus had a level, rising or falling tone; (3) received feedback and saw the tone graph associated with the correct answer; (4) heard the stimulus again and saw the tone and the graphic indication of the tone simultaneously; (5) Said the type of the tone (i.e. said “level”, “rising” or “falling”). The next trial started three seconds after the participant’s production in (5).

**Identification-plus-imitation training.** The participants in the identification-plus-imitation group received exactly the same training as the participants in the identification-only group, except that they were asked to imitate the stimulus instead of saying the tone type in step (5). The participants were encouraged to imitate the tones
as accurately as they could, and were only allowed to produce once. In order to match the identification-plus-imitation training to the identification-only training, the participants in the identification-plus-imitation group did not receive any feedback about their productions.

All participants’ productions were recorded by a Marantz PMD660 digital recorder for assessment of the participants’ performance. These production data were not analyzed. Regardless of the training type, a total of 192 trials was presented in the training session, which was divided into two blocks. Each block had 96 trials (8 syllables × 3 tones × 2 speakers × 2 tokens/speaker). As mentioned in the previous section, only the syllables [kʰa], [kʰɛ], [kʰi], [kʰo], [pʰa], [pʰɛ], [pʰi] and [pʰo] were used in the training session. The trials in both blocks were identical, but the orders of the trials were randomized in each block by the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). A break of at least 30 seconds was given after the first block in order to prevent fatigue.

On the second day, all participants did the same identification task as on the first day. The participants were encouraged to respond as quickly and accurately as possible in all the tasks. After the participants completed all the tasks, they were asked to fill out a debriefing questionnaire including their self-evaluations of their performances in the tasks and their suggestions. The entire experiment took about 2 hours for each participant.

**Data Analysis**

The percentages of correct identification for each participant in the pre-training task, the training session and the post-training task were computed. Then these percentages were converted to d-prime scores, according to the Table A5.2 Values of d’.
for m-interval forced choice or identification in Macmillan & Creelman (1991, p. 319). The percentage of correct identification equaled 100% were replaced by 99%, which was the maximum percentage that could be found in the Table A5.2. Therefore, the maximum d-prime score a participant could obtain was 3.62. This procedure for calculating d-prime scores was different from the procedures described in Experiment 1 &2, because there was no existing formula that could be used to directly calculate d’ for identification responses in the literature. Even though the d’ calculation procedures were different for the discrimination and identification paradigms, some researchers claimed that the d-prime scores could be compared directly (e.g. Keating, 2004). Furthermore, the percentages of correct identification for the trained stimuli (i.e. the syllables [kʰɛ], [kʰo], [pʰɛ] and [pʰi]) and the untrained stimuli (i.e. the syllables [tʰa], [tʰi], [tʰɛ] and [tʰo]) in the pre-training and the post-training tasks were calculated and were converted to d-prime scores separately. Besides the d-prime scores, each participant’s mean reaction times (RTs) for the correct responses in the pre-training task, the training session and the post-training tasks were computed. The RT outliers that were larger than the mean RT of all the correct responses plus 2.5 standard deviations for each participant were replaced by this value. After the outlier replacement each participant’s mean RTs for the total stimuli, the trained stimuli and the untrained stimuli in the pre-training task, the training session and the post-training task were calculated separately, in order to see whether the participants responded faster in the post-training task than in the pre-training task as a result of training. Moreover, the percentages of correct identification for each tone were calculated separately in order to investigate which tone had the greatest improvement and which tone was the most resistant for improvement.
All the data were then submitted to several repeated measures ANOVAs and paired \( t \) tests (when applicable) which will be described in more detail in the Results section below. The \( p \)-values were adjusted using the Greenhouse-Geisser correction, and the post hoc paired \( t \) tests were adjusted using the Bonferroni correction for multiple comparisons. The result was considered statistically significant when the \( p \)-value was less than .05. If the \( p \)-value was between .05 and .10, the result was described as marginally significant.

**Results**

**Accuracy**

Mean d-prime scores with the standard errors for the identification-only and the identification-plus-imitation groups in the pre-training task, the training session and the post-training task are illustrated in Figure 4-2. The mean d-prime scores of the pre-training task, the training session and the post-training task were 1.30 (SD = 0.98), 2.04 (SD = 1.05) and 2.11 (SD = 1.11) for the identification-only group and 1.35 (SD = 0.73), 2.33 (SD = 0.77) and 2.34 (SD = 0.92) for the identification-plus-imitation group.

Furthermore, each participant’s mean d-prime scores for the trained and untrained stimuli in the pre-training task, the training session and the post-training task were calculated separately in order to investigate whether the participants generalized the improvements gained through the trained stimuli to the untrained stimuli. The trained stimuli included the syllables \([k^h\epsilon],[k^h\o],[p^h\a] and [p^h\i]\), which were used in the pre-training task, the training session and the post-training task, while the untrained stimuli were the syllables \([t^h\a],[t^h\i],[t^h\epsilon] and [t^h\o]\), which were only tested in the pre- and post-training tasks. Besides, the syllables \([k^n\a],[k^n\i],[p^n\epsilon] and [p^n\o]\) were only used in the training session. Since the untrained stimuli were not included in the training session,
we first focused our analysis on the mean d-prime scores for the trained and untrained stimuli in the pre- and post-training tasks. Figure 4-3 presented the mean d-prime scores with the standard error of the trained and untrained stimuli in the pre- and post-training tasks for both groups. The mean d-prime scores for the trained stimuli in the pre- and post-training tasks were 1.26 (SD = .99) and 2.11 (SD = 1.15) for the identification-only group, and 1.34 (SD = .76) and 2.36 (SD = .85) for the identification-plus-imitation group. For the untrained stimuli the mean d-prime scores in the pre- and post-training tasks were 1.34 (SD = 1.00) and 2.14 (SD = 1.11) for the identification-only group, and 1.39 (SD = .81) and 2.30 (SD = .97) for the identification-plus-imitation group.

