PERCEPTUAL REORGANISATION FOR TONE: LINGUISTIC TONE AND NON-LINGUISTIC PITCH PERCEPTION BY ENGLISH LANGUAGE AND CHINESE LANGUAGE INFANTS

Karen Janette Mattock

BPsyc (Hons)

A thesis submitted for the degree of Doctor of Philosophy
University of Western Sydney
June 2004
PLEASE NOTE

The greatest amount of care has been taken while scanning this thesis,

and the best possible result has been obtained.
Multimedia item accompanies print copy
I hereby declare that this submission is my own work and, to the best of my knowledge, it contains no material previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at the University of Western Sydney, or any other educational institution, except where due acknowledgment is made in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project’s design and conception is acknowledged.
Dedicated to the memory of my cousin

Lauren Maree Ford

14th June 1985 - 24th April 2004
Acknowledgements

‘Appreciation is a wonderful thing: It makes what is excellent in others belong to us as well’ (Voltaire, 1694-1778).

This work would not have been possible without the wisdom, generosity and kindness of the following people, to whom I am eternally appreciative.

Thank you to my dedicated supervisor Prof. Denis Burnham for his inspiration, astuteness, patience, motivation, good humour, and for always believing in this project and my capability. Thank you Denis for your advice and encouragement when it seemed that the whole project would fall apart! I am extremely fortunate to have worked with you at MARCS Auditory Laboratories, and am so grateful for the many experiences I have had including the 6 weeks setting up this project at the University of Hong Kong during 2001 and 2002, for the visits to infant laboratories in Canada during 2003, Virginia during 2004, and for support in attending the International Conference on Infant Studies 2004 in Chicago.

I thank my co-supervisor Dr. Christine Kitamura, BabyLab leader, mentor, and friend. Thank you for trusting me with the responsibility that comes with being a part of an infant lab, and for sharing your expertise on infant testing, methods, and all things speech, thesis, and the rest!

Thank you to MARCS research assistant Joan Liang, Cantonese/Mandarin/English tri-lingual and translator/interpreter extraordinaire! I honestly could not have done this without you. Sincere thanks for all of your hard work liaising with the Chinese community, for giving me insight into the languages of Cantonese and Mandarin, and for your valuable assistance on testing days.

Thank you to MARCS Lab Manager, Colin Schoknecht, for enjoyable intellectual discussions, support with lab set-up, and technical assistance. My sincerest appreciation to past and present colleagues at MARCS: Dr. Kate Stevens for her wonderful research ethic; Dr. Michael Tyler for his countless contribution to and
comments on this project; Gail Charlton for organising my overseas visits; Dr. Caroline Jones, Rua Haszard Morris, and Nicole Lees for assistance with development of the nonspeech stimuli; Helen Tam for assistance with lab set-up in Hong Kong; Barbara Schwanhäusser for discussions about tone and advice on phonetics; Dr. Kimiko Tsukada for assistance with pitch plots of the Thai, Cantonese, and Mandarin tones; the MARCS Research Assistant’s for their patience and tolerance of noise on testing days, and my sincere gratitude to Mel Gallagher for her time and assistance with formatting this thesis.

I acknowledge and appreciate the great work of Angela McIntyre, Senior Media Officer at UWS for her efforts in liaising with the media for the recruitment campaign, and thank MARCS Research Assistant’s Kate Harle and Anita Doyle for their communication with the BabyLab sponsors. To my fellow postgrads at MARCS, I thank you for your friendship and support. It has been an honour to share the PhD experience with you.

I thank my honours degree supervisor Dr. Steven Rooodernys from the University of Wollongong for introducing me to the wonders of research, and for encouraging me to apply for a PhD position with Denis at MARCS.

I have learnt so much from working with my colleagues and friends from the Division of Speech and Hearing Sciences at the University of Hong Kong: Dr. Valter Ciocca, research assistant Betty Lee, Dr. Alexander Francis (now at Purdue), department heads Dr. Paul Fletcher and Dr. Tara Whitehill, and technicians Raymond Wu and Donald Chan. Thank you for your assistance with lab set-up, recruiting and testing infants, and for your willingness to share precious equipment and space for the head-turn lab. I hope there is an opportunity to work together again in the future. Thank you to Dr. Conrad Perry for comments on the nonspeech discrimination experiments.

Many thanks to Professor Alan Slater and Dr. Rachel Hayes from the University of Exeter for permission to use their Windows head-turn program and for their continued support and advice in getting the procedure up and running in the Sydney and Hong Kong labs. I also appreciate the time and effort of John Fowler from the University of
New South Wales for building the hand-buttons, foot-pedals, and junction boxes that are needed for the head-turn procedure. Thank you also to Dr. Sudaporn Luksaneeyanawin (Chulalongkorn University, Thailand) for permission to use her lexical tone stimuli, and to Dr. Hideki Kawahara (Wakayama University) for permission to use his STRAIGHT program for research purposes.

Thank you to Professor’s Robin Panneton Cooper, Linda Polka, Glenn Schellenberg, Sandra Trehub, and Janet Werker, for allowing me to present some of this work at their laboratories, and for the wonderful experience of seeing their research teams in action! Thank you especially to Robin Panneton Cooper for enjoyable intellectual discussions about topics in this thesis.

My appreciation is extended to the Early Childhood Health Centres and hospitals for their continued support of this project, and I sincerely thank all the parents and infants who volunteered their time for this study.

I am grateful to my proofreading team for their time: Sean Coward, Doğu Erdener, Bettina Keresztesi, Christine Kitamura, Nicole Lees, Sandra Mattock, Stuart Mattock, Brett Molesworth, Margaret Parker, Barbara Schwanhäusser, Iris-Corinna Schwarz, and Michael Tyler. I am also very fortunate to have the continued support and understanding of my dear friends, Kate Jarvis, Caroline Micallef and Kelly Wren. Thank you!

To my mum Sandra, dad Stuart, and sister Alison, thank you for always encouraging me to reach for the stars, and for actually believing I can. Thank you to Papa for opening his home to me and providing me with a quiet place to study, and to Nana who supported me from above.

Finally, all my love and deepest appreciation goes to Sean, who knows exactly what this takes. Thank you for making me smile.
Abstract

Young infants can discriminate a great variety of speech sounds both native and nonnative in their language environment. At approximately 4 months a process of perceptual reorganisation begins, indexed mainly by attenuation of infants’ perceptual performance for nonnative speech sounds. Accordingly, the focus of the perceptual reorganisation research to date has been on infants’ discrimination of nonnative segments, in particular, consonants and vowels. In tone languages (e.g., Cantonese, Mandarin, and Thai) phonemic distinctions are signalled not only by consonants and vowels, but also by lexical tone – consisting of variations in fundamental frequency (pitch) and related features. Although such languages are spoken by over half the world’s population, the development of lexical tone perception has been relatively neglected. This thesis addresses whether perceptual reorganisation occurs for tone in infancy.

Eight experiments (four cross-sectional and four longitudinal) are reported here. Infants of two ages, 6 and 9 months, and from two language backgrounds, English and Chinese (Cantonese and Mandarin), were tested for their discrimination of easy and difficult Thai lexical tone contrasts, and nonspeech analogues of these contrasts with identical fundamental frequency characteristics.

In line with the perceptual reorganisation literature, English-learning infants, for whom lexical tone is nonnative, were expected to show better discrimination of lexical tone at 6 months than at 9 months, but similar performance at the two ages for nonspeech discrimination. Chinese-learning infants, for whom tone contrasts are native, were expected to show no decline in lexical or nonspeech tone discrimination over age.

Overall, the results of the experiments support the hypotheses and the existence of perceptual reorganisation for tone in infancy, similar to that for consonants and vowels. Nine-month-old English-learning infants’ discriminated lexical tone contrasts more poorly than the 6-month-olds, however a similar decrement was not found for nonspeech tone discrimination. Chinese-learning infants’ discrimination performance level for both lexical and nonspeech tone was maintained across age. Implications of the results for speech perception development theories, ‘tone space’, tone acquisition, and early word learning are discussed, and future studies relating to these issues suggested.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... xix
LIST OF TABLES ............................................................................................................... xxi

CHAPTER 1  THE SPEECH SIGNAL AND HOW IT IS PERCEIVED ........ 1
  1.1 Phones and Phonemes ................................................................................. 2
  1.2 How Speech Sounds are Produced ........................................................... 2
    1.2.1 Laryngeal Variation ........................................................................ 3
      1.2.1.1 The Larynx and Vocal Folds .................................................. 3
      1.2.1.2 Voiceless Consonants .............................................................. 4
      1.2.1.3 Voiced Speech Sounds ............................................................... 4
      1.2.1.4 Voice Quality ........................................................................... 5
      1.2.1.5 Fundamental Frequency: Tone and Prosody .......................... 5
      1.2.1.6 Periodic Sounds ...................................................................... 5
    1.2.2 Supralaryngeal Variation .................................................................... 6
      1.2.2.1 Aperiodic Sounds ................................................................... 6
      1.2.2.2 Shape of Oral Cavity ................................................................. 7
      1.2.2.3 Manner and Place of Articulation ............................................. 8
  1.3 Types of Speech Sounds ............................................................................. 8
    1.3.1 Segments ......................................................................................... 8
      1.3.1.1 Consonants ............................................................................ 9
      1.3.1.2 Vowels .................................................................................. 10
    1.3.2 Tones .............................................................................................. 11
      1.3.2.1 Static and Dynamic Tones ....................................................... 12
      1.3.2.2 Features of Tone .................................................................. 12
    1.3.3 Suprasegmental Features .................................................................. 13
      1.3.3.1 Rhythm ................................................................................ 13
      1.3.3.2 Intonation ............................................................................. 14
      1.3.3.3 Stress .................................................................................... 14
    1.3.4 Summary: Are Tones Segmental or Suprasegmental? .................. 15
  1.4 The Nature of Speech Perception .............................................................. 15
    1.4.1 The Problem of Speech Perception ................................................ 16
      1.4.1.1 Invariance and Contextual Effects ......................................... 17
1.4.1.2 Talker Variability ......................................................... 17
1.4.1.3 Segmentation ................................................................ 18

1.5 CATEGORICAL AND CONTINUOUS SPEECH PERCEPTION ............. 18

1.5.1 Consonants ................................................................... 20
1.5.1.1 Voicing ...................................................................... 20
1.5.1.2 Place of Articulation ................................................ 20
1.5.1.3 Manner of Articulation ............................................. 21

1.5.2 Vowels ......................................................................... 21

1.5.3 Nonspeech ..................................................................... 22

1.5.4 Categorical Perception by Non-human Animals ....................... 25

1.6 SUMMARY AND DISCUSSION ............................................. 26

CHAPTER 2 DEVELOPMENTAL CHANGES IN SPEECH PERCEPTION DURING THE FIRST YEAR OF LIFE ................................................. 28

2.1 TECHNIQUES FOR MEASURING INFANT SPEECH PERCEPTION .......... 30

2.1.1 High Amplitude Sucking Technique .................................. 30
2.1.2 Visual Habituation Paradigm ......................................... 31
2.1.3 Head-turn Preference Procedure .................................... 33
2.1.4 Conditioned Head-turn Procedure .................................. 34
2.1.5 Summary ...................................................................... 36

2.2 METHODS FOR INVESTIGATING THE DEVELOPMENT OF SPEECH PERCEPTION ......................................................... 36

2.2.1 The Ontogenetic Method ................................................ 36
2.2.2 The Differential Experience Method ............................... 37

2.3 THE PROBLEM OF SPEECH PERCEPTION AND INFANTS’ SOLUTIONS .......... 38

2.3.1 Invariance and Contextual Effects ................................. 38
2.3.2 Talker Variability ......................................................... 40
2.3.3 Summary ..................................................................... 41

2.4 CATEGORICAL PERCEPTION BY INFANTS ........................................... 41

2.4.1 Consonants ................................................................. 42
2.4.1.1 Voicing ................................................................. 42
2.4.1.2 Place of Articulation ................................................ 43
2.4.1.3 Manner of Articulation ............................................. 44
2.4.1.4 Is Infants’ Perception of Consonants Really Categorical? ........ 44

2.4.2 Vowels ......................................................................... 45
### 2.4.3 Nonspeech ................................................................. 46
### 2.4.4 Summary ................................................................. 47

### 2.5 Early Linguistic Sensitivities ........................................ 47
#### 2.5.1 Preference for Mother’s Voice and Speech Patterns .......... 47
#### 2.5.2 Native Language Preferences ................................... 48
#### 2.5.3 The Role of Prosody in Early Speech Preferences .......... 49
#### 2.5.4 Infant-directed Speech and the Role of Emotion in Early ... 52

### 2.6 Recognising the Suprasegmental and Phonotactic Properties of the Ambient Language: The Origins of Word Recognition .......... 54
#### 2.6.1 Preference for Native Language Stress Patterns ................ 54
#### 2.6.2 Phrase Boundaries and Word Boundaries ..................... 56
#### 2.6.3 Word Segmentation ................................................. 57
#### 2.6.4 Statistical Learning ................................................. 59

### 2.7 Infants’ Perception of Segments .................................... 61
#### 2.7.1 Development of Consonant Perception ......................... 62
    #### 2.7.1.1 Development of Consonant Perception and Age-related Advances in Cognitive Abilities .......................... 64
#### 2.7.2 Development of Vowel Perception ................................ 65
#### 2.7.3 From Language-general to Language-specific Speech Perception .... 68

### 2.8 Theories and Models of Speech Perception .......................... 69
#### 2.8.1 The Three Periods of Infant Speech Perception Research .... 69
    #### 2.8.1.1 Universal, Attunement, and Perceptual Learning Theories .......... 70
    #### 2.8.1.2 Perceptual Assimilation Model .................................. 73
    #### 2.8.1.3 Native Language Magnet Theory .................................. 75
    #### 2.8.1.4 Word Recognition and Phonetic Structure Acquisition Model .... 77
    #### 2.8.1.5 An Epigenetic Model of Speech Perception .................... 78
    #### 2.8.1.6 PAUSE: mir ..................................................... 80
    #### 2.8.1.7 Summary ..................................................... 80

### CHAPTER 3 Lexical Tone .................................................. 81
#### 3.1 What is Lexical Tone? ................................................ 82
    #### 3.1.1 What Lexical Tone is Not ..................................... 83
#### 3.2 Tonogenesis ............................................................. 84
3.3 Tone Space in Thai, Cantonese and Mandarin ................................................. 86
  3.3.1 The Nature of Thai Tones ............................................................................. 86
  3.3.2 The Nature of Chinese Tone: Cantonese and Mandarin .............................. 88
    3.3.2.1 Description of Cantonese Tones ............................................................ 89
    3.3.2.2 Description of Mandarin Tones ............................................................. 92
  3.3.3 Tone Sandhi ................................................................................................. 94

3.4 The Perception of Tone ..................................................................................... 95
  3.4.1 Perceptual Dimensions of Tone ................................................................. 96
  3.4.2 Interaction Effects and Tone Perception .................................................... 99
    3.4.2.1 Lexical Tone versus Intonation ............................................................. 100
    3.4.2.2 Lexical Tone and Coarticulation .......................................................... 102
    3.4.2.3 Summary ......................................................................................... 103

3.5 The Acquisition of Tone ................................................................................... 103
  3.5.1 Production .................................................................................................... 105
  3.5.2 Perception .................................................................................................... 106
    3.5.2.1 The Use of Tone Cues Over Age ......................................................... 106

3.6 Issues in Tone Perception .................................................................................. 107
  3.6.1 Mode of Processing in Tone Perception ....................................................... 108
    3.6.1.1 Evidence from Inter-Stimulus Interval Studies .................................. 108
    3.6.1.2 Evidence from Hemispheric Differences and Aphasics ..................... 110
  3.6.2 Experiential Influences on Tone Perception ............................................. 111
    3.6.2.1 The Role of Linguistic Experience in Tone Perception ....................... 111
    3.6.2.2 The Role of Context in Tone Perception ............................................ 113
    3.6.2.3 The Role of Musical Training in Tone Perception ............................... 114
    3.6.2.4 Summary ......................................................................................... 114
  3.6.3 Is Tone Perceived Continuously or Categorically? ..................................... 115
  3.6.4 Are Tones Phones? ..................................................................................... 118
    3.6.4.1 Linguistic Considerations in the Tone as Segment Argument ............. 118
    3.6.4.2 The Relative Salience of Tones versus Consonants and Vowels .......... 119
    3.6.4.3 Tonological Awareness versus Phonological Awareness .................. 120

