# Categorical Perception of Lexical Tones: ERP and Behavioral Experiments

ZHENG, Hong-Ying

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in

Electronic Engineering



January 2010

*To my parents,*

*Tianhe ZHENG and Meiyu CHEN,*

*To my husband,*

*Jianyong, CHEN*

*for their love, considerations, and supports.*

# ACKNOWLEDGEMENTS

This thesis is a synthesis of the research during my PhD study at both City University of Hong Kong (CityU) and The Chinese University of Hong Kong (CUHK). The most important person in my academic life is my supervisor, Prof. William ShiYuan WANG. All the words in the world do not suffice to thank him. His special wisdom led me into the beautiful realm of linguistics. His comprehensive lectures on speech perception inspired me to choose my topic on perception on lexical tones. He has guided me to learn how to appreciate the beauty and cherish the joy of doing research in basic science, and his encouragement has been all along supporting me to overcome various difficulties in research. He encouraged me to go into the new, exciting area of neuroscience. Without his recommendation and support, I would never have the chance to attend several international conferences and summer schools during the past four years or have the chance to visit the University of California, Berkeley during the summer in 2008.

I would like to extend my gratitude to Prof. Peter WM TSANG, who acted as my co-supervisor during my study in CityU, and sincerely supported me to continue my research when Prof. WANG leaves CityU. Especial thanks should be given to Prof. Richard IVRY and Prof. Paul KAY, who kindly hosted me when I visited the University of California, Berkeley. The working experience with them enriched my knowledge about psychology. More importantly, they helped me to raise the ability to do research work scientifically, critically, and creatively. I also want to thank Prof. Thomas Hun-Tak LEE, whose great knowledge about language acquisition and critical comments on my work are valuable for me to reconsider the experimental design.

I own a lot to the technical support and moral encouragement given by all the members of LEL. It was my great honor to collaborate with Dr. James W. MINETT and Dr. Gang PENG. Their rich experiences and critical challenge greatly inspired me to fulfill my research. I am grateful to Dr. Jin-Yun KE, Dr. Tao GONG, Dr. Francis Chunkit WONG, and Ms. Yao YAO to help me to improve my awkward English and to reorganize my paper. In addition, I am especially grateful to Dr. Tao GONG, Dr. Ching-Pong AU and Ms. Ruo-Xiao YANG who kindly helped me to collect some field work data presented in this thesis. Furthermore, I enjoyed a lot discussion with my other colleagues, Dr. Susan Lan SHUAI, Mr. Yingwei WONG, Mr. Raymond Weiwen WU, Mr. Guang-Ting MAI, Mr. Manson

FONG, Ms. Lin ZHOU, and Ms. Hui CHEN on various research topics. I am very fortunate to have these close amigos in my life.

In the end, I would like to thank the members in a bigger family, the Digital Signal Processing and Speech Technology Laboratory (DSPSTL). The leaders of this laboratory, Prof. Pak-Chung CHING, Prof. Tan LEE, and Prof. Ken Wing-Kin MA have generously supported me to perform my basic science research in this engineering laboratory. And I enjoyed working, growing, and improving my skills with the colleagues in DSPSTL, Dr. Meng YUAN, Ms. Yujia LI, Dr. Nengheng ZHENG, Ms. Ning WANG, Dr. Feng TIAN, Ms. Houwei CAO, Mr. Jiaxian PAN, Mr. Li QIANG, Dr. Yvone Siuwa LI, and Ms. Natalie Li ZENG. Without these friends, my PhD life in CUHK would have had lost many colors.

# CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Speech sounds vary across different conditions and subjects; nevertheless, listeners perceive the phonemes without difficulties. *Categorical perception* (CP) occurs when listeners map the varying speech sounds into discrete phonemic categories. In CP, to discriminate a pair of stimuli that cross a category boundary is much easier than those that lie within the same category, even though both pairs are separated by an equal physical difference. CP is one of the important properties essential for speech perception.