The mean d-prime scores for each participant were submitted to a 2 (test time: pre- and post-training) × 2 (stimulus type: trained and untrained) × 2 (group) repeated measures ANOVA, with test time and stimulus type as the within subject factors and group as the between-subject factor. The result showed that there was a main effect of test time \(F(1, 18) = 93.28, p < .001\), indicating that both groups performed more accurately in the post-training task than in the pre-training task. No other significant effect or interaction was found. There was no difference between the mean d-prime scores for the trained and untrained stimuli, suggesting that participants generalized the improvements gained through the trained stimuli to the untrained stimuli.

Furthermore, each participant’s mean d-prime scores for the tested ([kʰɛ], [kʰo], [pʰa] and [pʰi]) and untested ([kʰa], [kʰi], [pʰɛ] and [pʰo]) stimuli in the training session were calculated separately for the identification-only and the identification-plus-imitation groups. Figure 4-4 presents the mean d-prime scores with the standard errors of the
tested and untested stimuli in the training session for both groups. The mean d-prime scores of the tested and untested stimuli were 2.05 (SD = 1.09) and 2.08 (SD = 1.03) for the identification-only group, and 2.32 (SD = .79) and 2.34 (SD = .79) for the identification-plus-imitation group. The mean d-prime scores for each participant were submitted to a 2 (stimulus type: tested and untested) × 2 (group) repeated measures ANOVA, with stimulus type as the within subject factor and group as the between-subject factor. No significant effect or interaction was found, indicating that the two groups did not differ in the identification of the tested and untested stimuli in the training session.

In sum, participants in both the identification-only and the identification-plus-imitation groups improved on the tone identification abilities through the training session. Even though the experimental design was biased toward the identification-only group, the identification-only group did not show more improvement than the identification-plus-imitation group after training. Moreover, the participants in both groups generalized the improvements gained through the trained stimuli to the untrained stimuli.

**Reaction Times**

The participants’ reaction times were measured from the onset of the stimulus and were rounded to full milliseconds. Mean reaction times with the standard errors of all the correct responses in the pre-training task, the training session and the post-training for the identification-only and the identification-plus-imitation group are presented in Figure 4-5. The mean reaction times in the pre-training task, the training session and the post-training task were 1675ms (SD = 611.13), 1946ms (SD = 694.48) and 1587ms (SD = 804.97) for the identification-only group, and 1943ms (SD = 560.09),
1840ms (SD = 722.22) and 1510ms (SD = 567.42) for the identification-plus-imitation group. Because the training session had different stimuli and experimental procedure from the pre- and post-training tasks, we first performed analysis on the reaction times in the pre- and post-training tasks to test whether the participants identified tones more quickly in the post-training task than in the pre-training task as a result of training.

The mean response times for the trained and untrained stimuli were calculated separately for the identification-only and identification-plus-imitation groups. Figure 4-6 illustrates the mean reaction times with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups. For the trained stimuli, the mean reaction times in the pre- and post-training tasks were 1670 (SD = 661) and 1590ms (SD = 814) for the identification-only group, and 1894ms (SD = 524) and 1458ms (SD = 511) for the identification-plus-imitation group. For the untrained stimuli, the mean reaction times in the pre- and post-training tasks were 1677 (SD = 576) and 1586ms (SD = 800) for the identification-only group, and 1458ms (SD = 511) and 1568ms (SD = 636) for the identification-plus-imitation group. The reaction times in the pre- and post-training tasks were submitted to a 2 (test time: pre- and post-training) × 2 (stimulus type: trained and untrained) × 2 (group) repeated measures ANOVA, with test time and stimulus type as the within-subject factors and group as the between-subject factor. The results showed that there was a main effect of test time \( F(1, 18) = 14.11, p = .001 \) and a significant interaction of test time × group \( F(1, 18) = 6.24, p = .022 \). Two independent-samples t tests showed that the two groups did not differ in the reaction times for the trained stimuli \( t(18) = -0.84, p = .41 \) or for the untrained stimuli \( t(18) = -1.16, p = .26 \) before training. Further paired samples t tests revealed that only the
identification-plus-imitation group identified the tones more quickly in the post-training task than in the pre-training task \( t (9) = 4.71, p = .001 \); the identification-only group did not show any improvement with respect to reaction time \( t (9) = .89, p > .39 \) after training.

Moreover, the mean response times for the tested and untested stimuli in the training session were also calculated for the identification-only and the identification-plus-imitation groups respectively. Figure 4-7 presents the mean response times with the standard errors of the tested and untested stimuli in the training session for both groups. The mean reaction times of the tested and untested stimuli in the training session were 1927ms (SD = 659) and 1967ms (SD = 735) for the identification-only group, and 1842ms (SD = 688) and 1836ms (SD = 760) for the identification-plus-imitation group. The reaction times for each participant were submitted to a 2 (stimulus type: tested and untested) × 2 (group) repeated measures ANOVA, with stimulus type as the within-subject factors and group as the between-subject factor. The result did not yield any significant effect or interaction.

To sum up, although the experimental design was biased toward the identification-only group, the identification-plus-imitation group gained more improvement in terms of reaction time than the identification-only group after training. Moreover, the reaction times for the trained and untrained stimuli did not differ.