3.7 Infants’ Perception of Tone ................................................................................. 121
  3.7.1 Threshold Sensitivity to Pure Tones ............................................................ 121
  3.7.2 Pitch in Music ............................................................................................. 122
CHAPTER 4  EXPERIMENTAL RATIONALE, FRAMEWORK, HYPOTHESES, AND GENERAL METHOD .......................... 131

4.1  THE NEED FOR RESEARCH ON LEXICAL TONE DISCRIMINATION BY ................................................................. 132

   Infants ......................................................................................................................................................... 132

4.1.1  Framework for Experiments: Perceptual Reorganisation ................................................................. 133

4.1.2  Language Background, Age of Participants, and Stimulus ......................................................... 135

        Selection ............................................................................................................................................ 135

4.1.3  Task and Method .......................................................................................................................... 137

4.1.4  Aim and Purpose of Studies ....................................................................................................... 138

4.1.5  General Hypotheses ................................................................................................................... 138

4.2  METHOD ........................................................................................................................................ 140

4.2.1  Participants .................................................................................................................................. 140

4.2.2  Design .......................................................................................................................................... 140

        4.2.2.1  Experimental design ............................................................................................................ 141

4.2.3  Stimuli .......................................................................................................................................... 141

4.2.4  Procedure .................................................................................................................................... 142

        4.2.4.1  Traditional Conditioned Head-turn Procedure .................................................................. 142

        4.2.4.2  Windows Conditioned Head-turn Procedure .................................................................... 144

4.2.5  Laboratories and Equipment ........................................................................................................ 145

        4.2.5.1  Testing Chambers .................................................................................................................. 145

4.3  EXPERIMENTERS AND EXPERIMENTAL TRAINING ................................................................. 150

4.3.1  Training the Experimenter and Testing in Sydney .................................................................. 151

4.3.2  Training the Experimenter and Testing in Hong Kong .............................................................. 151

CHAPTER 5  ENGLISH-LEARNING INFANTS’ DISCRIMINATION OF TONE................................................................. 153

5.1  EXPERIMENT I: ENGLISH-LEARNING INFANTS’ DISCRIMINATION OF LEXICAL TONE CONTRASTS ................................................................. 154

5.2  METHOD ........................................................................................................................................ 154

      5.2.1  Design ........................................................................................................................................ 154
5.7 Discriminating Lexical Tone versus Nonspeech Tone: Comparing
Experiment 1 and Experiment 2 ..................................................... 176

5.7.1 Results .................................................................................. 176
5.7.2 Summary .............................................................................. 178

CHAPTER 6 Chinese-Learning Infants' Discrimination of Tone
........................................................................................................ 179

6.1 Experiment 3: Chinese-Learning Infants' Discrimination of Lexical Tone Contrasts ..................................................... 180

6.2 Method .................................................................................... 180
6.2.1 Design .................................................................................... 180
6.2.2 Participants ........................................................................... 180
6.2.2.1 Participant Numbers ............................................................. 180
6.2.2.2 Assessment of Infants' Exposure to a Chinese Language .... 181
6.2.3 Stimuli ..................................................................................... 182
6.2.4 Materials and Apparatus ....................................................... 183
6.2.5 Procedure ............................................................................. 183

6.3 Results ..................................................................................... 184
6.3.1 Proportion Reaching Criterion .............................................. 184
6.3.2 Discrimination Performance .................................................. 186
6.3.2.1 Entire Sample ........................................................................ 186
6.3.2.2 Subsample of Infants Reaching Criterion ......................... 187
6.3.3 Experimental 1 Stage Data .................................................... 188
6.3.3.1 Discrimination Performance ............................................. 188
6.3.4 Performance Across Stages ................................................... 188
6.3.5 Summary of Results, Experiment 3 ....................................... 189

6.4 Experiment 4: Chinese-Learning Infants' Discrimination of Nonspeech Tone Contrasts ..................................................... 189

6.5 Method .................................................................................... 190
6.5.1 Design ..................................................................................... 190
6.5.2 Participants ........................................................................... 190
6.5.3 Stimuli ..................................................................................... 190
6.5.4 Materials and Apparatus ....................................................... 191
6.5.5 Procedure ............................................................................. 191
6.6 RESULTS .............................................................................................................. 191
6.6.1 Proportion Reaching Criterion ................................................................. 191
6.6.2 Discrimination Performance .................................................................. 193
   6.6.2.1 Entire sample ...................................................................................... 193
   6.6.2.2 Subsample of Infants Reaching Criterion ........................................ 194
6.6.3 Experimental 1 Stage Data ................................................................... 195
   6.6.3.1 Discrimination Performance ............................................................. 195
6.6.4 Performance Across Stages ..................................................................... 195
6.6.5 Summary of Results, Experiment 4 ......................................................... 195

6.7 DISCRIMINATING LEXICAL TONE VERSUS NONSPEECH TONE: COMPARING
   EXPERIMENT 3 AND EXPERIMENT 4 ............................................................... 196
6.7.1 Results ........................................................................................................ 196
6.7.2 Summary ................................................................................................... 197
6.8 ORDER EFFECTS .......................................................................................... 197
6.9 QUESTIONNAIRE DATA ............................................................................... 198
   6.9.1 Coding Responses ............................................................................... 198
   6.9.2 Correlations ............................................................................................ 199
   6.9.3 Factor Analysis ....................................................................................... 200
   6.9.4 Regression ............................................................................................... 201
   6.9.5 Summary ................................................................................................ 202

CHAPTER 7 INTERLUDE: THE STORY SO FAR....................................................... 203

CHAPTER 8 LONGITUDINAL DATA: ENGLISH- AND CHINESE-
   LEARNING INFANTS’ DISCRIMINATION OF TONE .................................. 206
8.1 LONGITUDINAL STUDY OF ENGLISH-LEARNING INFANTS’ DISCRIMINATION OF
   LEXICAL TONE: EXPERIMENT 5 ................................................................. 207
8.2 METHOD .......................................................................................................... 207
   8.2.1 Participants ............................................................................................. 207
   8.2.2 Stimuli ...................................................................................................... 208
   8.2.3 Materials and Apparatus ....................................................................... 208
   8.2.4 Procedure ................................................................................................. 208
8.3 RESULTS .......................................................................................................... 208
   8.3.1 Proportion Reaching Criterion ............................................................... 209
   8.3.2 Discrimination Performance ................................................................. 209
8.3.2.1 Entire Sample ................................................................. 209
8.3.2.2 Subsample of Infants Reaching Criterion ......................... 210
8.3.3 Experimental 1 Stage Data ................................................. 211
     8.3.3.1 Discrimination Performance ..................................... 211
8.3.4 Performance Across Stages .............................................. 211
8.3.5 Summary of Results ...................................................... 211

8.4 Longitudinal Study of English-Learning Infants' Discrimination of Nonspeech Tone: Experiment 6 ........................................... 212

8.5 Method ............................................................................. 212
     8.5.1 Participants ............................................................... 212
     8.5.2 Stimuli .................................................................... 213
     8.5.3 Materials and Apparatus ............................................ 213
     8.5.4 Procedure .................................................................. 213

8.6 Results .............................................................................. 214
     8.6.1 Proportion Reaching Criterion ..................................... 214
     8.6.2 Discrimination Performance ....................................... 214
     8.6.2.1 Entire Sample ....................................................... 214
     8.6.2.2 Subsample of Infants Reaching Criterion .................. 215
     8.6.3 Experimental 1 Stage Data .......................................... 216
     8.6.3.1 Discrimination Performance .................................... 216
     8.6.4 Performance Across Stages ......................................... 216
     8.6.5 Summary of Results .................................................. 216

8.7 Longitudinal Study of Chinese-Learning Infants’ Discrimination of Lexical Tone: Experiment 7 ............................................. 217

8.8 Method ............................................................................. 217
     8.8.1 Design ..................................................................... 217
     8.8.2 Participants .............................................................. 218
     8.8.3 Stimuli ..................................................................... 218
     8.8.4 Materials and Apparatus ............................................ 218
     8.8.5 Procedure ................................................................. 218

8.9 Results .............................................................................. 219
     8.9.1 Proportion Reaching Criterion ..................................... 219
     8.9.2 Discrimination Performance ....................................... 219
8.9.2.1   Entire Sample .......................................................... 219
8.9.3    Summary of Results .................................................. 221

8.10   Longitudinal Study of Chinese-learning Infants’ Discrimination of Nonspeech Tone: Experiment 8 .............................................. 221

8.11   Method ................................................................. 221
8.11.1   Design .................................................................... 221
8.11.2   Participants ........................................................... 222
8.11.3   Stimuli ................................................................... 222
8.11.4   Materials and Apparatus .............................................. 222
8.11.5   Procedure .............................................................. 222

8.12   Results ................................................................. 223
8.12.1   Proportion Reaching Criterion .................................... 223
8.12.2   Discrimination Performance ...................................... 223
8.12.2.1   Entire sample ......................................................... 223
8.12.2.2   Subsample of Infants Reaching Criterion ................. 224
8.12.3   Summary of Results .................................................. 224

CHAPTER 9   THE COMPLETE PICTURE ........................................... 226

9.1   An All-Inclusive Analysis of Variance ................................... 227
9.1.1   A Problem with the Analysis and a Proposed Solution ........... 227
9.1.2   Results of the 2 x 2 x 2 x 2 factorial ANOVA ..................... 228
9.1.3   Simple Effects .......................................................... 230
9.1.4   Summary ................................................................. 231

9.2   Overview of Main Findings ................................................... 232
9.2.1   Discrimination of Lexical Tone by English-learning Infants .......... 234
9.2.1.1   Experiment 1: Cross-sectional Study .............................. 234
9.2.1.2   Experiment 5: Longitudinal Study ................................ 234
9.2.2   Discrimination of Nonspeech tone by English-learning Infants ...... 235
9.2.2.1   Experiment 2: Cross-sectional Study .............................. 235
9.2.2.2   Experiment 6: Longitudinal Study ................................ 235
9.2.3   Discrimination of Lexical tone by Chinese-learning Infants ........ 235
9.2.3.1   Experiment 3: Cross-sectional Study .............................. 235
9.2.3.2   Experiment 7: Longitudinal Study ................................ 236
9.2.4   Discrimination of Nonspeech Tone by Chinese-learning Infants .... 236
LIST OF FIGURES

Figure 1.1. The human speech production system................................................. 6

Figure 1.2. The vowel quadrilateral. The vowel quadrilateral represents the extreme points of articulation........................................... 7

Figure 1.3. Chart of consonant sounds (pulmonic and non-pulmonic)............... 9

Figure 2.1. Plot of F1/F2 frequencies showing asymmetries in infants’ vowel discrimination................................................................. 68

Figure 2.2. Models of Perceptual Learning Development: Universal theory, Attunement theory and Perceptual Learning theory ...................... 72

Figure 3.1. Time-normalised fundamental frequency contours of the five Thai tones (spoken by a male speaker of Thai)................................. 89

Figure 3.2. Time-normalised fundamental frequency contours of the six Cantonese tones (spoken by a female speaker of Cantonese)............... 93

Figure 3.3. Time-normalised fundamental frequency contours of the four Mandarin tones (spoken by a female speaker of Mandarin)........... 95

Figure 4.1. Schematic diagram of the test and control rooms, MARCS BabyLab, University of Western Sydney........................................ 147

Figure 4.2. Schematic diagram of the test and control rooms, Infant Perception Laboratory, University of Hong Kong............................. 149

Figure 4.3 Circuit diagram of the head-turn apparatus...................................... 151

Figure 5.1. Mean percent correct discrimination of the easy and difficult lexical tone contrasts by 6- and 9-month-old English-learning infants .......................................................................................... 166

Figure 5.2. Mean percent correct discrimination of the easy and difficult nonspeech tone contrasts by 6- and 9-month-old English-learning infants .......................................................................................... 175

Figure 6.1. Mean percent correct discrimination of the easy and difficult lexical tone contrasts by 6- and 9-month-old Chinese-learning infants .......................................................................................... 188

Figure 6.2 Mean percent correct discrimination of the easy and difficult nonspeech tone contrasts by 6- and 9-month-old Chinese-learning infants .......................................................................................... 195
Figure 8.1. Mean percent correct discrimination of the easy lexical tone contrast by English-learning infants tested at 6 and 9 months of age .... 211

Figure 8.2. Mean percent correct discrimination of the easy nonspeech tone contrast by English-learning infants tested at 6 and 9 months of age .... 216

Figure 8.3. Mean percent correct discrimination of the easy and difficult lexical tone contrasts by Chinese-learning infants tested at 6 and 9 months .............................................. 221

Figure 8.4. Mean percent correct discrimination of the easy and difficult nonspeech tone contrasts by Chinese-learning infants tested at 6 and 9 months .............................................. 225

Figure 9.1. The age by task interaction collapsed across difficulty and language background .......................................................... 230

Figure 9.2. The interaction between difficulty and language for lexical tone discrimination .......................................................... 232

Figure 9.3. Summary of infants’ mean percent correct discrimination in Experiments 1 to 4 .......................................................... 234

Figure 10.1. Asymmetries in English- and Chinese-learning infants’ discrimination of the easy lexical tone contrast ..................... 254

Figure 10.2. Asymmetries in English- and Chinese-learning infants’ discrimination of the difficult lexical tone contrast ..................... 254
LIST OF TABLES

Table 2.1 The separate and combined features of the Ontogenetic (amount) and the Differential Experience (type) methods for investigating linguistic development ........................................ 39

Table 3.1. The five lexical tones of Standard Thai and an example of a five-way lexical contrast on the syllable /kʰa/ .................................................. 88

Table 3.2. The six lexical tones of Cantonese and an example of a six-way lexical contrast on the syllable /fan/ ...................................................... 92

Table 3.3. The four lexical tones of Mandarin and an example of a four-way lexical contrast on the syllable /ma/ ........................................... 94

Table 5.1. Proportion of English-learning infants reaching criterion for discrimination of the easy and difficult lexical tone contrasts ............. 162

Table 5.2. Proportion of 6- and 9-month-old English-learning infants reaching criterion for discrimination of the lexical tone contrasts .......... 163

Table 5.3. Proportion of 6-month-old English-learning infants reaching criterion for discrimination of the easy and difficult lexical tone contrasts ............................................................. 164

Table 5.4. Proportion of 9-month-old English-learning infants reaching criterion for discrimination of the easy and difficult lexical tone contrasts ............................................................. 164

Table 5.5. Proportion of 6- and 9-month-old English-learning infants reaching criterion for discrimination of the easy lexical tone contrast ... 165

Table 5.6. Proportion of 6- and 9-month-old English-learning infants reaching criterion for discrimination of the nonspeech tone contrasts ... 173

Table 5.7. Proportion of English-learning infants reaching criterion for discrimination of the easy and difficult nonspeech tone contrasts .......... 173

Table 5.8. Proportion of 6- and 9-month-old English-learning infants reaching criterion for discrimination of the easy and difficult nonspeech tone contrasts ............................................................. 174

Table 5.9. Proportion of 6-month-old English-learning infants reaching criterion for discrimination on the lexical tone and nonspeech tone tasks ................................................................. 178

Table 5.10. Proportion of 9-month-old English-learning infants reaching criterion for discrimination on the lexical tone and nonspeech tone tasks ................................................................. 178
Table 6.1. Proportion of 6- and 9-month-old Chinese-learning infants reaching criterion for discrimination of the lexical tone contrasts .................... 186

Table 6.2. Proportion of Chinese-learning infants reaching criterion for discrimination of the easy and difficult lexical tone contrasts .................... 186

Table 6.3. Proportion of 6- and 9-month-old Chinese-learning infants reaching criterion for discrimination of the lexical tone contrasts ............ 187

Table 6.4. Proportion of 6- and 9-month-old Chinese-learning infants reaching criterion for discrimination of the nonspeech tone contrasts ............ 193

Table 6.5. Proportion of Chinese-learning infants reaching criterion for discrimination of the easy and difficult nonspeech tone contrasts ............ 193

Table 6.6. Proportion of 6- and 9-month-old Chinese-learning infants reaching criterion for discrimination of the easy and difficult nonspeech tone contrasts ................................................. 194

Table 6.7. Proportion of 6-month-old Chinese-learning infants reaching criterion for discrimination on the lexical tone and nonspeech tone tasks ............ 197

Table 6.8. Proportion of 9-month-old Chinese-learning infants reaching criterion for discrimination on the lexical tone and nonspeech tone tasks ............ 198

Table 6.9. Pearson product-moment correlations between parents’ language use and infants’ mean percent correct lexical tone discrimination .......... 201

Table 6.10. Principle component analysis loadings for Chinese language use questionnaire variables ................................................................. 202

Table 8.1. Proportion of English-learning infants reaching criterion for discrimination of the lexical tone contrasts at 6 and 9 months of age ....... 210

Table 8.2. Proportion of English-learning infants reaching criterion for discrimination of the nonspeech tone contrasts at 6 and 9 months of age ......................................................................................... 215

Table 8.3. Proportion of Chinese-learning infants reaching criterion for discrimination of the lexical tone contrasts at 6 and 9 months of age ....... 220

Table 8.4. Proportion of Chinese-learning infants reaching criterion for discrimination of the nonspeech tone contrasts at 6 and 9 months of age ......................................................................................... 224

Table 9.1. English-learning infants’ mean percent correct discrimination of the easy and difficult lexical and nonspeech tone contrasts .......... 230

Table 9.2. Chinese-learning infants’ mean percent correct discrimination of the easy and difficult lexical and nonspeech tone contrasts .......... 231
CHAPTER 1

The Speech Signal and How It is Perceived

‘The intellect pierces the form, overleaps the wall, detects intrinsic likeness between remote things and reduces all things into a few principles’

(R. W. Emerson, 1841).
Speech consists of small units or segments that make up words - consonants, vowels and tones - and in turn, the words and suprasegmental prosodic information that make up utterances. This thesis is concerned with the word level of speech perception, and in particular, tones. Nevertheless an understanding of all three word level units, and the suprasegmental features of speech, is necessary. This chapter consists of two parts (a) an articulatory description of speech sounds and (b) speech perception.