Pitch contour or its acoustic correlate, *fundamental frequency* (F0), distinguishes lexical meanings in tone languages. Two topics on CP of lexical tones were studied in the thesis: (1) the factors influencing CP, and (2) the temporal process of CP. These two topics were investigated through both behavioral and *event-related-potential* (ERP) methods on Cantonese and Mandarin tones.

Four factors were studied. They were (1) intrinsic acoustic properties of pitch contours by comparison between continua of level tones and contour tones; (2) positions of target syllables relative to context (without contextual sentence, at the beginning and at the end of the contextual sentence); (3) language backgrounds by comparison between listeners with different tone experiences; and (4) carrier syllables (real word, non word, and nonspeech). Three temporal stages were studied in the same experimental paradigm. They were (1) the preattentive stage investigated through the *mismatch negativity* (MMN); (2) the attentive stage investigated through the P300; and (3) the overt response stage investigated through the hit rate data.

All these four factors influence the degree of CP. In the discussion, both general auditory processing and language specific processing are suggested to be responsible for the various types of exhibition of CP, although they have different weights for different factors. Different patterns of CP were also observed in three temporal stages due to different weights of these two types of processing. In summary, a multistage model which includes both general auditory processing and language specific processing is proposed to explain the CP of lexical tones. This model improves previous models by proposing that the weights of these two types of processing in speech perception depend on the types of factors, and the temporal processing stages.

Finally, for the first time in the literature, the thesis also reported that even though a tone contrast (i.e., level vs. rising) is present in both tone systems, the same contrast is perceived differently by the two groups of subjects by virtue of their different language experiences.

# 摘要

同一句話，不同人或者相同的人在不同場合下發出的語音信號都是不一樣的，但是聽衆總能夠毫不困難地識別出這句話。語音的最小單位是音素。當人們將所聽到的變化多端的語音信號對應到有限的幾個音素上時，一種稱爲範疇感知的現象出現了。範疇感知意味著兩种類別交界處的分辨率遠遠高於類別內部的分辨率。範疇感知是語音感知的一個重要特性。

在聲調語言中，聲音的音高或者基頻曲綫也可以用來區別語義，因此也是一種音素。本論文討論兩個關於聲調範疇感知的議題：(1) 影響聲調範疇感知的因素 (2) 聲調範疇感知的時間動態特性。這兩個議題將通過行爲和事件相關電位兩种實驗手段進行研究。研究的對象是普通話和廣東話（粵語）的聲調。

本論文研究了四种因素。它們分別是(1) 基頻曲綫的內在聲學特性，包含了平聲調和曲折調 ；(2) 目標音節在上下文句子中的位置，包含孤立詞（沒有上下文），在句子開頭，和在句子結尾這三种情況；(3) 被試的聲調語言背景；和 (4) 載波音節的類型 (真字、 假字和非語音)。本論文還在同一种試驗範式下，研究了範疇感知在三种不同時間階段的表現。這三种時間階段是(1) 前知覺階段，主要通過失匹配負波來探測；(2)知覺階段，主要通過失 P300 來探測；(3)外縣反映階段，主要通過命中率來探測。

研究結果表明所有測試因素都影響了範疇感知的程度。在論文討論部分，不同的範疇感知表現被認爲是因爲測試因素中兩种處理過程的不同加權的體現。這兩种過程是普遍聲音處理過程和語言特有的處理過程。範疇感知在三個不同時間階段，表現形式也不一樣。同理，這些不同也是因爲上述兩种過程的不同加權的體現。總之，本論文提出一種包括兩种過程的新的語音感知模型來解釋聲調的範疇感知問題。這種新的模型與以前的模型一樣，都包含了普遍聲音處理和語言特有的處理過程，但是這個模型首次提出這兩种過程在不同的狀態和不同的時間階段有不同的加權表現。

最後，本論文首次發現一種新的語言現象：即使兩种不同的聲調系統都包含同一對聲調對照物，因爲被試的母語有不同的聲調系統，因此他們對這一對相同的聲調對照有不同的感知模式。這種語言現象据我們所知從未被報道過。