**Percentages of Correct Identification for Each Tone**

The mean percentages of correct identification for each tone type (Tone1: T1; Tone2: T2; and Tone4: T4) were computed separately for each participant. Because the training session had different procedure and stimuli from the pre- and post-training tasks, analyses were first performed on the percentages of correct identification in the
pre- and post-training tasks. Moreover, we did not perform analyses for the trained and untrained stimuli separately because the number of the experimental trials in the pre- and post-training tasks was too small to be separated for the trained and untrained stimuli and for each tone type.

Figure 4-8 illustrates the mean percentages of correct identification for each tone type in the pre- and post-training tasks for the identification-only and the identification-plus-imitation groups. For the identification-only group, the mean percentages of correct identification of the T1, T2 and T4 were 60.00% (SD = 33.59%), 74.69% (SD = 24.45%) and 66.88% (SD = 23.49%) in the pre-training task, and 85.63% (SD = 21.86%), 83.75% (SD = 17.79%), 76.56% (SD = 21.51%) in the post-training task. For the identification-plus-imitation group, the mean percentages of correct identification of the T1, T2 and T4 were 71.56% (SD = 26.86%), 70.63% (SD = 14.82%) and 68.44% (SD = 16.23%) in the pre-training task, and 86.56% (SD = 16.86%), 87.81% (SD = 10.97%), 86.56% (SD = 11.88%) in the post-training task. The mean percentages of correct identification for each participant were submitted to a 2 (test time: pre and post-training task) × 3 (tone type: T1, T2 and T4) × 2 (group) repeated measures ANOVA with test time and tone type as the within-subject factors and group as the between-subject factor. The result showed that there was a main effect of test time \( [F(1, 18) = 64.13, p < .001] \), indicating that participants in both groups identified all the three tones more accurately in the post-training task than in the pre-training task. No other effect or interaction was significant.

Furthermore, the mean percentages of correct identification for each tone type in the training session were also computed for the identification-only and the identification-
plus-imitation groups respectively. Figure 4-9 presents the mean percentages of correct identification with standard errors for each tone type in the training session for the two groups. The mean percentages of correct identification for the T1, T2 and T4 were 80.63% (SD = 20.77%), 83.75% (SD = 16.22%) and 81.09% (SD = 17.94%) for the identification-only group, and 87.66% (SD = 12.20%), 89.53% (SD = 6.83%) and 87.97% (SD = 8.90%) for the identification-plus-imitation group. The mean percentages of correct identification in the training session were submitted to a 3 (tone type) × 2 (group) repeated measures ANOVA with tone type as the within-subject factor and group as the between-subject factor. The result did not yield any significant effect or interaction, suggesting that the two groups did not differ in the identification accuracy for any of the tone type during the training session.

We further calculated the number of errors each participant made for each error type in the pre- and post-training tasks. Figure 4-10 depicts the error distribution for the identification-only and the identification-plus-imitation groups. Before training the participants were more likely to misidentify T1 as T2, T2 as T1, T2 as T4, and T4 as T1. After training participants made fewer errors for these error types: T1 as T2: \( t(19) = 4.56, p < .001 \); T2 as T1: \( t(19) = 2.88, p = .010 \); T2 as T4: \( t(19) = 2.59, p = .018 \); and T4 as T1: \( t(19) = 5.23, p < .001 \).

**Summary**

This experiment compared the effectiveness of a laboratory identification-only training and an identification-plus-imitation training on tone identification. The results demonstrated that the participants’ identification accuracy improved after both the identification-only and the identification-plus-imitation training. Moreover, the participants in both groups generalized the improvements gained through the trained
stimuli to the untrained stimuli. In terms of the reaction time, the identification-plus-imitation group identified the tones more quickly in the post-training task than in the pre-training task but the identification-only group did not show any improvement. These results indicated that even though the experimental design was biased toward the perception-only group, the identification-plus-imitation training was more effective to improve the tone identification than the identification-only training.
Table 4-1. Mean accuracy percentages of the tonal test, the rhythmic test and the composite performance in AMMA for the identification-only and identification-plus-imitation groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Tone</th>
<th>Rhythm</th>
<th>Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification-only group</td>
<td>48%</td>
<td>47%</td>
<td>46%</td>
</tr>
<tr>
<td>Identification-plus-imitation group</td>
<td>57%</td>
<td>48%</td>
<td>52%</td>
</tr>
</tbody>
</table>

Figure 4-1. Step illustration of the identification-only training and the identification-plus-imitation training
Figure 4-2. Mean d-prime scores with the standard errors for the identification-only and the identification-plus-imitation groups in the pre-training task, the training session and the post-training task respectively.

Figure 4-3. Mean d-prime scores with the standard errors for the trained and untrained stimuli in the pre- and post-training tasks for the identification-only and the identification-plus-imitation group respectively.
Figure 4-4. Mean d-prime scores with the standard errors of the tested and untested stimuli in the training session for the identification-only and the identification-plus-imitation groups respectively.

Figure 4-5. Mean reaction times with the standard errors of all the correct responses in the pre-training task, the training session and the post-training for the identification-only and the identification-plus-imitation group.
Figure 4-6. Mean reaction times with the standard errors of the trained and untrained stimuli in the pre- and post-training tasks for both groups respectively

Figure 4-7. Mean response times with the standard errors of the tested and untested stimuli in the training session for the identification-only and identification-plus-imitation groups respectively
Figure 4-8. Mean percentages of correct identification with the standard errors for each tone type in the pre- and post-training tasks for the identification-only and the identification-plus-imitation groups.

Figure 4-9. Mean percentages of correct identification with standard errors for each tone type in the training session for the identification-only and the identification-plus-imitation groups respectively.
Figure 4.10. Error distribution for the identification-only and the identification-plus-imitation groups. (Note: the tone before the dash refers to the correct answer while the tone after the slash indicates participants’ response)
CHAPTER 5
CONCLUSIONS

This dissertation sought to compare the effectiveness of perception training and production training on lexical tone perception and to examine whether the effects were domain specific or domain general.