1.1 Phones and Phonemes

A phone is a speech sound that can be distinguished on the basis of phonetic or articulatory features. A phoneme is a family of sounds used to distinguish meaning in a particular language. For example /p/ and /b/ are different phonemes in English and thus ‘pat’ and ‘bat’ differ in meaning. Phonemes may be realised as one of several different phones (allophones). An allophone is a member of a phoneme category. For example, the [pʰ] in ‘past’ the [p] in ‘spoon’ and the [p] in ‘chop’ are all allophones of the phoneme /p/ even though the sounds are articulated differently and are acoustically as well as phonetically distinct. In written text, including this thesis, slashes are used to indicate phonemes /p/, and brackets are used to indicate phones [p].

From the array of possible phones each language differs in which particular set of phones it uses, and how these phones are grouped into phonemes for that language. Consequently, it is important for the learner of a language to discover which phones serve as functional sound units for forming words in that language, and moreover which differences among the sounds are relevant for distinguishing meaning and which differences are irrelevant.

1.2 How Speech Sounds are Produced

This section is devoted to a discussion of the three word level units- consonants, vowels and tones, and is in two parts. The first part reviews how speech sounds are produced and focuses on laryngeal and supralaryngeal variation. This forms the basis for the second part, which reviews the perception of consonants, vowels and tones.
1.2.1 Laryngeal Variation

Laryngeal variation refers to vocal fold adjustments for speech that are responsible for voiced and voiceless speech sounds, fundamental frequency, voice quality, and periodic and aperiodic sounds. During speech production there is separation of the vocal folds for voiceless sounds and adduction and vibration of the vocal folds for voiced sounds, and the pressure wave of sound produced by the vocal folds is periodic (i.e., repeating). Fundamental frequency, perceived by the human listener as the speaker’s pitch, relates directly to vocal fold vibration rate that is mainly controlled by the cricothyroid muscles in the larynx. Voice quality or timbre is partly due to differences in the fundamental frequency of voices, but also relates to differences in vocal fold vibration such as breathy voice, which is achieved when vocal folds are not adducted sufficiently for full voicing. In this section the role of laryngeal variation in voiced versus voiceless speech sounds, fundamental frequency, voice quality, and periodic pressure wave patterns are discussed in turn.

1.2.1.1 The Larynx and Vocal Folds

The larynx is used for producing speech\(^1\) and forms the upper end of the trachea. An understanding of laryngeal function is enlightening with respect to the articulation of various phonetic and tonetic variations. The larynx is a tube composed of cartilages connected by ligaments and membranes, and covered by mucous membranes. The space within the tube contains two sets of vocal folds, the false vocal cords (the ventricular folds) and the true vocal cords used for voicing. The vocal folds themselves are made up of the vocal ligaments, the elastic protuberances of tendon, muscles, and mucous membrane in the larynx.

When relaxed, the vocal folds are thick and they open and close in an undulating manner. Speech is powered by expired air from the lungs that the upper airways convert into audible sound waves for speech. Speakers use two methods of

\(^1\) The larynx also controls flow of air into and out of the lungs, it prevents substances from entering the lungs (e.g., food, water), assists in swallowing, and enables air pressure build up for such functions as coughing and vomiting.
transforming the expired air into speech sounds. The first, voicing, uses the air pressure to set the vocal folds into vibration by bringing the vocal folds together, producing a repeated pattern of vibration. The second method involves allowing the air to pass through the larynx into the vocal tract where noises and bursts are produced.

1.2.1.2 Voiceless Consonants
When producing voiceless consonants such as /s/ and /t/, the vocal folds are wide open in order to obtain sufficient air from the lungs to create noise in the oral cavity. This is the simplest adjustment that the vocal folds need to make for speech. In continuous speech where there is a speedy transition between voiceless and voiced sounds, the vocal folds must open and close rapidly to interrupt voicing.

1.2.1.3 Voiced Speech Sounds
To produce voiced consonants such as /z/ and /d/, and vowels such as /u/ and /i/ as in ‘shoe’ and ‘sheep’, the vocal folds must vibrate, that is, the normally separated folds must be adducted. Stronger adduction of the folds is needed for vowel production. Thus, unlike voiceless sounds where the sounds are created in the oral cavity, the vocal cords vibrate, causing a basic source sound, and the particular vowels are formed by the acoustic resonance produced by the shape of the oral cavity. For speech sounds such as /z/, that require voicing but also a sound source above the glottis, the vocal cords are less closely adducted than in vowel production.

Voicing is a phonetic feature that specifies particular phones and classes of phones, phonemes. For example, a bilabial stop can be voiced /b/, voiceless /p/ or even prevoiced /-p/. Lisker and Abramson (1964) showed that voice-onset-time (VOT), the period between the complete obstruction of the airstream and the onset of vocal fold vibration, while not the sole cue, is a powerful cue for perception of voicing. Among the multiple acoustic cues that comprise the phonetic cue of voicing are for example F1 cutback - the degree to which the first formant of the vowel following the consonant is shortened, or ‘cut back’ (see Stevens & Klatt, 1974).

---

2 The voiced /b/ and voiceless /p/ are present in English. The prevoiced /-p/ is present in various other languages, e.g., Thai.
1.2.1.4 Voice Quality
Differences in voice quality depend upon the mode of vocal fold vibration. For example, a breathy voice is achieved by not completely adducting the vocal cords for full voicing, so air is continuously released. A hoarse voice is caused by irregularities in the folds due to irritation, swelling (such as in laryngitis), nodules due to vocal abuse or overuse, or when there is decreased resistance to the air pressure because of a reduced amount of stiffness in the vocal cords.

1.2.1.5 Fundamental Frequency: Tone and Prosody
The elasticity and tension of the vocal folds can be varied; the folds can be lengthened, shortened, opened wide, closed together, elevated or depressed. In continuous speech these adjustments occur very rapidly. The number of times the vocal cords open and close per second is the frequency of vocal cord vibration. The frequency of vibration determines the fundamental frequency (F0), the lowest frequency of the sounds produced. The human listener perceives F0 as the speakers’ pitch and this varies during speech as the speaker alters the frequency of vocal fold vibration. For example, the different intonation patterns we hear in sentences such as ‘Do you know about Bob?’ (rising intonation) and ‘I know about Bob’ (falling intonation) are due to changes in F0. Additionally, differences between /mā/ and /mâ/ in a tone language such as Thai are substantially based on F0 (discussed later in section 1.3.2). The average F0 for male, and female and children’s voices, is 125Hz and 200Hz + respectively. Men typically have larger vocal folds than women (17-24mm and 13-17mm respectively) and the larger the mass of the vocal folds, the lower the frequency.

1.2.1.6 Periodic Sounds
The term periodic pertains to vocal fold vibrations occurring at equal intervals of time. Vowels are periodic sounds. It is the air flow of voicing in vowel production that creates the periodic sound wave. The rapid opening and closing of the vocal folds converts the steady flow of air from the lungs (subglottal system) into a series of
audible puffs\textsuperscript{3}. Sustained vowels are also periodic, which means that the vocal folds open and close in a repeating pattern of movement. This action produces a barrage of airbursts that create an audible sound wave at the glottis. However as specified in the following section speech is a complex waveform and the sound produced at the glottis is modified by the particular shape of the vocal tract, giving rise to a range of frequencies in the waveform over and above the fundamental frequency.

1.2.2 Supralaryngeal Variation

In addition to laryngeal variation there is also supralaryngeal variation or vocal tract variation. The vocal tract, shown in Figure 1.1, consists of all of the air passages above the larynx from the glottis to the lips (i.e., the pharyngeal cavity, the oral cavity, and the nasal cavity). In this section, the role of the oral cavity in speech sound production is considered, followed by a discussion of manner and place of articulation for speech sounds.

![Diagram of the human speech production system](image)

*Figure 1.1. The human speech production system.*

1.2.2.1 Aperiodic Sounds

Aperiodic sounds are produced by allowing air to pass through the open glottis and into the vocal tract where constrictions of the tract produce localised air turbulence.

\textsuperscript{3} There are unvoiced vowels in Japanese, where the vowels between two voiceless consonants tend to be unvoiced.
Consonants are aperiodic sounds. The consonant sound /z/ is a combination of periodic and aperiodic sound. It is a prolonged /s/ (aperiodic) combined with voicing periodic. In contrast, ‘clicks’, ‘bursts’, and ‘pops’ do not require air pressure from the subglottal system and are produced by localised changes in the vocal tract. Clicks, bursts and pops are classified as neither periodic nor aperiodic sounds.

1.2.2.2 Shape of Oral Cavity

The vowel quadrilateral, shown in Figure 1.2, represents tongue height on the ordinate and backing on the abscissa. The vowel sounds /i/, /a/, and /u/ form the extremes of the English vowel triangle. The front of the tongue is high for /i/, and the back of the tongue is high for /u/. For the front vowels, [i] as in ‘eat’, [ɪ] as in ‘it’, [ɛ] as in ‘ate’, [ɛ] as in ‘Ed’, and [ə] as in ‘at’, there is barely any lip movement and tongue height determines the resonance of the front vowels. The back vowels (from high to low), ‘shoe’ [u], ‘soot’ [ʊ], ‘sew’ [ou], ‘saw’ [ɔ] are articulated with changes in lip rounding and cavity shape, and ‘sock’ [ɔ] has maximum mouth opening.

![Vowels Quadrilateral Diagram](image)

*Figure 1.2. The vowel quadrilateral. The vowel quadrilateral represents the extreme points for articulation. Vowels on the same horizontal line have equal tongue height (F1). Vowels on the vertical are articulated with equal fronting or backing (F2). Figure reproduced from the Handbook of the International Phonetic Association (1999).*
1.2.2.3 Manner and Place of Articulation

For vowels, the source of the sound and the resonance of the vocal tract are most important. However for consonants, the vocal tract is the source of the sound and must become sufficiently tensed so that the air within the vocal tract will vibrate with greater amplitude at certain frequencies (resonances), while also producing decreased energy in certain frequency ranges (anti-resonances).

Place of articulation refers to the place of articulatory contact or constriction and speech sounds can be classified on this basis. For example /b/ has a bilabial (two lips together) and /t/ has an alveolar (tongue in contact with the alveolar ridge) place of articulation. Place of articulation for vowels and consonants are typically charted separately. The international phonetic chart for consonants is shown in Figure 1.3 in which place of articulation is represented on the horizontal axis.

Manner of articulation is a means of classifying consonant sounds based on the strategy used for producing the sound. For example, nasals, stops, fricatives, and affricates are distinguished from one another by manner of articulation. Manner of articulation is represented on the vertical axis in Figure 1.3, and different manners of articulation are discussed further in section 1.3.1.1.
THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

<table>
<thead>
<tr>
<th>Plosive</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retruded</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>t</td>
<td>d</td>
<td>c</td>
<td>j</td>
<td>k</td>
<td>g</td>
<td>q</td>
<td>g</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>nj</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Trill</td>
<td>B</td>
<td>b</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>Tap or Flap</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>Fricative</td>
<td>φ</td>
<td>β</td>
<td>f</td>
<td>v</td>
<td>θ</td>
<td>ð</td>
<td>s</td>
<td>z</td>
<td>s</td>
<td>z</td>
<td>s</td>
</tr>
<tr>
<td>Lateral fricative</td>
<td>l</td>
<td>b</td>
<td>l</td>
<td>b</td>
<td>l</td>
<td>b</td>
<td>l</td>
<td>b</td>
<td>l</td>
<td>b</td>
<td>l</td>
</tr>
<tr>
<td>Approximant</td>
<td>u</td>
<td>[</td>
<td>ɾ</td>
<td>j</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

<table>
<thead>
<tr>
<th>Clicks</th>
<th>Voiced implosives</th>
<th>Ejectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ć</td>
<td>Bilabial</td>
<td>as in:</td>
</tr>
<tr>
<td>l</td>
<td>Dental</td>
<td>Dental/alveolar</td>
</tr>
<tr>
<td>!</td>
<td>(Post)alveolar</td>
<td>Palatal</td>
</tr>
<tr>
<td>†</td>
<td>Palatoalveolar</td>
<td>Velar</td>
</tr>
<tr>
<td>‖</td>
<td>Alveolar lateral</td>
<td>Uvular</td>
</tr>
</tbody>
</table>

Figure 1.3. Chart of consonant sounds (pulmonic and non-pulmonic). Consonants are charted by manner of articulation on the vertical axis, and place of articulation on the horizontal axis (The International Phonetic Alphabet, 1993).

1.3 Types of Speech Sounds

In section 1.2 a distinction was made between laryngeal and supralaryngeal variation in speech sound production. This section specifies the types of laryngeal and supralaryngeal variation in consonants and vowels, tones, and the suprasegmental features of speech rhythm, intonation and stress.
1.3.1 Segments

Sections 1.3.1.1 and 1.3.1.2 introduce the different classes of consonants and vowels respectively.

1.3.1.1 Consonants

There are 11 classes of consonants, as shown in Figure 1.3 (section 1.2.2.3) but only the consonant classes of English - nasals, stops, fricatives and affricates - are introduced in this section.

Nasal resonance is needed for the production of /m/, /n/ and /ŋ/ in English. In the production of such nasals the entrance to the nasal cavities are open producing a larger resonator in which the air flow may operate, and the oral cavity is occluded. For /m/ the lips are closed (bilabial). For /n/ the tongue touches the upper alveolar ridge of the hard palate and the sides of the tongue touch the molars. When producing /ŋ/ the top of the tongue touches the back of the palate (the velum) allowing less of the oral cavity to resonate (the velar place of articulation).

Stops, fricatives and affricates are three aperiodic speech sounds resonated in the vocal tract and can be produced with various places of articulation, for example, bilabial, alveolar and velar. There are six stop consonants in English, /p, b, t, d, k, g/. In the production of stops the oral cavity rapidly creates a build-up of air pressure at a particular place of articulation, which is suddenly released by relaxing the occlusion. Stops can be either voiced (e.g., in English /b, d, g/) or voiceless (e.g., /p, t, k/), and can be aspirated as in the [pʰ] in ‘pin’ or unaspirated, as in the [p] phone ‘spin’.