Three main research questions were investigated in the present study:

- Are the effects of production training and perception training domain general or domain specific? That is, can the instrumental and vocal expertise percolate to lexical tone perception and production?
- Is laboratory perception-plus-production training more effective than perception-only training in facilitating lexical tone perception at the intentional level?
- Does perception-plus-production training affect the unintentional perception of lexical tones differently from perception-only training?

If the music expertise carries over to the language domain, we predicted that vocal musicians should have more advantage than instrumental musicians in imitating lexical tones. If production indeed facilitates perception as suggested by the motor theory (e.g. Liberman & Mattingly, 1989), the vocalists might also be superior to the instrumentalists in the perception of lexical tones. Moreover, the laboratory perception-plus-production training should be more effective than perception-only training in facilitating lexical tone perception at both intentional and unintentional levels. However, if the perception-only group yielded better results after training than the perception-plus-production group, the additional production training did not enhance perception, which was contrary to the Motor Theory.

In order to test these predictions, we conducted three experiments which were presented in Chapters 2, 3, and 4. The first experiment examined the production and perception training in the music domain while the second and third experiments focused
on the language domain. In the following sections, I will summarize the main findings of the three experiments and discuss the results in terms of the effectiveness of perception training and production training and the relationship between perception and production.

**Summary of Results**

Experiment 1 compared English vocalists and instrumentalists to see which type of musical training is more advantageous to lexical tone perception and production. Participants were tested in a same/different discrimination task and a production task. The results demonstrated that the English vocalists outperformed the English non-musicians in overall tone discrimination while the English instrumentalists did not exhibit such advantage. However, the English instrumentalists and vocalists did not show significant difference in lexical tone discrimination. Moreover, the English vocalists and instrumentalists discriminated the ‘different’ tone pairs (pair 12, pair13, pair14, pair23, pair24, and pair34) almost as accurately as the Chinese non-musicians, and both musician groups outperformed the English non-musicians. Nevertheless, the English musicians did not show any advantage in the discrimination of the ‘same’ tone pairs (pair11, pair22, pair33 and pair44) compared to the English non-musicians. For the pair22, the English vocalists were marginally more accurate than the English instrumentalists. Unexpectedly, we did not find any difference between the English musicians and the English non-musicians in lexical tone production according to the evaluation by native Mandarin speakers. Specifically, the English musicians and non-musicians produced Tone1 (level tone) and Tone4 (falling tone) as accurately as the native Mandarin speakers. Taken together, these results suggest that the music training improves the pitch perception in the language domain. Moreover, to some extent the vocal training in addition to instrumental training was an advantage for lexical tone
perception. However, lexical tone production was not significantly facilitated by either type of music training.

Experiment 2 collected both behavioral and electrophysiological (ERPs) data to examine the effectiveness of a perception-plus-production training and a perception-only training on the intentional and unintentional processing of tones by native English speakers. In the behavioral tasks, both the perception-only and the perception-plus-production groups improved on the tone discrimination abilities through the training session. Moreover, the participants in both groups generalized the improvements gained through the trained stimuli to the untrained stimuli. In the ERP tasks, the MMN was smaller in the post-training task than in the pre-training task. Moreover, for the $T41$ condition the two groups differed in the lateralization of the MMN component, with the perception-only group showing a larger MMN at the left-frontal electrodes while the perception-plus-production group displaying a larger MMN at the right-frontal electrodes. Besides, both groups showed an increased late negativity at the posterior electrodes in the post-training task compared to the pre-training task. In sum, analyses on the ERP components did not show any significantly different training effects between the perception-only and the perception-plus-production groups.

As described in previous chapters, Experiment2 was a little biased toward the perception-plus-production group, because the imitation task probably required the participants to pay more attention to the stimuli than the perception-only group. However, the perception-plus-production group did not gain more improvement through the additional imitation task compared with the perception-only group.
Experiment 3 was designed to be biased toward the perception-only group. If the perception-plus-production group could show more improvement than the perception-only group, it indicated that the perception-plus-production training was more effective than the perception-only training and the employment of the articulatory system benefits the perceptual skills. In Experiment 3, participants were trained and tested in an identification task, instead of a discrimination task. The results demonstrated that the tone identification accuracy improved in both the identification-only and the identification-plus-imitation groups after training. Moreover, the participants in both groups generalized the improvements gained through the trained stimuli to the untrained stimuli. In terms of the reaction time, the identification-plus-imitation group identified the tones more quickly in the post-training task than in the pre-training task while the identification-only group did not show any improvement. These results indicated that the identification-plus-imitation training was more effective to improve the tone identification than the identification-only training, even though the experimental design was biased toward the perception-only group.

**Discussion**

**Effects of Vocal Experience and Instrumental Experience on Lexical Tone Perception and Production**

Consistent with previous studies (e.g. Alexander et al., 2005; Lee & Hung, 2008), the current study demonstrated that the English musicians outperformed the English non-musicians on lexical tone perception, suggesting that the processing systems associated with lexical tone perception and musical pitch perception overlap to a great extent. Specifically, the English vocalists and instrumentalists outperformed the English non-musicians on the discrimination of the ‘different’ tone pairs, but not on the
discrimination of the ‘same’ tone pairs. In order to correctly discriminate the ‘same’ tone pairs, participants need to form categories for the pitch contours and to ignore other within-category acoustic and articulatory features. To the contrary, the discrimination of ‘different’ tone pairs might encourage participants to pay more attention to the within-category properties (e.g. Jamieson & Morosan, 1989). Therefore, our results suggested that even though the English musicians had acuter discrimination skills of lexical tone, their pitch categorization skills were not significantly better than those of the non-musicians. Our results were highly comparable with the findings in Wayland, Herrera & Kaan (2010), which demonstrated that although the musicians outperformed the non-musicians in the identification of non-linguistic pitch contours, their ability to quickly categorize the non-linguistic pitch contours across different tokens was not superior to that of non-musicians both before and after 3 days of identification training. Taken together, we concluded that music training only facilitated musicians’ discrimination skills but not the categorization skills of lexical tone.