Fricatives are created in the vocal tract by the air stream passing through constrictions in the tract at a particular place of articulation. English has five places of articulation used to produce fricatives: labiodental (/f/ as in ‘fan’ and /v/ as in ‘van’), interdental (/θ/ as in ‘thought’ and /ð/ as in ‘thy’), alveolar (/s/ as in ‘Sue’ and /z/ as in ‘zoo’),
palatal (/ʃ/ as in ‘shoe’ and /z/ as in ‘measure’), and glottal /h/. The voiced fricatives are /v, z, ʒ, h/ and the voiceless fricatives are /f, ɾ, θ, s, ʃ, h/.

An affricate is a stop with a fricative release. There are two affricates in English, [tʃ] and [dʒ], as in ‘chair’ and ‘jump’ respectively. Alveolar closure occurs for [t] or [d], but constriction noise [ʃ,ʒ] is produced when this closure is released.

1.3.1.2 Vowels

For the /i/ as in the word ‘sheep’, the oral cavity is constricted, the speaker fronts and elevates the tongue toward the alveolar ridge, and there is no lip protrusion, hence /i/ is classified as a high, front unrounded vowel. The configuration of the vocal tract for the low back vowel [o] is the reverse, the tongue is lowered in the back of the oral cavity. For the high back rounded vowel /u/, the vocal tract is elongated, by means of lip protrusion and the tongue is elevated towards the palate.

Vowels with greater tongue adjustments and longer duration are tense vowels, which appear in open syllables as in the words ‘see’, ‘say’, and ‘sow’. Vowels that occur in syllables ending with consonants are called lax vowels, for example ‘sit’, and ‘sat’, and are produced with less extreme movements than tense vowels.

Steady vowels are called monophthongs. Diphthongs are vowels of changing resonance, for example, /ai/, /ou/ /ei/ /au/ /ɔi/. Diphthongs ending with /i/, entail tongue movements forward and up from the [e], [a] and [ɔ] positions. Diphthongs ending with vocal tract cavities appropriate for /u/ involve moving the tongue back and up concurrently with lip protrusion.

---

4 The aspirate /h/ is a fricative. It is usually voiceless as in the word ‘hat’ but can be voiced, as in the word ‘ahead’.
The sounds /w/ ('way'), /j/ ('yes'), /r/ ('red') and /l/ ('lamp') are called semivowels, because they have the properties of both vowels and consonants. As for vowels the vocal tract is open for semivowels and they are highly resonant, particularly when in syllable final positions\(^5\). On the other hand, like consonants, semivowels function in the language to release vowels and diphthongs, for example 'win' /win/ and 'spring' /sprɪŋ/. Semivowels are divided into glides, /j/ and /w/, requiring movements of the tongue and lips to change the shape of the vocal tract from the starting position; and liquids /r/ and /l/ are produced by raising the tongue toward the alveolar ridge while voicing.

**1.3.2 Tones**

Lexical tone and lexical stress\(^6\) are the two methods by which some languages use fundamental frequency, amplitude, and duration to affect meaning at the lexical level (Cutler & Chen, 1997). In tone languages such as Thai and Cantonese, lexical forms are distinguished not only by vowel and consonant variations but also by differences in level and/or movement of F0 contours on a particular vowel phoneme (Nootenboom, 1997). In pitch accent languages, such as Swedish and Japanese, the relative F0, amplitude and duration in polysyllabic words is used to contrast meaning. In stress or 'intonation' languages such as English, pairs of unrelated words differing only in stress pattern are rare. In English semantically related words differing in their grammatical function are more common. For example, SUBject and subJECT are distinguished by vowel differences in the first syllable [ʌ] and [ə] and stress patterns; strong-weak and weak-strong respectively. Hence in English, stress can determine word meaning but it rarely does.

It is lexical tone that is of greatest interest here. Tone in the linguistic sense refers to that group of acoustic and articulatory features loosely centred on pitch (and not on vowel or consonant features) that are used to distinguish meaning. In the phonetic description of tone, beginning, middle and endpoint F0 values must be specified, and

\(^5\) This is the case for American English but not Australian English.

\(^6\) Lexical stress refers to the stress patterns of words and also to pitch accent.
where relevant, the point of inflection (Anderson, 1978) (the point at which the tone contour changes direction, for example in the falling tone of Thai; see also section 3.3.1).

### 1.3.2.1 Static and Dynamic Tones

Tones may be divided into the level tones, which are heard as having no pitch movement, that is, static pitch, and contour or dynamic tones, which feature rapid F0 movement and an audible rise or fall in pitch (Abramson, 1978b). Languages with a high proportion of static tones are called register-tone languages, and languages in which a high proportion of the tones are dynamic tones are called contour-tone languages. Yoruba is an example of a register-tone language, with three static tones and no dynamic tones (Gandour & Harshman, 1978). Thai and Cantonese are contour-tone languages with five (Abramson, 1962) and six (So, 1996) contrastive tones respectively. Thai has three level and two contour tones, and Cantonese has three level and three contour tones.

### 1.3.2.2 Features of Tone

The principal phonetic feature of lexical tone is pitch, whose primary acoustic correlate is the fundamental frequency (F0) of the speech signal. Nevertheless tone is not uniquely specified by F0 and pitch; other acoustic features that figure in tone distinctions are physical duration, F2 values, voice quality and amplitude (Abramson, 1978b; Henderson, 1981; Tseng, Massaro, & Cohen, 1985), perceived respectively as vowel length, vowel height, vowel quality and loudness (Abramson, 1978b; Tseng et al., 1985). For example, vowel length and breathy or creaky voice quality are well integrated into the tone systems of Mandarin and Thai respectively. A good analogy is that tone is to pitch as voicing is to voice-onset-time (see section 1.2.1.3 for a description of VOT) - it is the primary physical correlate, though not necessarily the only one. Anderson (1978) states that distinctions in voice quality, such as breathy voiced versus ordinary vowels, somewhat resemble the function of a tone distinction and both the

---

7 While Yoruba does not have lexically contrastive dynamic tones, contour tones can occur phonetically as a function of rules that combine two of the level tones, but this will not be discussed here (Gandour & Harshman, 1978).
tone and voice quality oppositions are closely related to laryngeal control. In fact Haudricourt (1972) and Hombert, Ohala, and Ewan (1979) go as far as saying that specific tone registers are actually associated with specific voice qualities, for example, that voiced consonants are related to low tone and voiceless consonants are related to high tone.

1.3.3 Suprasegmental Features

The suprasegmental features of a language are features larger than individual segments that are overlaid upon a word, phrase or sentence. Suprasegmental features are a direct bridge to meaning; they reveal the attitude and feelings of the speaker that cannot be conveyed by the segmental information alone and also convey syntactic information. Stress, rhythm and intonation are three suprasegmental features that will be discussed in this section.

1.3.3.1 Rhythm

Rhythm, as it applies to language, is the alternation of strong and weak beats in speech. Nooteboom (1997) points out that rhythm exists either in overall utterances, for example ‘The MAN in the STREET’ (where capitalised words are accented) or within words, for example ‘CONtent’ vs. ‘conTENT’. An important determinant of rhythm in English-language word units is the ‘stress foot’, also called the ‘within word foot’ (Shattuck Hufnagel & Turk, 1996). A stress foot is a rhythmic unit in speech that contains a lexically stressed syllable, followed by (if available) a lexically unstressed syllable (Echols, Crowhurst, & Childers, 1997). In English, stress feet are generally trochaic, that is they have a strong-weak stress pattern of a lexically stressed syllable followed by zero or one unstressed second syllable (see Cutler, Dahan, & van Donselaar, 1997; Nooteboom, 1997; Shattuck Hufnagel & Turk, 1996). The opposition between strong and weak features is an important feature of English, and word segmentation is facilitated on the basis of this utterance rhythm (Cutler et al., 1997).

The English lexical stress system is governed by two principles. One is the principle of syllable weight where heavy syllables (words containing long vowels) are those that are stressed, (e.g., table) or closed by a consonant that is not a permissible initial consonant
cluster in English. For example, in the word ‘monkey’ the first syllable is heavy because the vowel is followed by the consonant cluster /nk/ (Turk, Juszczyk, & Gerken, 1995). The second principle governing English stress is the principle of rhythmic alternation. In this case, stress in a word is assigned to alternating syllables from left to right, starting with an unstressed/light syllable. The word ‘beckon’ is an example of this second principle. The first syllable of ‘beckon’ is light (contains a lax vowel) but rhythmic alternation assigns stress to this syllable. In the word ‘telephone’ the first syllable is light, the final syllable is heavy, and the principle of rhythmic alternation assigns stress to these syllables.

1.3.3.2 Intonation

In section 1.3.2 tone was described as F0 characteristics carried on a syllable. In contrast, intonation is F0 modulation that is perceived as the pitch or intonation contour of a phrase or sentence. Both tone and intonation convey differences in meaning, for tone this is meaning between words, and for intonation, syntactic meaning. In English, a sentence produced with rising intonation usually signals a question, for example ‘He is happy?’ whereas the declarative sentence ‘He is happy’ has sentence declination.

Intonation patterns also signal attitudes and feelings. Anger and high levels of enthusiasm are marked by large shifts in intonation, while calm, subdued states including boredom, are characterised by a narrow range of intonation variation. Intonation patterns can be imposed on a sentence, phrase, or word (Abercrombie, 1968; Kramer, 1963). English declarative sentences and open-ended questions are mostly characterised by a rise-fall intonation curve (e.g., ‘Where have you visited in Sydney?’). Another common intonation curve is the end of utterance pitch rise (e.g., ‘Is it finished?’) characterised by an increased fundamental frequency and wider pitch range than declarative utterances. Specific speech registers have more exaggerated intonation patterns than others. For example, mothers and others communicate with infants using infant-directed speech (IDS), which has heightened fundamental frequency, relatively larger pitch excursions, and wider pitch range compared with adult speech (Burnham, Kitamura, & Vollmer-Conna, 2002). The exaggerated intonation
patterns communicate affect, and serve to attract, engage and maintain infant attention (Kitamura & Burnham, 1998).

1.3.3.3 Stress

There are two uses of stress in English - rhythmic stress and lexical stress (i.e., stress that signals meaning differences). Stress in rhythm was reviewed above in section 1.3.3.1, so the focus of this section is how linguistic stress is used contrastively in English.

In some two-syllable words, moving stress to the second syllable changes lexical class, that is, nouns into verbs. For example, ‘PERmit’ with the first syllable stressed is a noun meaning a document of authorisation, but ‘perMIT’ with the second syllable stressed is a verb meaning ‘to allow’. Other examples are ‘EXtract’ versus ‘exTRACT’, and ‘DIgest’ versus ‘diGEST’. Second syllable stress tends to occur for nouns as in ‘to estiMATE’ but is less likely to occur in verb form, for example ‘an estimate’. Stress is characterised by increased articulatory effort, pitch, duration, intensity, and a change in the pattern of formants. The fundamental frequency of stressed syllables usually increases, while for de-stressed syllables the vowels are neutralised, that is, they become more schwa-like. Vowels are longer in duration in stressed syllables and tend to be of higher intensity due to the greater subglottal air pressure.

Stress can be used to emphasise words in a sentence. For example, ‘take him to the shop NOW’ versus ‘take him to the shop now’ changes the meaning of sentences. Without stress some written sentences may be ambiguous, for example ‘They are eating chickens’.

1.3.4 Summary: Are Tones Segmental or Suprasegmental?

Tone can be considered to be both segmental and suprasegmental. Like intonation and stress, tones can be considered suprasegmental because they overlay vowels. However tones also serve to contrast lexical meaning and therefore function like segments (Clumeck, 1980). In support of the segmental function of tone, tone is carried mainly
on the vowel of a syllable although some involvement of surrounding consonants may also be implicated.

1.4 The Nature of Speech Perception

In general, the field of speech perception originated from formal linguistic theory in which speech was originally viewed as an idealised sequence of discrete symbolic units that must be abstracted from the stimulus environment and constructed by the perceiver. Chomsky (1965) remarking on the idealised forms of language in linguistic theory said,

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. We thus make a fundamental distinction between competence (the speaker-hearer’s knowledge of his language) and performance (the actual use of language in concrete situations).

In line with this abstractionist view of speech, most early researchers in speech perception assumed that the stimulus environment was impoverished and that the perceiver made sense of their environment via constructive processing (Pisoni & Lively, 1995). Thus much of the early work on speech cues was concerned with identifying acoustic invariants in the speech signal that correspond to phonemes. However, it was soon realised that a one-to-one correspondence between speech cues and phonemes was unlikely to be found, as the acoustic properties associated with a particular phone were dependent on the nature of the surrounding phonetic segments (Cooper, Dellatre, Liberman, Borst, & Gerstman, 1952). This has been termed the invariance problem, and is one of a number of problems in speech perception that are discussed in the next section.
1.4.1 The Problem of Speech Perception

Observing adults engaged in conversation would suggest that decoding speech is effortless. Indeed, the subjective experience of speaking and comprehending one’s native language is seemingly effortless. Nevertheless, on deeper analysis this apparently simple task is much more complex, a truth to which any second language (L2) learner will testify. Vihman (1996) enumerated three characteristics of speech that complicate the decoding task of the listener; the invariance problem, talker variability, and the segmentation problem. Another developmentally oriented problem of speech perception was introduced earlier, namely the fact that first language (L1) learners (infants) and L2 learners (infants, children or adults) must determine the subset of all possible speech sounds that are used in the specific language and uncover their segmental and prosodic structures. This developmental issue is discussed in chapter 2 and the three problems of speech perception, invariance, talker variability and segmentation are discussed below.

1.4.1.1 Invariance and Contextual Effects

The invariance problem is characterised by the same identifiable phonetic segment being manifested differently even within an utterance, depending on the position of the phonetic segment in an utterance and the phonetic context. This is due to coarticulation, the overlapping of articulatory movements, lips, tongue, face, and velum, associated with consecutive phonetic segments. For example in ‘tap’ the [t], [æ], and [p] are different from those in ‘pat’.

The invariance problem is underlined by attempts to develop Automatic Speech Recognition (ASR) systems. Early in the development of this technology it became apparent that speech was very difficult for machines to decode, even though children and adults conduct this task effortlessly. In fact as section 2.3 outlines, infants also have a remarkable ability to cope with variations in the speech signal, which far surpass that of machines.

---

8 Indeed, ASR technology may even learn from infants’ speech perception (see Burnham, 1998).
1.4.1.2 Talker Variability

The talker variability problem relates to how two tokens of a single utterance type are recognised as the same despite great variations in acoustic output between speakers. For example the size and shape of the vocal tract, F0, loudness of delivery, voice quality, and speed and accuracy of articulation all influence the signal in different ways resulting in marked acoustic differences between the same utterance produced by different speakers. Human listeners must be able to compensate for these differences to be able to correctly perceive the linguistic message (for a review see Jusczyk, 1995) in a similar manner to how they cope with other perceptual constancies.

1.4.1.3 Segmentation

Fluent perceivers of a language tend to hear speech as a sequence of discrete words, despite considerable acoustic overlap between successive words, syllables and phones in the speech stream. When listeners perceive a foreign language they often have difficulty identifying where one word ends and the next one begins because utterances are often produced without interruptions, and any interruptions that occur may be due to the production of stop consonants rather than the marking of a word boundary. This lack of an overt reliable marking of boundaries is the segmentation problem.

1.5 Categorical and Continuous Speech Perception

In section 1.4, the means by which listeners deal with the problems of speech perception - lack of invariance, talker variability and contextual information - were discussed. These means involve the identification of relevant similarities between speech tokens and classification on the basis of these dimensions. To classify in this way, sounds must of course be discriminable.

A related point of interest is when physical differences are not discriminable - when there are discontinuities in the perception of speech sounds which vary along a physical continuum. This is called categorical perception and refers to a case in which two sounds differing by \( d \) units along a continuum may be discriminated in a particular
region of a continuum but another two sounds differing in $d$ units in another region of the continuum may not be discriminated.

In this section categorical speech perception is considered in relation to what it reveals about the nature of speech, the human (and non-human) auditory system, and the way in which consonants, vowels and tones are perceived. The origins of categorical and continuous perception in infancy will be discussed in chapter 2 (section 2.4).

Speech processing by adults is characterised by the ability to extract perceptual invariance from the extremely variable speech signal. One solution to the problems of speech perception is categorical speech perception, in which sounds varying along a certain continuum are perceived to belong to discrete categories, with discrimination between tokens at the boundaries of these categories (i.e., between-category differences), occurring more easily than discrimination of differences between tokens within the same phoneme category (i.e., within-category differences) (Burnham & Jones, 2002). Categorical perception is not limited to the domain of speech. For example, both adults and infants categorically perceive the visible wavelengths of light as distinctly different hues with sharp category boundaries although some within-hue discrimination is possible (Bornstein, Kessen, & Weiskopf, 1976). Both identification (labelling) functions and discrimination functions are required to determine whether categorical perception of a continuum is present (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

Tests of categorical speech perception involve the use of a series of synthetic speech sounds that are created by altering the acoustic dimension of interest in small equal steps. Listeners are then tested on their discrimination of neighbouring speech sounds drawn from either all along the continuum, either within categories or from opposite sides of the category boundaries; and/or for their identification of individual speech sounds as members of one or the other category. A large amount of research with stop consonants, and some with place of articulation features, shows that consonants tend to be perceived categorically (Blumstein & Stevens 1980). Vowels on the other hand, are perceived continuously (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973). The difference between the perception of consonants and vowels may be linked to the
different roles they play in language. Werker and Polka (1993b) state, "consonants carry information specifying lexical contrast … whereas vowels carry information about speaker identity, stress, intonation, emotional state etc…"

The evidence for and against the categorical perception of consonants and vowels is discussed below, with the debate surrounding how tone is perceived, reserved until chapter 3 (section 3.6.3.).

1.5.1 Consonants

1.5.1.1 Voicing

There is evidence that the native linguistic environment exerts a profound influence on the ability to categorise sounds differing in VOT into the phonological categories of the native language. Lisker and Abramson (1964) examined the VOT and aspiration differences of word-initial stop consonants produced by adult native speakers of 11 different languages. Using spectrographic analysis they identified three categories of voicing that form the basis for phonological categories, and showed that boundary locations for stops are slightly different across the languages. These categories of voicing are prevoiced (−100ms⁹) where voicing onset precedes stop release, short-lag or voiced (+10ms) in which voicing onset and release burst are almost simultaneous, and a long-lag or voiceless mode (+70ms) where voicing onset occurs following the release burst. Adults perceived the three categories of voicing based on psychoacoustic thresholds for the perception of simultaneity (Jusczyk, Pisoni, Walley, & Murray, 1980), but were able to discriminate the speech contrasts only if the contrasts were from relevant phonological categories in the native language.

1.5.1.2 Place of Articulation

Several acoustic properties, such as a graded change in second formant transition burst spectra, and direction of rapid spectra change following consonantal release, have been implicated in the perception of place of articulation (e.g., Cooper et al., 1952; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Stevens & Blumstein.

⁹ These are mean values for the voicing categories.
1975). No one of these appears to be the primary perceptual feature for place of articulation because all can signal a place of articulation distinction in different phonetic contexts.

In a landmark study, Liberman et al. (1967) used the Pattern Playback speech synthesiser to produce 14 two-formant sound (vowel-like) patterns differing only in the direction and extent of F2 transition. Subjects heard /ba/ for those stimuli having the most rapidly rising transitions, then as the transitions became less sharp or began to fall, subjects reported a sudden change to hearing /da/, their percepts finally changing to /ga/ at the falling transition end of the continuum. Similarly Mattingly, Liberman, Syrdal, and Halwes (1971) created a synthetic place of articulation continuum that ranged from /bae/ to /dae/ to /gae/ using two formant synthetic speech sounds that varied only with respect to F2 starting point. These second-formant transitions were sufficient to cue categorical perception of /bae/, /dae/ and /gae/ and the discrimination functions were accordingly marked by sharp peaks in the regions of the phonetic boundaries.

1.5.1.3 Manner of Articulation

Adult listeners continue to perceive consonant contrasts categorically when they differ in manner of articulation. For example syllable-initial /b/ and /w/ are typically distinguished in a categorical manner by listeners on the basis of duration of the initial formant transitions into the following vowel: /b/ is typically produced with short transitions and /w/ with long transitions (Liberman, Delattre, Gerstman, & Cooper, 1956).

Miller and Liberman (1979) have established that the transition duration cue that maps onto percepts of /b/ and /w/ is dependant on syllable duration (length) which is known to vary with changes in speaking rate. In their /ba/ to /wa/ continuum, Miller and Liberman adjusted the steady-state vowel portions of the stimuli so that within-category syllables varied in transition duration from 16ms to 64ms and between-category syllables ranged from 80 to 296ms duration. The findings highlight that rate is a cue to manner of articulation because as the syllable became longer, the location of the category boundary between /b/ and /w/ shifted towards a longer transition duration.
The main aim of this section was to show that consonants are perceived categorically. In the next section, data on vowel perception is considered.

1.5.2 Vowels

The perception of steady-state vowels is quite different from the perception of stop consonants. Studies of vowel perception indicate that, in contrast to stops, vowels are perceived continuously, that is, that listeners can discriminate many more vowel stimuli even if they belong to the same category (see Fry et al., 1962). There is also evidence that phonemic vowel duration is perceived continuously. Eimas (1963) demonstrated that vowels, vowel duration, and tones (see section 3.6.3) are all perceived continuously, in essentially the same manner as continuous variations in nonspeech stimuli.

The story for vowels is different when they are not isolated. The more usual situation for vowels is that they occur between consonants with rapid articulation. Stevens (1966) proposes that the perception of vowels in ecologically-valid contexts (rapid changes in formant position) is more categorical than that of steady-state vowels presented in isolation, and generally the more natural the sound the more categorical the perception (Hessen & Schouten, 1999).

To quote Liberman et al. (1967), categorical perception means that “Listeners ... discriminate only slightly better than they can identify absolutely ” (p.443) and occurs only when a physical continuum is perceived discontinuously. This appears true for consonants as reviewed in section 1.5.1, but the research reviewed here on vowel perception suggests that vowels are perceived continuously.

1.5.3 Nonspeech

Liberman, Harris, Kinney, & Lane (1961) found perception of nonspeech to be more continuous than speech perception. When they had adults discriminate between stop consonants in nonspeech instead of speech contexts, within-category differences were more discriminable, perceived continuously as a function of physical or acoustic
differences. These and similar findings by Lisker and Abramson (1964), were taken to indicate that categorical perception is a special feature of speech processing, that speech categories are influenced by the ambient phonological environment and are therefore linguistic rather than acoustic (Liberman et al., 1961). For a recent discussion of this issue refer to Burnham, Tyler, and Horlyck (2002). A good test-bed for this theory is infants (this issue is discussed in chapter 2, section 2.4.1.4), and non-human animals (section 1.5.4).

This so-called ‘speech is special’ argument has been challenged by findings of categorical perception for certain nonspeech continua (Cutting & Rosner, 1994; Cutting, Rosner, & Foard, 1976; Pisoni, 1977). Cutting and Rosner (1974) varied the rise-time of sawtooth wave stimuli to create a continuum of stimuli that sounded as if they had been generated by a musical stringed instrument. Categorical identification and discrimination functions for these nonspeech stimuli were found, with fast rise-time sawtooth wave stimuli (less than 35ms) perceived as ‘plucked’ and those with gradual rise-times (greater than 35ms) perceived as ‘bowed’ sounds. In a speech context, chop-shop and chad-shad rise-time cues were also perceived categorically. Subsequent research further quashed the long held assumption that only speech sounds were perceived categorically. For example, Pisoni (1977) tested English-speaking adults for their perception of a series of two-component nonspeech tones (each tone was composed of a 500Hz tone and a 1500Hz tone) differing in tone-onset-time (TOT). The subjects displayed categorical perception by grouping the tones into three different categories, with boundaries for these at approximately –25ms TOT (low 500Hz tone onset preceding high 1500Hz tone onset by 25ms) and +25ms TOT (low 500Hz tone onset lagging 25ms behind high 1500Hz tone onset). Given the similarity of these values to VOT values in voicing distinction, the similarity of the low frequency 500Hz tone to voicing, and the high 1500Hz tone to release, Pisoni raised the possibility that general auditory mechanisms are responsible for detecting temporal-order differences underlying the perception of voicing contrasts. However, Bailey, Summerfield, and Dorman (1977) found categorical boundaries for a set of nonspeech frequency-and-amplitude-modulated sine waves, modelled on stop consonant-vowel (CV) speech syllables, but the boundaries did not correspond with the phonetic boundaries obtained
for the speech stimuli. Bailey and colleagues argued that this was evidence for a specialized speech processing mechanism.

Clearly, this latter research highlights parallels between the way adults perceive speech and nonspeech sounds. Why might the earlier studies of categorical perception (e.g., Liberman et al., 1961) have shown differences in the way speech and nonspeech are perceived? One possibility is that the nonspeech stimuli used as the controls lacked important speech-like information. Indeed Liberman et al. (1961) admitted that this was a possible explanation for the differences between the perception of speech and nonspeech continua. Furthermore when Mattingly et al. (1971) investigated categorical perception in speech and nonspeech (chirps and bleats) modes they found categorical perception for the former but not for the latter stimuli. Interestingly, both their chirp and bleat stimuli lacked first formant transition information and since then Jusczyk, Smith, and Murray (1981) have suggested that the presence or absence of first formant transitions essentially affects adults perceptual classification of chirps as categorical or continuous. It seems then, that a decision about whether nonspeech control stimuli are appropriate depends on the results of one’s investigation. In some cases, nonspeech sounds may act as ‘superstimuli’ that trigger speech-processing mechanisms when they possess speech-like characteristics, for example changes in speech rate or temporal-order differences. However if the pattern of results for speech and nonspeech diverge then one can claim that critical information was omitted from the nonspeech sounds.

Recent research by Vouloumanos, Kiehl, Werker, and Liddle (2001) sheds some light on the ‘good’ and ‘bad’ of nonspeech stimuli. They presented speech, high and low frequency simple tones (simple nonspeech), and matched sine wave analogues of speech (complex nonspeech) to adults, and measured cortical activation using event-related fMRI to investigate the neural substrates mediating the initial processing of speech - when the listeners detect a change in the auditory stream from speech to nonspeech. Vouloumanos and colleagues argued that the complex nonspeech stimuli, unlike the simple tones, were a more suitable control for the speech stimuli because they were more ‘speech like’ by being matched to speech along temporal and spectral

---

10 Vouloumanos et al. (2001) used a simple oddball detection paradigm and equated the stimuli for familiarity and controlled task-related attentional demands, thus differences in cortical activation were unlikely to be confounded by these factors (refer to their manuscript for a discussion of these issues)
dimensions. The speech stimuli elicited more activation in brain regions that have previously been identified as classic receptive language areas including the left hemisphere, Wernicke's area and Broca's area\textsuperscript{11}, than the complex and simple nonspeech stimuli. In turn, complex nonspeech activated the bilateral language regions more so than the simple nonspeech but to a lesser extent than the differential activation in response to speech. Thus the complex nonspeech stimuli were processed in the cortex in a more speech-like manner than the simple tones.

It seems premature to draw any conclusions as to whether the underlying capacities responsible for speech perception are specialised for speech perception, or whether there is a general auditory mechanism. While behavioural evidence suggests that speech and nonspeech are processed similarly in some cases, it does not automatically imply that the same perceptual mechanism is invoked for both. It may be that the nonspeech stimuli are so close to speech in some critical respect that even though the sounds are not perceived as speech, the speech processor is engaged (Miller & Eimas, 1994).

One aim of this thesis, in light of the 'speech is special' debate, is to determine whether lexical tones are perceived in a similar fashion to pitch changes, and if not, to determine the critical difference(s) between these speech (tone) versus nonspeech (pitch) sounds. Thus the evidence for and against the 'speech is special' argument is of great interest here. To address this issue, we turn to instances in which human and non-human animals diverge in their processing of speech sounds. This issue is addressed next.

1.5.4 Categorical Perception by Non-human Animals

Humans and non-human animals respond to speech, but non-human animals do not make meaningful use of speech. It is unclear whether this capacity depends on a unique neurobiological mechanism or whether a part of this mechanism is found in other animals. One strategy for addressing whether speech is special has been to compare the response patterns of humans and evolutionary lower animals to speech stimuli. The

\textsuperscript{11} Broca's area of the brain is predominately associated with speech production rather than speech perception.
rationale here is that any parallel between the way that humans and non-human animals perceive speech would be support for the view that the mechanisms subserving human speech are general rather than highly specialised mechanisms dedicated to processing speech. The first suggestion that non-human mammals may display similar speech processing abilities to humans came from a study with chinchillas. Kuhl and Miller (1978) reported that chinchillas display categorical-like\(^{12}\) discrimination functions for several voicing contrasts, and that these functions were similar to the way humans discriminate perceptual boundaries for voicing contrasts. Subsequently, Kuhl and Padden (1982) found that macaque monkeys (non-human primates) also discriminate voicing and place of articulation contrasts in a similar manner to human listeners. Moreover, Kluender, Diehl, and Killeen (1987) trained Japanese quail to categorise speech sounds according to place of articulation contrasts used in English. These results do not support the existence of specialised speech mechanisms in humans; rather they suggest that human and nonhuman species process speech signals similarly, using a general auditory mechanism. However Miller and Eimas (1994) warn against assuming that identical mechanisms are responsible for speech processing across different species. For example, monkeys do not always discriminate speech contrasts in a way that matches human listeners. Waters and Wilson (1976) noted that although rhesus monkeys discriminated voicing contrasts categorically, unlike humans their boundaries were affected by the initial training stimuli. Similarly, Sinnott, Beecher, Moody, and Stebbins (1976) found that macaque monkeys could discriminate place of articulation differences, but they required much greater stimulus differences to detect these than human listeners. Finally, Kuhl (1991) tested whether monkeys demonstrate a ‘perceptual magnet effect’ (see chapter 2, section 2.8.1.3 for a discussion of this effect and Native Language Magnet theory) for certain vowel contrasts as humans do (i.e., less accurate discrimination of vowel pairs when they are phonetically close to a vowel prototype or ‘magnet’). The perceptual magnet effect was not evident in monkeys’ responses to the vowel stimuli. Monkeys treated the variants surrounding the prototype in exactly the same way regardless of their distance from the prototype. These results are interesting because unlike the case of categorical perception, they reveal dissociation between humans and monkeys in a test of phonetic perception.

\(^{12}\) Note that the term ‘categorical-like’ is used for describing studies that have assessed either discrimination or identification, but not both.
1.6 Summary and Discussion

In summary, the first part of this chapter introduced the speech signal and its component speech sounds, consonants, vowels and tones. The second part was devoted to human adults and non-human animals’ perception of the speech signal, with particular focus on consonants, vowels, tones, and nonspeech sounds. Chapter 2 is concerned with infants’ early sensitivity to suprasegmental variations in speech, and to their increased attention to phonemic segments over the course of the first year.
CHAPTER 2
Developmental Changes in Speech Perception During the First Year of Life

‘One of the major accomplishments of early childhood is learning to understand one or more languages’ (Polka, Colantonio & Sundara, 2001)
Over the past three decades a substantial body of research has accumulated concerning infant’s speech perception abilities. In this chapter, the focus is on how infants deal with the problems of speech perception, categorical perception, prenatal speech perception, and the development of language specific speech perception. Surrounding this content, the chapter begins with methods for testing infants, and concludes with a consideration of the models of infants’ speech perception.

Evidence suggests that low frequency acoustic information, in the form of the mother’s voice, is transmitted to the unborn infant through air and bone conduction, and during the last trimester of gestation influences the newborn infant’s propensity to process global language properties (DeCasper & Fifer, 1980). There is considerable support for the ability of human infants to discriminate between an array of phonetic contrasts, even phonetic contrasts that are not phonemically relevant in the infant’s native language (Aslin, Pisoni, Hennessy, & Perey, 1981; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; for a review see Werker & Tees, 1992). In comparison, adults have difficulty discriminating nonnative speech distinctions (Werker, Gilbert, Humphrey, & Tees, 1981). There is a small improvement in adults’ ability to discriminate nonnative contrasts when short inter-stimulus intervals (≥ 500ms) are used (Werker & Logan, 1985; Werker & Tees, 1984b) and following training, however performance levels fail to equal the ability of native speakers and infants. During the second half of the first postnatal year there is a reorganisation of speech perception abilities. As a result of linguistic experience, early discriminative abilities are aligned and sharpened and infants become increasingly attuned to processing phonetically relevant contrasts of the ambient language-learning environment. As a trade-off, the ability to perceive nonnative phonetically irrelevant contrasts is attenuated (Werker & Tees, 1992).