In terms of the effects of different music experience, we found that the English vocalists and instrumentalists did not differ in the discrimination skills of lexical tone. Using behavioral and electrophysiological methods, Nikjeh (2006) examined the discrimination of non-linguistic pitch by vocalists and instrumentalists and also failed to find a difference between the two types of musicians. These results suggested that regardless of training type long-term music background benefited auditory pitch discrimination in both music and language domain. Interestingly, our result demonstrated that the discrimination of pair22 was marginally better in the English vocalists than in the English instrumentalists. This result seemed to provide some
indication that the vocal training was more effective than the instrumental training to improve the musicians’ categorization skills of lexical tone. But why did we only find the difference for pair22 rather than other tone pairs? One possible explanation for this question is that the English listeners are more familiar with the linear tones (Tone1 and Tone4) than the contour tones (Tone2 & Tone3), since the level and falling tones are very similar to the level and falling intonation used at sentence final position in English. It might be speculated that the different effect of music background only shows up for unfamiliar pitch contours rather than familiar pitch contours. Pair 33 did not show such difference because of the stimulus manipulation, which resulting in Tone 3 being more likely to be perceived as Tone4.

As for lexical tone production, the current data showed that the English musicians did not differ significantly from the English non-musicians as judged by native Mandarin speakers. Specifically, the English musicians and non-musicians produced Tone1 (level tone) and Tone4 (falling tone) as accurately as the native Mandarin speakers, while produced Tone2 (rising tone) and Tone3 (falling-rising tone) less accurately than the native Mandarin speakers. Moreover, the English musicians did not differ from the English non-musician in the production of Tone2 and Tone3. This result was inconsistent with previous studies (e.g. Amir, Amir, & Kishon-Rabin, 2003; Nikjeh, 2006), which demonstrated that regardless of training type music background enhanced pitch production abilities. The discrepancy between the findings in the current study and previous research might be attributed to the acoustic and articulatory features of lexical tone. Burnham & Mattock (2007) claimed that besides pitch, lexical tone comprises several other acoustic and articulatory features, including duration, amplitude and
phonation type. The English musicians might only have advantages in the production of
pitch, but not other acoustic or articulatory features associated with Mandarin tones.
Therefore, their productions were not judged as more accurate than that of the English
non-musicians by the native Mandarin speakers. Further acoustic analysis on
participants’ productions would be needed to examine the validity of this explanation.

In addition, the production of lexical tone did not differ in the English vocalists
and instrumentalists. This result was compatible with Nikjeh (2006) which also failed to
find significant differences between vocalists and instrumentalists in music pitch
production. Nikjeh (2006) argued that this result was probably due to that the self-paced
imitation task was too easy for the musicians to show the different training effect
between vocalists and instrumentalists.

Another possible explanation for the lack of distinction between the English
vocalists and instrumentalists in both the discrimination and imitation tasks is the effects
of different types of instrumental training. Spiegel & Watson (1984) reported that
musicians whose primary instrument was brass, string or woodwind had smaller
discrimination thresholds than musicians whose primary instrument was keyboard.
Spiegel & Watson claimed that the musicians who needed to tune their own instruments
might be more sensitive to small frequency changes than the musicians who did not
tune their own instruments. In the current study, most of the instrumentalists played
brass, string or woodwind instruments. Moreover, the vocal musicians also need to tune
their ‘instrument’, which were their articulatory apparatus and their sensation of
production. Therefore, the English vocalists and instrumentalists in the current study did
not differ in either the discrimination or the production of lexical tones. Further
investigations on vocalists and instrumentalists with same instrumental expertise should provide more insights on the different effects of vocal training and instrumental training on lexical tone perception and production.

**Effects of Laboratory Perception training and Production Training on Tone Perception**

Since the perception training and production training in the music domain did not show significant difference in improving the perception and production of lexical tone, we further proposed a behavioral and electrophysiological investigation to examine the effects of laboratory perception-only training and perception-plus-production training on tone perception. If production indeed facilitated perception, as implied by the Motor Theory (e.g. Liberman & Mattingly, 1989), we anticipated that the perception-plus-production training should be more effective to improve participants’ intentional and unintentional discrimination of tone than the perception-only training. To the contrary, if production relied on perception, as claimed by the Speech Learning Model (Flege, 1995a), we would expect that the perception-only training might be more effective to improve participants’ intentional and unintentional discrimination of tone than the perception-plus-production training. In order to investigate these predictions, we conducted two experiments using two training paradigms.