This chapter is concerned with the development of speech perception in infancy, early linguistic sensitivities, and the period of reorganisation/alignment towards the native language. The chapter will focus on developmental changes in the perception of nonnative prosodic patterns, consonants, and vowels and emphasises the need for the investigation of lexical tone perception in infancy. Lexical tone is a segmental and
suprasegmental element of speech that has been neglected in studies of infant speech perception development.

2.1 Techniques for Measuring Infant Speech Perception

A pervasive problem that a language researcher faces is just how to ask nonverbal infants about their perception of speech, a problem not unique to infant speech and language research, but also pertinent in any studies of cognitive and perceptual capacities in animals and infants (Jusczyk, 1997).

This section outlines the main methods for measuring speech perception in the first year postpartum, the High Amplitude Sucking Technique, the Head-turn Preference Procedure, the Visual Habituation Paradigm and the Conditioned Head-Turn Procedure, with the aim to provide a basis for discussion of the speech perception development literature.

2.1.1 High Amplitude Sucking Technique

The High Amplitude Sucking (HAS) technique was first used by Eimas et al. (1971) and is suited to testing infants aged from 1 to 4 months. This study will be used as a model to explain the procedure. In this procedure the infant is placed in a reclining chair directly facing a blank wall approximately 1 metre away. A colourful slide is projected on the wall above an audio speaker that plays the sounds. Infants suck on a non-nutritive nipple that is attached via a pressure transducer to a polygraph that provides a digital output of the infants’ high amplitude sucks.

In the pre-shift phase of the experiment, if the rate of high amplitude sucking responses reaches criterion for a trial, one speech syllable is presented as reinforcement and this is continued for every burst of sucking that reached criterion. The maximum stimulus presentation rate is between one and 2 syllables per second. Typically under these conditions high-amplitude sucking habituates as the infant learns about the repeated speech sound.
When the habituation criterion is reached (at least a 20-33% decrease in each infants' rate of high amplitude sucks) in the pre-shift phase for 2 consecutive minutes, the post-shift phase begins. At this point the auditory stimulus changes for infants in the experimental condition, but remains the same for infants in the control condition. It is expected that experimental but not control infants will show recovery of sucking rate in the post-shift phase.

Some variations to the HAS procedure have involved assessing infants’ abilities to handle stimulus variability such as talkers’ voices and multiple stimuli while still recognising shared phonetic segments (e.g., Kuhl, 1983), and to investigate speech categorisation in young infants (Bertoncini, Floccia, Nazzi, & Mehler, 1995). A more recent modification has been to insert delay periods of two minutes between the pre- and post-shift phases of the experiment (Jusczyk, 1997) to investigate infants’ ability to encode and remember speech sounds.

One major disadvantage of the HAS procedure is the high rate of attrition due to restlessness, crying, sleeping or failing to produce and maintain sucking rates above criterion (Polka, Jusczyk, & Rvachew, 1995). The procedure is also only suitable for very young infants.

2.1.2 Visual Habituation Paradigm

In the visual habituation procedure, infants’ fixation of a visual stimulus results in presentation of repeating auditory stimuli. The paradigm rests on the premise that when there is a change of auditory stimulus, infants will exhibit reliable increases in visual fixation if they are able to differentiate the change (Jusczyk, 1997).

Polka, Jusczyk and Rvachew (1995) outline the procedure. The infant views a visual display on a blank background. Sounds are presented through a loudspeaker whenever, and only if, the infant visually fixates the display. A flashing light is used to initially centre the infant’s attention towards the visual display and is extinguished once fixation is established. Hidden observers record when the infant is visually fixating on the pattern and their responses lead to presentation of audio sounds via computer. The
visual pattern and auditory stimuli are either presented simultaneously and terminated when the infant looks away (e.g., Polka & Werker, 1994), or the visual pattern is presented before the presentation of the auditory stimuli and terminated when the infant looks away from the visual pattern (Best, McRoberts, & Sithole, 1988). The duration of auditory stimulus presentation for each trial is controlled by the infant, but across successive trials with the same visual pattern and auditory stimulus, the duration of visual fixation tends to decline (Polka et al., 1995). When a habituation criterion is reached, (typically a 50% reduction in the infant’s looking time), infants in the experimental group are presented with a change in the auditory stimulus, whereas infants in the no-change group are presented with the same auditory stimulus. Infants’ discrimination of the auditory change is indexed by comparing the duration of fixations at the end of the pre-shift habituation period with those in the post-shift period. The experimental group is expected to show a larger increase in visual fixation following the auditory change, relative to the no-change control group. However, Cohen (2001) suggests that rather than having an experimental group and a control group, each infant should be tested on change versus no-change trials to determine whether looking times are greater to a new novel stimulus compared to a familiar habituated stimulus.

The visual habituation procedure has been successfully used to provide information about infants’ ability to discriminate global differences in native and nonnative speech patterns (Mehler et al., 1988), contrasts differing by a single phonetic feature (Best et al., 1988; Polka & Werker, 1994), changes in talker gender (Miller, 1983), and has been modified as a word learning procedure (Stager & Werker, 1997; Werker, Cohen, Lloyd, Casasola, & Stager, 1998; Werker, Corcoran, Fennell, & Stager, 2002). In the word learning ‘switch’ task infants are habituated to two word-object pairings, for example, ‘lif’ paired with object 1, and ‘neem’ paired with object 2, and are tested on their ability to recognise a switch in the pairing, for example ‘lif’ with object 2 (Werker et al., 1998).

An advantage of using the visual fixation method as a dependent variable is that it enables the visual habituation method to be used with infants from across the age range of 2 to 14 months (Jusczyk, 1997) and has been adapted to assess adult discrimination (Best et al., 1988). The method also allows for the presentation of speech samples
longer than a single word. Unfortunately high attrition rates have been reported by a number of experimenters in both the infant controlled (Polka & Werker, 1994) and fixed trial duration procedures (Mehler et al., 1988).

2.1.3 Head-turn Preference Procedure

Procedures like the HAS and the operant head-turn method (discussed in sections 2.1.1 and 2.1.4 respectively) typically only allow the presentation of brief stretches of speech (of a few syllables in length). However, the head-turn preference procedure (HPP) is very useful for investigating cues which are distributed over long utterances, such as prosody (Polka et al., 1995), and phonotactic patterns in native language words (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). The HPP is also well suited for studying infants’ perception of the information contained in short utterances such as lexical stress (Jusczyk, Cutler, & Redanz, 1993), and has been used to obtain repeated measures from the same subject in longitudinal studies (Jusczyk et al., 1992). The HPP has been modified slightly across experiments depending on the nature of the speech samples to be presented (Kuijpers, Coolen, Houston, & Cutler, 1998). The HPP developed by Jusczyk and Aslin (1995) provides the basis for the following description.

The infant is seated on the parent’s lap inside a three-sided pegboard test booth open on one side. A hidden observer blind to which stimuli are presented to the infant on each side, records the duration and direction of the head-turns and terminates trials when the infant looks away for more than two seconds. Each session is videotaped to check reliability of looking times.

Trials commence with the onset of a blinking centre light to orient the infant’s attention. During the familiarisation phase, the infant is familiarised with repetitions of the target stimuli playing on each side, until they have listened to each item for 30 seconds. In the test phase, the infant must produce a head-turn of at least 30 degrees in the direction of each sidelight for the corresponding auditory stimulus to play. Trials are terminated when the infant turns away from the target by 30 degrees in any direction, or when the stimuli for that trial have finished.
The HPP has proved adaptable to many experimental questions and this is one of its great advantages. Furthermore, drop out rates are typically lower than in many other procedures and average around 15-20% (Polka et al., 1995). The procedure has been used successfully with babies ranging in age from 4 to 11 months.

2.1.4 Conditioned Head-turn Procedure

The Conditioned Head-Turn (CHT) procedure is a technique for assessing speech discrimination in infants between 5½ and 18 months of age (Polka et al., 1995). It involves operantly conditioning an infant to produce a head-turn response toward a visual reinforcer when there is a change in speech sound. The CHT procedure has been reviewed by a number of researchers including Polka et al. (1995) and (Juszczyk, 1997) and these form the basis of the following description. General details about the traditional two-experimenter procedure are given below with more details of this method to follow in section 4.2.4.1. The new one-experimenter version of this procedure - the procedure of choice for this thesis - is introduced in section 4.2.4.2.

The CHT procedure typically used for speech research (e.g., Grieser & Kuhl, 1989) has a conditioning phase and a discrimination phase. During conditioning infants are trained to discriminate between two speech sounds, for example [ba] and [da]. At the start of the session, a repeating background stimulus (one of the speech sounds) is played from a loudspeaker. When the infant is looking towards the experimenter the sound may change from the background stimulus for a brief interval (e.g., three to four seconds) to the target sound (a change trial). On the first few trials the reinforcer is activated simultaneously with target onset to encourage the infant to turn their head towards the reinforcer when they hear a change in sound. Gradually a delay is introduced between the target onset and activation of the reinforcer, so that the infant learns to make anticipatory head-turns towards the reinforcer if they perceive a change in the sound. An observer (blind) in an adjacent room observes the infant (via one-way mirror or real-time video), and judges when they produce a head-turn. In order to proceed to the discrimination phase, infants must reach some criterion of correct responding (e.g., three consecutive correct head-turn responses on trials in which a target sound occurs).
In the discrimination phase the infant receives an equal number of change (test) and no-change (control) trials. Successful discrimination of sound change is indexed by a criterion number of ‘hits’ (correct head-turns) set by the experimenter. Kuhl (1983) and Grieser and Kuhl (1989) set this criterion at 9 correct out of 10 consecutive trials; while Werker and Lalonde (1988) used a criterion based on how many correct responses occurred within a fixed number of trials. A ‘miss’ is scored if infants fail to turn their head on test trials (Grieser & Kuhl, 1989), a ‘correct rejection’ is scored if infants do not turn towards the visual reinforcer on control trials, and a ‘false-positive’ is scored if the infant produces a head-turn to a control trial.

Various modifications of the CHT have been used. Discrimination of native and nonnative phonetic contrasts has been investigated at various ages from infancy to adulthood using the CHT procedure and variations of it. For adults the CHT procedure was adapted such that a button-press was required (rather than a head-turn) to a stimulus change, and the reinforcement consisted of lights illuminated in a box but no toy activation (Werker & Tees, 1984a).

Werker and Tees (1984a) also used the CHT to investigate generalisation of discrimination. They trained infants to discriminate between a pair of exemplars from different phonetic categories and subsequently tested discrimination of the contrast using multiple exemplars produced by a single talker. Thus infants were assessed on their ability to detect between-category phonetic differences and ignore within-category acoustic variability. In addition to being modified for generalisation of discrimination, the CHT has been adapted to study infant categorisation and perceptual organisation within a single vowel category, as opposed to cross-category categorisation (see Grieser & Kuhl, 1989; Kuhl, 1991; Kuhl & Iverson, 1995; Kuhl, Williams, Lacerda, & Stevens, 1992). The introduction of a third phase, the generalisation phase, has made it possible to study such issues as within and cross-language perceptual equivalence, for example, perceived similarity between a target variant of /i/ and the prototype of /i/.

The CHT procedure is a powerful technique for studying speech perception in infants, and with some modifications, subjects across a broad range of ages. However, the
procedure is not suitable for testing infants younger than 5½ months due to the need for voluntary control of head-turns, and the duration of the procedure (typically around 10-15 minutes); nor is it suitable for conveying speech information longer than one word because it is difficult to measure head-turn responses over an extended temporal window.

2.1.5 Summary
This section has described some of the common methods for obtaining speech perception data with young infants. Polka and colleagues (1995) state that the strongest inferences of speech perception ability can be obtained from the results of multiple test trials gathered using the CHT procedure and that a particularly powerful aspect of this method is the separation of the stimuli and reinforcers in this paradigm, so that head-turns in response to a stimuli change are rewarded as appropriate by the independent presentation of a visual reinforcer. This is not the case in the visual habituation paradigm where the stimulus and reinforcer are intertwined. The CHT procedure is the procedure of choice in this thesis and its particular instantiation here is outlined in more detail in chapter 4.

2.2 Methods for Investigating the Development of Speech Perception
Given newborn infants’ remarkable sensitivities to speech, how does speech perception change as a function of age and linguistic experience? The two main methods for investigating development, including speech perception development, are the Ontogenetic Method and the Differential Experience Method (Burnham & Sekiyama, in press).

2.2.1 The Ontogenetic Method
The ontogenetic method involves comparing the performance on a particular task by individuals of different ages reared in functionally similar environments (Burnham & Sekiyama, in press). In this way the relative roles that amount of experience (perceptual, linguistic, cultural etc) and maturation (cognitive skills and abilities), may have, can be identified. With regard to speech perception, the ontogenetic method is

37
useful for investigating such developmental changes as the perceptual dimensions governing infant preferences (e.g., prosody, phonetics), changes in infants’ perceptual ability for perceiving native and nonnative speech sounds, and matching sound and word meanings.

2.2.2 The Differential Experience Method

The differential experience method involves comparing performance on a particular task by individuals of the same age brought up in functionally different environments on a particular task, in order to investigate the impact of the type of experience (perceptual, linguistic, cultural, etc) on development (Burnham & Sekiyama, in press).

Regarding the differential experience method and speech perception, the world’s languages differ with respect to phonology, syntax, semantics and pragmatics; and experience with a particular language structure modifies the manner in which speech is perceived. For example, Burnham and Sekiyama (in press) have found that English-speaking adults have difficulty perceiving lexical tone differences (‘pitch’ variations) that are used contrastively in tone languages such as Thai and Cantonese, but do not signal word meaning in English.

The characteristics of the two types of developmental studies are summarised in Table 2.1, taken from Burnham and Sekiyama (in press). As can be seen, a particularly powerful method would be to combine the ontogenetic and experiential methods. In this thesis a combination of both methods is used. Armed with technique and method for testing infants’ speech perception capacities, it is apt to move onto the literature, but first the questions raised in chapter 1 regarding the problem of speech perception and categorical perception, need to be answered.
Table 2.1

*The Separate and Combined Features of the Ontogenetic (Amount) and the Differential Experience (Type) Methods for Investigating Linguistic Development*.

<table>
<thead>
<tr>
<th>Ontogenetic (Amount)</th>
<th>Differential Experience (Type)</th>
<th>Ontogenetic Plus Differential Experience (Amount and Type)</th>
</tr>
</thead>
<tbody>
<tr>
<td>By the judicious choice of ages, the effect of linguistic experience and maturation on development can be evaluated</td>
<td>By the judicious choice of languages, the effect of particular linguistic structures on development can be evaluated</td>
<td>By the judicious choice of ages and languages, the effect of linguistic experience, maturation and linguistic structures on development can be evaluated</td>
</tr>
</tbody>
</table>

*Table reproduced from Burnham and Sekiyama (in press). Permission to reproduce this table was obtained from the first author.

2.3 The Problem of Speech Perception and Infants’ Solutions

Earlier, section 1.4.1 dealt with how ‘expert’ users of language, adults, cope with the problems of speech perception. However as discussed in the following sections, there is evidence that even infant language learners can handle these problems of speech perception, including invariance, contextual effects, and talker variability.

2.3.1 Invariance and Contextual Effects

The acoustic properties associated with a particular phoneme can vary greatly depending on its context, meaning that different acoustic information can signal the same phonetic contrast across contexts (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). This phonetic equivalence of acoustic correlates across contexts suggests that listeners ‘trade-off’ these cues. Both adults (Best, Morrongiello, & Robson, 1981; Moore, 1997; Repp, 1982) and infants have been found to do this.
Jusczyk and Derrah (1987) habituated 2-month-old infants' sucking rate using a series of syllables sharing the same phonetic segment, for example [bi], [ba], [bo], and then introduced either a new syllable sharing the same initial phoneme, for example [bu], or a syllable not sharing the same initial phoneme with the other syllables, for example [du]. It was found that infants increased their sucking rate to both the [bu] and [du] type tokens showing that infants could recognise the perceptual identity of a phonetic segment across phonetic contexts. Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, and Mehler (1988) confirmed this finding with 2-month-olds, who noticed the addition of a new syllable differing in either consonant or vowel. However, tests of newborn infants showed that the newborns only increased their sucking rate to novel syllables when the vowel but not the consonant differed, indicating that vowels are more important than consonants in early infancy. Nonetheless, Kuhl (1980) found that 6-month-old infants identify changes in fricative contrasts despite changes in vowel context, so it appears that the ability for invariant perception across phonetic variability may develop sometime between 2 and 6 months of age.

The ability to categorise similar speech sounds is also an important skill in coping with the invariance of speech, and infants have been found to do this. The crucial issue of categorisation is the recognition of similarity between phonetically equivalent but acoustically different members of the same category (Kuhl, 1983, 1993). In other words, categorisation involves mapping complex acoustic cues onto the perceived attributes of sound in order for the infant to learn about which changes in phonetic category are meaningful and important for communicating successfully in the native language (Jusczyk, 1997). Kuhl (1983) enumerates two characteristics of speech categorisation. Firstly, the stimuli used in speech categorisation studies vary along a number of dimensions such as phonetic identity, pitch contour and talker identity and these differences must be perceived by the infant. Secondly, infants must produce equivalent responses to stimuli perceived to be similar in order for successful categorisation to be inferred. Kuhl (1980) and Hillenbrand (1983) discovered that infants demonstrate an ability to perceive categories based on the initial or the final consonant of sets of syllables, such as those that start with /s/ as opposed to those that start with /sh/ and /m/ vs. /n/ respectively. While there is perceptual constancy for
phone classes, other categorisation skills are also important. One of these is speaker
gender and Miller (1983) has shown that 2-month-old and 6-month-old infants
categorise voices on the basis of the speaker’s gender.

Together these studies show that infants perceive the phonetic and voice quality
(gender) equivalences by 6 months of age and that this develops some time between 2
and 6 months.

2.3.2 Talker Variability

Without the ability to recognise the same word produced by different talkers, every
acoustic difference could potentially signal a difference in meaning, and language
acquisition would be virtually impossible. Here variability due to talker identity, pitch
and speaking rate will be considered. Kuhl and Miller (1982) used the High-
Amplitude Sucking (HAS) procedure to conduct the first investigation of how infants
deal with talker variability. In the pre-shift phase of the experiment, 1- to 4-month old
infants were exposed to two tokens of the vowel [a], one spoken in a monotone and the
other with a rise-fall pitch contour. In the post-shift phase infants were found to be able
to discriminate the vowel contrast [a] vs. [i] despite irrelevant variations in pitch (flat
vs. rise-fall), but were unable to discriminate a pitch change when the vowel colour was
varying. The results show that young infants can deal with variability in speech,
although this ability may be limited to particular parameters, in this case for vowel
identity but not pitch contour.

More recently, Jusczyk, Pisoni, and Mullennix (1992) discovered that infants as young
as 2 months demonstrate some capacity to handle talker variability. Infants were
presented with 12 tokens of the words /bug/ (produced by six male speakers and six
female speakers) and 12 tokens of /dug/ by the same 12 speakers. Infants significantly
discriminated the words despite the talker variation. However, when a two-minute
delay period was introduced between the presentation of the /bug/ and /dug/ tokens,
infants could not discriminate the contrast. Thus infants show perceptual normalisation
for talker variation only when the processing resources for encoding and retrieving
speech are not restricted.

41
Kuhl (1979) conducted a more extensive investigation of how older infants handle talker and pitch variability. Six-month-old infants were trained using the conditioned head-turn procedure to discriminate the vowel contrast [a] and [i] when talker voice identity and pitch contour were fixed, then they were tested for their ability to maintain discrimination of [a] and [i] in the face of talker and pitch variability. Variability along these dimensions was increased throughout the session and infants continued to discriminate the vowel contrast even with the maximum level of variability (three talkers, two pitch levels). In a further study, Kuhl (1983) tested infants on the discrimination of the two vowels [a] and [o], which have considerable acoustic overlap in productions by some speakers. It was clear that infants could discriminate this contrast even with the maximum degree of speaker and pitch variability (as in Kuhl, 1979).

2.3.3 Summary

The aim of this section was to show that young infants demonstrate perceptual constancy for speech sounds across talkers and pitch variance. Although 2-month-old infants find it difficult to perceive phonetic identity in the face of contextual variability, they appear to take account of such contextual variability by 6 months of age.

2.4 Categorical Perception by Infants

The initial impetus for investigating phonetic perception in infants arose out of an attempt to understand adult speech perception and to determine whether infants display the same kind of categorical perception for speech sounds. Before experimental study of infants was possible, categorical perception of speech sounds was assumed to be a result of extensive linguistic experience (Liberman, Harris, Hoffman, & Griffith 1957). It was the arguments of Chomsky and Fodor, that certain linguistic abilities were species-specific and had an innate rather than an experiential basis, which led to the investigation of speech perception abilities in infants.
2.4.1 Consonants

This section reviews the categorical perception of consonantal voicing (2.4.1.1), manner (2.4.1.2) and place of articulation (2.4.1.3) distinctions. This is followed by a discussion of whether infants’ speech perception is really categorical, with due regard to the lack of reliable perceptual identification data as the reason (2.4.1.4).

2.4.1.1 Voicing

Eimas et al. (1971) conducted the first study investigating infants’ perception of voicing contrasts. They used the HAS procedure to test pre-linguistic infants as young as one month of age for their ability to discriminate VOT differences in bilabial stops. Following the presentation of stop consonants from one phonetic category, the infants elicited greater sucking rates to a new stimulus from a contrasting (adult) phonetic category (between-category contrasts, /ba/ vs. /pa/) but not from the same adult phonetic category (within-category contrasts, /ba1/ vs. /ba2/). Similar results have been found in later studies by Aslin and Pisoni (1980), Eilers (1980), and (Aslin, Pisoni, and Jusczyk (1983), and together these findings have been taken to suggest that infants, who had not been subjected to a long period of specific language experience, perceive speech in a manner approximating categorical perception. Consequently, Eimas et al. (1971) and later Eimas (1975) proposed that the mechanisms underlying speech perception are innate and that humans are genetically endowed to perceive speech signals in a linguistic mode.

The convergence of several lines of research has led to a reappraisal of the conclusions originally drawn by Eimas et al. (1971) that categorical perception is a purely linguistic phenomenon. Subsequent research strongly supports the role of experiential factors in mediating the perception of phonetic categories (Aslin et al., 1981; Walley, Pisoni, & Aslin, 1981). Indeed the early work by Lisker and Abramson (1964) demonstrates that there are cross-language similarities and differences (i.e., experiential influences) in the discrimination of adult VOT boundaries (see section 1.5.1.1). Thus, regardless of any genetic predisposition at some point categorical speech perception must become less general and more language specific. For example, it appears that the surrounding
linguistic environment influences the location of VOT boundaries and cross-language differences between voiced and voiceless phoneme classes.

In this regard, cross-language research by Aslin et al. (1981) using a staircase testing version of the head-turn procedure showed that English language environment infants (6-12 months) are able to discriminate synthetic bilabial stops that straddle the voiced/voiceless (phonemic for English listeners), and prevoiced/voiced (non-phonemic for English listeners) boundaries. That is, English-language environment infants were capable of discriminating phonologically irrelevant voicing contrasts in both the positive (native) and negative (nonnative) regions of the VOT continuum. Likewise, adults from an English-language environment were also found to discriminate VOT differences in the negative VOT region at above chance levels. Two aspects of the comparison of infants and adults are important. First, adults who unarguably have greater linguistic experience than infants, showed heightened sensitivity to VOT differences at regions of the continuum associated with boundaries between phonemic categories. Secondly for infants to discriminate VOT differences successfully, the magnitude of difference between the target and background was consistently greater than it was for adults. This provides evidence that the native phonological system is selectively modified over development and there is a transition from acoustic to phonemic processing. Aslin and Pisoni (1980) address this in their model of speech perception, discussed later in section 2.8.1.1.

2.4.1.2 Place of Articulation

Moffitt (1971) and Morse (1972) found that 2- and 5-month-old infants respectively discriminated the /ba/ vs. /ga/ place contrast. Eimas (1974) built on these findings by testing 2- to 3-month-old infants on a similar contrast [bae] vs. [dae] and found that this was perceived in a categorical manner. In a study with French newborn infants Bertoncini, Bijeljac-Babic, Blumstein, and Mehler (1987) demonstrated that infants discriminate place contrasts, suggesting that the mechanisms for discriminating place of articulation contrasts are in place at birth.
Research into the categorical discrimination of place of articulation contrasts has not been limited to the [b], [d], [g] type. Two-month-olds can also reportedly discriminate [ma] vs. [na] (Eimas & Miller, 1980b), [wa] vs. [ja] (Jusczyk, Copan, & Thompson, 1978), [v] vs. [f], and [f] vs. [g] (Levitt, Jusczyk, Murray, & Carden, 1988), while 6-month-olds have been found to discriminate [fâ] vs. [ga] (Holmberg, Morgan, & Kuhl, 1977) in a categorical manner.

2.4.1.3 Manner of Articulation

The first investigation of manner of articulation discrimination was conducted by Hillenbrand, Minifie, and Edwards (1979) who found that 6- to 8-month-olds could discriminate [ba] vs. [wa]. In a study of younger infants, Eimas and Miller (1980a) and Miller (1983) found that 2-month-olds were also sensitive to the [ba] vs. [wa] contrast. Moreover Eimas (1975) found that [ra]-[la], could be discriminated categorically by 2-month-old American English language environment infants and adult native speakers of American English, whereas Eimas and Miller (1980b) found inconclusive evidence for categorical perception when they tested 2- to 4-month-old infants on discrimination of [ba] vs. [ma]. Thus for infants’ discrimination of manner of articulation contrasts, the evidence is both for and against categorical perception.

2.4.1.4 Is Infants’ Perception of Consonants Really Categorical?

The results of the Eimas et al. (1971) study and similar studies that followed were taken to indicate categorical perception in infancy. This claim has been criticised by Burnham, Earnshaw, and Quinn (1987) who argue that infant identification data is required in order to maintain a consistent definition of categorical perception which includes both discrimination and identification functions. They propose that different functions are involved in identification and discrimination: Discrimination functions indicate areas of heightened perceptual ability but cannot be used to interpret whether the tokens have been categorised into distinct categories. Burnham, Earnshaw, and Clark (1991) developed an infant identification procedure (ISI) and tested 9- to 11-month-old infants on the identification of bilabial stop contrasts varying in VOT from –
70ms to +70ms in 10ms steps\textsuperscript{13}. Infants were able to perceive the contrast in the positive VOT region more easily than in the negative region but this identification of bilabial stops was not categorical\textsuperscript{14}. Thus while Aslin et al. (1981) argue that infants discriminate \textipa{/b/} vs. \textipa{/\textipa{-}b/}, Burnham and his colleagues provide evidence that infants have difficulty with identification of this non-phonemic contrast. The dissociation between discrimination and identification of these bilabial contrasts is also evidence that different underlying processes mediate discrimination and identification.

\subsection*{2.4.2 Vowels}

Much of the research on infants’ early speech perception has focussed on consonantal contrasts, but there are also studies investigating infants’ vowel perception. In the first investigation Trehub (1973) presented natural speech tokens of two different vowel pairs, [a] vs. [i] or [i] vs. [u] to 1- to 4-month-old infants. The infants were able to discriminate both pairs of vowels. Kuhl and Miller (1982) repeated Trehub’s findings for the [a] vs. [i] pair and also found that infants continue to discriminate this pair even in the face of irrelevant pitch variation. While these studies did not investigate whether or not infants’ perceive vowels in a categorical manner, they were an important first step in highlighting infants’ discrimination of steady-state vowels.

Swoboda, Morse, and Leavitt (1976) presented 2-month-olds with a more subtle contrast - [i] and [I] - and tested infants’ discrimination of this contrast and whether perception was categorical or continuous. Not only did the infants show between-category discrimination but they also perceived vowels continuously- they were able to discriminate within-category differences between the vowels. Hence, just like adults (e.g., Pisoni, 1973) the infants displayed continuous perception of a vowel series.

\textsuperscript{13} 2-year-olds, 6-year-old and adults were also tested. Discrimination data was not obtained in this study but was compared to previous discrimination data, for example Aslin et al. (1981) and Elmas et al. (1971).

\textsuperscript{14} Speech perception became more categorical across development and with linguistic experience. Children older than 2 years of age, and adults, elicited categorical-like identification functions.
2.4.3 Nonspeech

In an effort to circumvent the possibility that humans may respond categorically to all complex acoustic signals rather than using a ‘special’ mode to process speech, Mattingly, Liberman, Syrdal, and Halwes (1971) assessed infants’ discrimination for speech and nonspeech sounds of equal complexity. The nonspeech sounds (chirps) were isolated presentations of second-formant transitions. Flat discrimination functions were obtained for nonspeech indicating that the infants discriminated within-category pairs as well as between-category pairs, whereas the discrimination functions for the speech stimuli showed sharp peaks in the regions of the phonetic boundaries. Thus infants’ discrimination of the nonspeech sounds appeared continuous but their discrimination of speech appeared categorical.

In a similar vein, Morse (1972) attempted to determine how infants respond to the same acoustical differences in speech and nonspeech contexts. He examined the perception of the place contrast [ba] vs. [ga]. The speech stimuli were F2 to F3 transitions followed by vowel formants, while the nonspeech stimuli were isolated chirps composed of only the transition portions from the second and third formants. Investigation of differences in the infants’ discrimination performance for the speech and nonspeech stimuli (rather than categorical perception) showed that there was no difference in the level of responsiveness to the speech and nonspeech stimuli. However, Morse argued that infants were responding in a different way to the speech and nonspeech sounds because half of the infants gave evidence of discrimination and half did not. In an earlier study Eimas et al. (1971) suggested that infants have innate and specialised mechanisms for speech processing and respond to speech in a special way, but their study did not allow for a direct comparison of infants’ responses to speech and nonspeech sounds. Morse’s findings (1972) however, were taken as evidence that infants possess innate and specialised mechanisms for speech processing.

As discussed in section 1.5.3, the long standing ‘speech is special’ argument has been challenged by demonstrations that speech and nonspeech stimuli are processed similarly. Following on from the Pisoni (1977) study showing categorical perception for two-component tones, Jusczyk, Pisoni, Walley, and Murray (1980) tested 2-month-olds on stimuli from the same nonspeech series, and found infants could discriminate
only between-category differences. A subsequent investigation by Jusczyk, Rosner, Reed, and Kennedy (1989) directly compared 2-month-olds’ discrimination of voicing differences for speech sounds with their discrimination of temporal-order differences with nonspeech sounds. The location of the category boundaries for the speech and nonspeech stimuli were similar, leading Jusczyk et al. (1989) to conclude that the same general auditory mechanisms may underlie infants’ perception of both speech and nonspeech sounds.

2.4.4 Summary
Despite extensive research, complete understanding of the mechanisms of speech perception has eluded researchers. Nevertheless, data from categorical speech perception studies strongly suggest that a definable mechanism, whether speech specific, or general, and whether human-specific, or general, exists in humans and is operational in early infancy.

2.5 Early Linguistic Sensitivities
Moving onto speech perception development, this section reviews what capacities exist at birth and how they change over age. Infants have three clear preferences at birth - their native language, their mother’s voice, and infant-directed speech (IDS), and each appears to be based on prosody. Ramus and Mehler (1999) define prosody as rhythm plus intonation. Infants’ preference for mother’s voice, native language, and IDS are discussed below, and the section concludes by looking at the role of prosody in early speech perception.