**Discrimination training paradigm**

In the behavioral tasks, both the perception-only and the perception-plus-production groups improved on the tone discrimination abilities as a result of training. Moreover, the participants in both groups generalized the improvements gained through the trained stimuli to the untrained stimuli. These results suggested that the additional production component in the perception-plus-production training did not result in more
improvement in the tone discrimination. There were several possible accounts for this result. First, it was possible that the participants’ own productions affected their perceptual learning. However, this seems to be unlikely since Adank et al. (2010) found that no matter whether people heard their own voice or not, vocal imitation significantly improved the comprehension of unfamiliar accent. Moreover, Kraljic & Samuel (2005) demonstrated that participants’ perceptual learning was not hindered by their own bad productions. Second, the lack of additional improvement in the perception-plus-production group may have been due to our training session being too short to show the effect of the production component. In other training studies (e.g. Wang et al., 1999; Wayland & Li, 2008; Baese-Berk, 2010), participants usually received at least two sessions of training with each session took at least one hour. In our study, the perception-plus-production training only took one hour, during which the participants were only allowed to imitate each stimulus once. Moreover, the participants did not receive any feedback on their productions. Therefore, the perceptual learning through the production process was very limited. Further research should include longer training sessions and provide feedback to participants’ productions (e.g. visual feedback) in order to really facilitate perceptual learning. Third, the lack of significant difference between the two groups may also be attributed to the individual differences, with some fast learners demonstrating large improvement while others revealing small or no improvement after training. An analysis on individual improvements would be necessary to further investigate the difference between the two training groups. Fourth, if the lack of distinction between the two groups is not due to the short training session or the individual differences, we could surmise that the production leaning might be hard
to transfer to the perceptual learning, indicating that speech perception does not rely on speech production.

In the ERP task, both groups showed a smaller MMN in the post-training task than in the pre-training task. This result was comparable with Kaan et al. (2008), which demonstrated that after training the MMN decreased for the low-falling tone in English speakers. Kaan et al. (2008) claimed that the larger MMN in the pre-training task was due to the differences in F0 onset rather than the F0 direction. The English speakers were more sensitive to the F0 onset (e.g. Gandour, 1983) before training and started paying more attention to the F0 direction after training. However, the detection of F0 direction was harder than the detection of F0 onset for the English speakers, therefore, the MMN amplitude decreased in the post-training task compared to the pre-training task. This explanation could account for the result in the current study as well. In the current study, there were six conditions in the ERP task: T12, T14, T21, T24, T41 and T42. In four (T12, T21, T24 and T42) of the six conditions, the deviants had different F0 onset from the standards, resulting in a larger MMN in the pre-training task in both the perception-only and perception-plus-production groups. From Figure 3-13 we also observed that the MMN was larger in the T21, T41 and T24 conditions than in the T12, T14 and T42 conditions respectively. Francis & Ciocca (2003) found that Cantonese listeners were more accurate to discriminate the tone pairs when the first token in the pair had a lower F0 (low-high order) than when the first token had a higher F0 (high-low order). Given these findings, it might be that, the MMN was more prominent if the pitch onset of the deviant stimuli was lower than the pitch onset of the standard stimuli; and the MMN was most obvious if the pitch offset of the deviant stimuli was lower than the
pitch onset of the standard stimuli. In the T21 and T24 conditions, the pitch onset of the deviant stimuli (T2) was lower than the pitch onset of the standard stimuli (T1 & T4). Therefore, the T21 and T24 conditions showed larger MMN than the T12 and T42 conditions. In the T41 condition, the pitch offset of T4 was lower than the pitch onset of T1, therefore, this condition showed the most prominent MMN. Future analysis on the different stimulus presentation order in the behavioral data may yield further insights on this account.

In addition, both groups showed an increased late negativity at the posterior electrodes in the post-training task compared to the pre-training task. This finding was inconsistent with Kaan et al. (2008), which presented a decreased late negativity after training. They claimed that the smaller late negativity in the post-training task might indicate less intentional reorientation from deviant stimuli to standard stimuli. In other studies, the late negativity has also been associated with neural processes of auditory rule extraction (e.g. Zachau et al., 2005). Therefore, the increased late negativity in the current study might suggest that both the perception-only and the perception-plus-production groups improved at the unintentional level of tone abstraction as a result of training.

Nevertheless, the perception-only and the perception-plus-production groups did not show any significantly different training effects in any of the ERP components. This result suggested that the additional production component in the perception-plus-production training did not result in more improvement in the tone discrimination at the unintentional level. More research is needed to determine the cause of the lack of distinction between the two types of training. However, these findings provide an
important first step in examining the effectiveness of perception training and production training on tone perception.

**Identification training paradigm**

Using discrimination training paradigm, the participants who received the perception-plus-production training did not show different improvements compared to the participants who received the perception-only training in the intentional and unintentional discrimination of tone. We further examined the effectiveness of these two types of training using an identification training paradigm. Conforming to the discrimination training paradigm, the results demonstrated that the tone identification accuracy improved in both the identification-only and the identification-plus-imitation groups after training. Moreover, the participants in both groups generalized the improvements gained through the trained stimuli to the untrained stimuli. One difference observed between the two training groups was that the identification-plus-imitation group identified the tones more quickly in the post-training task than in the pre-training task while the identification-only group did not show any improvement in terms of reaction time. It should be remembered that this experiment was a little biased toward the identification-only group because the identification-only training reinforced the tone types of the stimuli during the production process while the identification-plus-imitation training directed the participants to pay more attention to the sounds rather than the tone types. Even though the experimental design was biased toward the identification-only group, the identification-plus-imitation group showed more improvement suggesting that the identification-plus-imitation training was more effective to improve the tone identification than the identification-only training.
Final remarks

In the current study, two training paradigm yielded different results: using the discrimination training paradigm, the participants who received the perception-plus-production training did not show more improvements compared to the participants who received the perception-only training in the intentional and unintentional discrimination of tone; however, using the identification training paradigm, the participants who received the identification-plus-imitation training showed a larger decrease in reaction times than the participants who received the identification-only training in tone identification. Our results were inconsistent with some previous research. For example, Flege (1995b) trained native Mandarin speakers to identify the English word-final /t/ and /d/ using both identification training and discrimination training. He found that both training paradigms were equally effective in improving participants’ identification. Moreover, Wayland & Li (2008) trained native English and Chinese speakers to discriminate Thai mid- and low-tone contrast, using both identification and discrimination training. They also found that the two training procedures were equally effective in promoting the participants’ discrimination of the two Thai tones. Nevertheless, these two studies only focused on the accuracy data but did not analyze the reaction time data. Therefore, it was unclear whether reaction time could yield and different effects between the two training paradigms. Additionally, the identification training used in these two studies consisted of a two-alternative forced-choice identification procedure, which might be too easy for the participants to gain significant improvements. In our study, the identification training employed a three-alternative forced-choice procedure, which might be more effective to show the different training effects.
Our results suggested that the identification training was more effective than discrimination training in improving participants’ categorization skills. The discrimination training might encourage participants to pay more attention to within-category acoustic differences that are irrelevant to the key identify of the categories (e.g. Jamieson & Morosan, 1989), while the identification training encourages participants to form distinct categories and to focus more on the differences between categories. However, it was still unclear why different training paradigm presented different results in the current study. A possible account for this question is that the categorization ability might be crucial for production learning to be able to transfer to the perceptual modality. The discrimination training paradigm did not yield different results between the perception-only and the perception-plus-production groups is probably due to that the participants did not gain enough categorization skills through the discrimination training. Therefore, the transfer from the production learning to the perceptual learning did not take place. On the contrary, the participants who received the identification training gained enough categorization skills, which enable the production learning to take place and in turn to facilitate the perceptual learning. In one word, the categorization skill might be the prerequisite for the occurrence of the production learning, which in turn may facilitate the perceptual learning. By this account, speech production may rely on speech perception since the categorical learning precedes the production learning, which was contrary to the Motor theory. Future research can compare the effect of perceptual learning on speech production with the effect of production learning on perceptual learning in order to uncover the relationship between speech perception and production.
## APPENDIX A
### MUSICAL BACKGROUNDS FOR THE ENGLISH MUSICIANS IN EXPERIMENT 1

Table A-1. Musical backgrounds for the American English musicians in Experiment 1

<table>
<thead>
<tr>
<th>Musician number</th>
<th>Years of music training (vocal/instrumental)</th>
<th>Age onset of training (vocal/instrumental)</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>V01</td>
<td>4/10</td>
<td>18/12</td>
<td>Voice, piano</td>
</tr>
<tr>
<td>V02</td>
<td>7/16</td>
<td>14/5</td>
<td>Voice, piano</td>
</tr>
<tr>
<td>V03</td>
<td>10/NA</td>
<td>10/NA</td>
<td>Voice</td>
</tr>
<tr>
<td>V04</td>
<td>3/7</td>
<td>19/10</td>
<td>Voice, piano, guitar</td>
</tr>
<tr>
<td>V05</td>
<td>7/5</td>
<td>14/5</td>
<td>Voice, piano</td>
</tr>
<tr>
<td>V06</td>
<td>3/1</td>
<td>18/6</td>
<td>Voice, piano</td>
</tr>
<tr>
<td>V07</td>
<td>3/6</td>
<td>16/10</td>
<td>Voice, piano</td>
</tr>
<tr>
<td>V08</td>
<td>15/NA</td>
<td>6/NA</td>
<td>Voice</td>
</tr>
<tr>
<td>V09</td>
<td>14/3</td>
<td>11/8</td>
<td>Voice, piano</td>
</tr>
<tr>
<td>V10</td>
<td>12/1</td>
<td>6/5</td>
<td>Voice, piano, guitar, ukulele, vibraphone</td>
</tr>
<tr>
<td>V11</td>
<td>14/15</td>
<td>16/8</td>
<td>Voice, piano, guitar</td>
</tr>
<tr>
<td>V12</td>
<td>6/10</td>
<td>16/6</td>
<td>Voice, piano, guitar, ukulele</td>
</tr>
<tr>
<td>V13</td>
<td>10/3</td>
<td>11/7</td>
<td>Voice, guitar, piano</td>
</tr>
<tr>
<td>V14</td>
<td>9/17</td>
<td>14/5</td>
<td>Voice, piano, flute</td>
</tr>
<tr>
<td>V15</td>
<td>8/7</td>
<td>15/11</td>
<td>Voice, saxophone, flute, clarinet</td>
</tr>
<tr>
<td>I01</td>
<td>NA/10</td>
<td>NA/9</td>
<td>Flute, piccolo</td>
</tr>
<tr>
<td>I02</td>
<td>NA/13</td>
<td>NA/6</td>
<td>Piano, clarinet</td>
</tr>
<tr>
<td>I03</td>
<td>NA/10</td>
<td>NA/11</td>
<td>Clarinet</td>
</tr>
<tr>
<td>I04</td>
<td>NA/6</td>
<td>NA/9</td>
<td>Piano, flute, guitar</td>
</tr>
<tr>
<td>I05</td>
<td>NA/4</td>
<td>NA/14</td>
<td>Guitar</td>
</tr>
<tr>
<td>I06</td>
<td>NA/8</td>
<td>NA/10</td>
<td>Trumpet</td>
</tr>
<tr>
<td>I07</td>
<td>NA/9</td>
<td>NA/11</td>
<td>Clarinet, piano</td>
</tr>
<tr>
<td>I08</td>
<td>NA/7</td>
<td>NA/11</td>
<td>Trumpet, flute</td>
</tr>
<tr>
<td>I09</td>
<td>NA/3</td>
<td>NA/11</td>
<td>Flute, piccolo</td>
</tr>
<tr>
<td>I10</td>
<td>NA/10</td>
<td>NA/10</td>
<td>Trumpet, sax, euphonium, clarinet, piano</td>
</tr>
<tr>
<td>I11</td>
<td>NA/12</td>
<td>NA/10</td>
<td>Horn, piano</td>
</tr>
<tr>
<td>I12</td>
<td>NA/11</td>
<td>NA/8</td>
<td>Piano, flute, piccolo</td>
</tr>
<tr>
<td>I13</td>
<td>NA/12</td>
<td>NA/6</td>
<td>Piano, flute, guitar</td>
</tr>
<tr>
<td>I14</td>
<td>NA/6</td>
<td>NA/11</td>
<td>Trumpet, bass guitar</td>
</tr>
<tr>
<td>I15</td>
<td>NA/12</td>
<td>NA/7</td>
<td>Piano, saxophone, clarinet, bass guitar</td>
</tr>
</tbody>
</table>
APPENDIX B
LANGUAGE AND EDUCATION QUESTIONS

Note: Your participation is voluntary; you may decline to answer questions.

Age:___ Gender: M / F Hispanic? Y / N Race: ______________

Handedness: R / L / both

**Education**

1a. What is the highest level of education you completed (e.g., “High school”, “two years of college”, “one year of grad school”, “Ph.D”)?

1b. What is your major (if declared)?

**Language**

2. Please indicate what languages you speak (including English), whether you learned them at school and at what age/grade you learned them. Next, for each language, rate how fluent you are in writing, reading, speaking and listening, using a scale from 1 (not fluent at all) to 5 (native or near native).

<table>
<thead>
<tr>
<th>Language</th>
<th>Learned at school?</th>
<th>At home?</th>
<th>Age/grade acquired</th>
<th>Writing</th>
<th>Reading</th>
<th>Speaking</th>
<th>Listening</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Which language did you learn first as a child? __________

4. Do you consider yourself bi(multi)lingual? ______

5. Parents’ languages spoken and ages acquired:
   Mother: ___________________________________________ Age acquired: __________
   ___________________________________________ Age acquired: __________
   ___________________________________________ Age acquired: __________

   Father: ___________________________________________ Age acquired: __________
   ___________________________________________ Age acquired: __________
   ___________________________________________ Age acquired: __________
6a. If you were not born in the US, at what age did you move to the States? __________
6b. In which country or countries did you live before you moved to the US? ____________________________________________________________

7. If you were born in China, what part of China are you from? ________________
What part of China are your parents from? __________________________________
What dialects of Chinese do you speak? __________________________________

8. Have you had, or do you have, any known problems with reading, hearing, speech?_______________If yes, please explain: __________________________________

Music experience
9. Do you have perfect (absolute) pitch? Y / N

10a. Have you played or do you play a musical instrument (including voice) or have you ever received musical training? Y / N

b. If ‘yes’: Please complete the following table. You can use the last column and the space below for comments

<table>
<thead>
<tr>
<th>Instrument, Including voice</th>
<th>Started at age</th>
<th>Stopped at age</th>
<th>Started with formal training at age</th>
<th>Stopped with formal training at age</th>
<th>Still playing?</th>
<th>Still receiving formal training?</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

133
APPENDIX C
PERSONAL HISTORY FORM FOR LANGUAGE EXPERIMENTS WITH EVENT-RELATED POTENTIALS OR FMRI

The following form contains questions about your health. You may request that this form be seen only by the Principal Investigator in charge of the project. If you do not request this, the form may also be seen by the research assistant.

If you request that the form be seen only by the Principal Investigator, check here: _____
Health related questions.

Note: Your participation is voluntary; you may decline to answer questions.

Please check the appropriate response.

1a. Are you taking any medications?  
\[
\begin{array}{cc}
\text{Yes} & \text{No} \\
( ) & ( ) \\
\end{array}
\]

2b. If ‘yes’, are these medications that may affect the nervous system (e.g. beta-blockers, anti-depressants, epilepsy medication)?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

2. Do you easily get an allergic reaction on your skin?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

3. Are you allergic to band-aids, or stickers on your skin?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

4. Are you allergic to dust mite?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

5. Is your corrected vision significantly less than 20/20?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

6. Is your hearing ok, as far as you know?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

7a. Have you ever had a head injury or have you been knocked unconscious?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

b. If ‘yes’, were you hospitalized?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

8. Have you ever had a seizure?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

9. Have you ever suffered from a neurological or psychological disorder (e.g. epilepsy, depression, schizophrenia)?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

10. Have you ever been diagnosed with a language or reading-related problem (e.g. dyslexia)?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

11a. Have you participated in other scientific studies?  
\[
\begin{array}{cc}
( ) & ( ) \\
\end{array}
\]

b. If ‘yes’, when and which?
APPENDIX D
HANDEDNESS QUESTIONNAIRE

Note: Your participation is voluntary; you may decline to answer questions.

Which is your dominant hand?_________________________________________
Have you ever had any tendency to change your preferred hand?______________
Do you have any immediate family members that are left handed (parents, siblings?)

Please indicate your preferences in the use of hands in the following activities by putting + in the appropriate column. Where the preference is so strong that you would never try to use the other hand unless absolutely forced to, put ++. If in any case you are really indifferent, put + in both columns. Please try to answer all questions, and leave only a blank if you have no experience at all of the activity or task.

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Writing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Drawing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Throwing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Scissors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Toothbrush</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Knife (without fork)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Spoon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Broom (upper hand)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Striking match (match)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Opening box (lid)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

i. Which foot do you prefer to kick with?

ii. Which eye do you use when using only one?
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Shuang Lu was born in Jinan, Shandong Province, China. She received her B.A. in English from Shandong Jianzhu University in 2005. She earned her M.A. in Linguistics from Leiden University in the Netherlands in 2006. She joined the University of Florida as a Ph.D. student in the fall of 2008 and received her Ph.D. in Linguistics in the summer of 2013.