2.5.1 Preference for Mother’s Voice and Speech Patterns
It has been suggested that early auditory competency subserves bonding. DeCasper and Fifer (1980) provide the earliest demonstration that infants younger than four days prefer to listen to their mother’s voice over the voice of a stranger. In an adaptation of the HAS procedure, infants increased their inter-burst intervals (IBIs) when this was instrumental in producing mother’s voice over the voice of a female stranger. When the response requirements were reversed, the probability of the infant terminating IBIs
was highest when this led to the presentation of the maternal voice. In a second
experiment DeCasper and Fifer (1980) used a different discrimination task to measure
newborn’s preference for the maternal voice. Infants learnt that sucking when a pure
tone was presented terminated the tone and produced their own mother’s voice,
whereas sucking during the period of silence led to the presentation of a non-maternal
voice recording. These results show that within three days of postnatal development,
infants are able to discriminate between speakers, prefer their mother’s voice in
comparison to other female voices, and suggest that infants have learned about their
mothers’ voice in the womb.

A direct demonstration of the link between prenatal auditory experience and subsequent
postnatal preferences comes from DeCasper and Spence (1986) who hypothesised that
following prenatal experience with a target passage this passage would be more
reinforcing for postnatal newborns than a novel passage. The mother read a story to the
prenatal infant twice a day for the last 6 weeks of the pregnancy. Subsequently, the
postnatal newborns preferred the target story, presumably because in utero they heard
and remembered aspects of the target passage. Moreover, infants preferred the target
story independent of who recited it. The findings imply that 3-day-old infants retain
two types of information from prenatal sensory experience for at least several days after
birth: Specific acoustic information about their mother’s voice (perhaps fundamental
frequency patterns) and general language-relevant cues from prenatal exposure to
specific language input.

2.5.2 Native Language Preferences

Moon, Cooper and Fifer (1993) employed a HAS preference procedure to determine if
2-day-old infants preferred to listen to their native language over a foreign language.
Sucking patterns controlled auditory presentation. Sucking to one signal, for example,
/a/, produced the native language recording and sucking to another signal, /i/, produced
a nonnative speech recording. Infants activated recordings of the native language for
longer periods than the foreign language. This differential responding is evidence for a
voluntary preference for general properties of the native language. Moon and
colleagues findings suggest that 2-day-old infants recognise their language and prefer to
listen to it than an unfamiliar language, quite possibly from prenatal experience with the language.

Bosch and Sebastian-Galles (2001) expanded previous work on infants’ native-language recognition abilities by testing the capacity of 4-month-old infants either from Catalan or Spanish speaking language environments to distinguish between the native and nonnative languages, and to determine if discrimination is achieved using prosodic or segmental information. Catalan and Spanish languages are prosodically similar at the phonological phrase level, but are different at the segmental and syllable levels. They found that infants discriminated between the native and nonnative utterances, and showed a native-language preference. In addition, infants made this language distinction even when the utterances were low-pass filtered (which has the effect of removing segmental cues, and leaving only prosodic cues). Thus the differential prosodic cues in the two languages were sufficiently salient to be used by infants for native language recognition. Additionally, differences at the syllable level between the languages, for example vowel reduction in Catalan but not Spanish, and differences in stress patterns - Spanish has strong-weak stress (as in the English word DOctor) and Catalan has weak-strong stress (as in docTOR) - may have contributed to infants’ ability to detect differences on the basis of language rhythm. There are also differences in stress patterns between the two languages at the phonological word level that could also influence preferences for listening to the native language. Stress patterns in language are discussed in detail later in section 2.6.1.

In conclusion, there is a large body of evidence indicating that during the first year of life, infants discover many facets of their native language including prosodic, phonetic and phonotactic patterns. Recognising the properties of the native language will not only influence learning about the phonological organisation of the language, but also the development and organisation of the mental lexicon.

### 2.5.3 The Role of Prosody in Early Speech Preferences

There is considerable evidence that the late-term human fetus is responsive to the prosodic characteristics of the mother’s speech patterns because the auditory system is
mature at 6 months conceptional age. The uterine wall low-pass filters acoustic signals meaning that prenatal infants receive only low frequency information (Jusczyk, 1997). The argument is that this low-pass filtered information bootstraps the infant to attend preferentially to acoustic signals exhibiting the same rhythmic patterning and pitch contours of the human voice. These prosodic cues appear to play a role in the infant’s acquisition of native language grammar (Jusczyk, Hohne, & Mandel, 1995), and information available prenatally has been shown to have an effect on subsequent postnatal perceptual sensitivities (Cooper & Aslin, 1989; 1990).

Further studies of early linguistic sensitivities have investigated more precisely the basis for infants’ preference for a familiar language. Mehler et al. (1988) showed that 4-day-old native French newborns could discriminate Russian from French in both natural speech and low-pass speech samples, but there was no evidence of their ability to distinguish utterances from two foreign languages, English and Italian. Hence, when distinctive phonetic information is removed, prosodic information, that is, rhythm and intonation, is sufficient for distinguishing native language strings from foreign language strings. Similarly, native English 2-month-olds could easily discriminate English from Italian but were unable to respond differentially to two nonnative languages. It seems unlikely that the failure to discriminate two nonnative languages is due to the infant’s inability to process segmental differences, as young infants have the ability to discriminate various speech contrasts be they native or nonnative (Aslin et al., 1981; Eimas et al., 1971; Werker, 1994). Rather, discrimination between two languages may be dependent on familiarity with at least one of the two languages. It is also possible that young infants use prosodic more than phonetic information. The aim of this section is to show that early speech preferences are based on prosody.

Ramus, Hauser, Miller, Morris, and Mehler (2000) tested human newborn Dutch infants and cotton-top tamarin monkeys’ on the discrimination of natural Dutch (stress-timed) and Japanese (mora-based) utterances. Tamarins and human newborn infants were able to discriminate the Dutch and Japanese utterances despite speaker variability, suggesting that both groups could extract linguistic invariants amongst the speakers. However after sentence resynthesis, which involved changing all fricatives to /s/, vowels to /a/, liquids to /l/, plosives to /t/, nasals to /n/ and glides to /j/ while preserving
Speech perception during the 1st year

F0 and phoneme duration, infants but not tamarins, could discriminate the languages. Ramus et al. (2000) state that one possibility for this finding is that the infants were responding to prosodic features of speech whereas the tamarins were more sensitive to phonetic contrasts. In addition to these findings, both the tamarins and the infants failed the language discrimination task when sentences were played backwards, suggesting that discrimination is reliant on speech-specific cues rather than a general auditory capacity. This result reinforces the ‘rhythmic hypothesis’ - that young infants should discriminate nonnative rhythmic classes from the native rhythmic class - and reiterates that rhythm is sufficient for infants to discriminate languages.

In a study with older infants, Nazzi, Jusczyk and Johnson (2000) investigated 5-month-old American infants’ ability to discriminate native language passages from nonnative mora-timed Japanese passages. Five-month-old infants were successful in discriminating the stress-based native language from a different mora-timed rhythmic class. Furthermore, when 5-month-old infants were tested on their ability to discriminate Italian (syllable-timed) and Japanese (mora-timed), the infants were able to discriminate the two languages, even though their native language English does not have either of these rhythmic patterns. Interestingly, infants could not discriminate between two languages from the same nonnative rhythmic class, but were able to discriminate two foreign languages from the native (stress-based) rhythm class. The findings suggest that experience with foreign languages from different rhythmic classes is not required for infants to be able to distinguish between them, but it is difficult for infants to distinguish foreign languages from the same nonnative rhythm class.

Overall it appears that infants’ early speech preferences are based on the prosody of the speech, and more specifically on its rhythm. The abovementioned studies highlight three important points about infants’ sensitivity to rhythm:

1. Prenatal experience with native language rhythm influences postnatal language preferences

2. From birth, infants are capable of discriminating native language rhythm from nonnative language rhythm, and
3. Infants have difficulty discriminating two foreign languages from the same rhythm class.

The evidence suggests that infants' early sensitivity to rhythm may assist them to bootstrap other properties of the native language. It also appears that there is a rich display of the rhythmic and intonational properties in a special speech style, infant-directed speech or motherese, which parents use with infants.

2.5.4 Infant-directed Speech and the Role of Emotion in Early Preferences

Mothers and fathers communicate with infants using infant-directed speech, a speech register in which various acoustic, phonetic and affective characteristics are enhanced (Fernald et al., 1989). Across languages and cultures and to varying degrees, infant-directed speech (IDS) has an increased fundamental frequency, wider pitch range, exaggerated pitch contours, slower tempo, and more hyperarticulated vowels than typical adult-directed speech (e.g., Burnham, Kitamura & Vollmer-Conna, 2002; Fernald & Kuhl, 1987; Fernald et al., 1989; Grieser & Kuhl, 1988; Kuhl et al., 1997; Stern, Spieker, & MacKain, 1982). These exaggerated intonation patterns are thought to have three purposes - to communicate affect (Kitamura & Burnham, 1998; Werker & McLeod, 1989) and facilitate social interaction (Werker & McLeod, 1989); attract, engage and maintain infant attention (Stern et al., 1982); and facilitate language acquisition by exaggerating lexical and grammatical structure (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989) and aiding the acquisition of vowel categories in the language.

Regarding the third of these, Kuhl et al. (1997) showed that English-speaking mothers have bigger vowel triangles, that is, more hyperarticulation of the corner vowels /a/, /i/, and /u/, as compared with adult-directed speech (ADS). Vowel hyperarticulation appears to be specific to IDS rather than to highly emotional speech, because bigger vowel triangles have been found for IDS than pet-directed speech (Burnham et al., 2002), and there is evidence that greater vowel hyperarticulation in IDS assists infants’
speech discrimination skills (Liu, Kuhl, & Tsao, 2003). Thus IDS functions to increase speech discrimination. But are there other functions of IDS?

When given a choice of what they want to listen too, infants from as young as a few days old to several months of age prefer to listen to IDS over ADS (Cooper & Aslin, 1990; Fernald, 1985; Hayashi, Tamekawa, & Shigeru, 2001; Werker & McLeod, 1989). In an analysis of the acoustic cues responsible for this, Fernald and Kuhl (1987) showed that infants’ IDS preferences are based more on pitch than amplitude and duration pattern. However, more recently Kitamura and Burnham (1998) have shown that the affective qualities of IDS are important in determining infants’ IDS preferences; when pitch was equated and affect manipulated, infants preferred high affect over low affect utterances but when affect was equated and pitch manipulated, infants showed no preference. This shows that while pitch contours may be important in determining infants’ preferences for IDS, it may be that its importance lies in its ability to act as a vehicle for heightened affect. Because independent studies have documented the influences of acoustic and emotional features on infants’ attention to IDS, recent work by Cooper, Kitamura, Mattock, and Burnham (under review) directly compared the relative influence of speech tempo and degree of positive affect on infants’ preferences for infant-directed speech over age. Cooper et al. (under review) tested 16-week-old and 30-week-old infants’ speech preferences in three experiments: (a) normal tempo IDS vs. slow tempo IDS, both with high affect, (b) high affect vs. low affect IDS, both with slow tempo, and (c) high affect, normal tempo IDS vs. slow tempo, low affect IDS. When IDS was matched for affect but tempo varied, the younger infants preferred to listen to slow speech, but when tempo was matched and affect varied, these younger infants showed no preference. Cooper and colleagues suggest that for younger infants slow speech is perceived in a manner similar to speech that has positive affect, that is, as soothing or comforting (see also Trainor, Austin, & Desjardins, 2000). Older infants on the other hand, were found to attend more to high affect IDS that was normal tempo, a possible sign of their increased linguistic sensitivity (i.e., speech is not usually so slow) and their increased attention to phonetic over prosodic cues in speech (Burnham, Kitamura, & Lancuba, 1999).
The studies reviewed in this section show that prosody is an important factor in the development of infant speech perception across the first year of life. Young infants are perceptually sensitive to the suprasegmental aspects of speech and this figures centrally in young infants’ preferences for their mother’s speech, their native language, and infant-directed speech. Infants’ early sensitivity to the salience of prosody in speech and native-language rhythm opens the door for them to acquire knowledge about native-language linguistic structure, stress patterns and word segmentation, topics discussed in the following section. The next section focuses on infants’ perception between 6 and 9 months of age, the age at which they are beginning to ‘honed in’ on their native language.

2.6 Recognising the Suprasegmental and Phonotactic Properties of the Ambient Language: The Origins of Word Recognition

A number of major changes related to word and utterance recognition occur between 6 and 9 months of age, and this section is concerned with infants’ recognition of native language suprasegmental and phonotactic properties. For meanings of utterances to be understood the location and recognition of words and their syntactic function is critical. For adults, acquired lexical knowledge supports speech segmentation, whereas for pre-linguistic infants, segmentation of the speech stream must be based on non-lexical cues (Friederici & Wessels, 1993). Thus, while the infant comes into the world equipped to deal with language, it is unlikely that they can use information about word structure because every human language has a different word organisation and infants must first begin to segment the speech stream. To do this, infants need to learn where the words are in the speech stream. According to Friederici and Wessels (1993) there are two major cue types used by the infant for initial speech segmentation and identification of word boundaries: suprasegmental prosodic features and phonotactic features. The role of these cues is the focus of this section.

2.6.1 Preference for Native Language Stress Patterns

Results of several studies, (e.g., Jusczyk et al., 1999; Jusczyk, 1993; Jusczyk, 1992) support the notion that infants are sensitive to the stress patterns of native language
words - the kinds of regularities in native language input that could prove useful in segmenting words from fluent speech. Infants’ attention to the diverse rhythmic cues suggests that infants not only attend to rhythmic patterns in speech, but also to the frequency of occurrence of these rhythmic units (Jusczyk, Cutler et al., 1993; Morgan, 1996). Cutler and Mehler (1993) have referred to the infant’s rhythmic sensitivity as an instance of a general periodicity bias whereby infants acquire sensitivity to the prosodic characteristics of speech (and the native language) such as rhythm and stress that are exploited by word segmentation strategies.

In English about 90% of words have a strong-weak stress pattern and being able to recognise this is a good cue to word segmentation. It is important to start with a discussion of how infants notice stress in words. Specifically, 9-month-old English language infants demonstrate a trochaic bias, preferring to listen to lists of disyllabic words with trochaic (strong-weak) stress patterns compared to lists of words with iambic (weak-strong) stress patterns. This preference is likely to result from experience with the ambient language, particularly since 6-month-old infants failed to show this preference (Jusczyk, Cutler et al. 1993). Therefore it seems unlikely that 4-month-olds in Bosch and Sebastian-Galles (2001) (discussed earlier in section 2.5.2), make their native versus nonnative distinctions on the basis of stress patterns. However, since only English-learning infants took part in this study, it is debatable as to whether this developmental change is due to language experience.

By employing the HPP Turk, Jusczyk, and Gerken (1995) repeated the Jusczyk et al. (1993) experiment (detailed earlier in section 2.1.3), using lists of disyllabic nonword stimuli matched on phonetic and phonotactic properties, that is, all the items were strong syllables with heavy syllable weight. American 9-month-olds listened significantly longer to lists of strong-weak bisyllables than weak-strong bisyllables. Even when strong syllables did not have a heavy syllable weight, infants continued to prefer strong-weak to weak-strong patterns indicating that syllable weight does not play a necessary role in the preference for strong-weak words. However, the results of a third experiment indicate that 9-month-old infants prefer heavy strong syllables, independent of their preference for words that begin with a strong syllable. Turk and colleagues (1995) concluded that the sensitivity to surface linguistic patterns (i.e.,
lexical stress) and the principles underlying them (i.e., syllable weight), develop independently in the first year of life.

Morgan (1996) provides additional support for the enhanced cohesiveness of trochaic sequences and a preference for trochaic sequences over iambic stress patterns emerging between 6 and 9 months of age. Using a variant of the CHT procedure, Morgan presented infants with ‘buzz noises’ between pairs of syllables in multisyllabic speech, while manipulating the prosodic and sequential properties of the strings to see if this influenced perception of cohesive trochaic sequences. Consistent with previous research (Echols, Crowhurst, & Childers, 1997; Jusczyk, Cutler et al., 1993), 6-month-olds did not perceive bisyllables (trochaic and iambic) as cohesive when presented with rhythmically novel but segmentally familiar stimuli in the test phase. However, the ability of 6-month-olds to perceive segmentally novel but rhythmically familiar trochaic and iambic bisyllables cohesively was evident. On the contrary, the 9-month-olds were differentially sensitive to different rhythmic patterns. These older infants were biased toward perceiving novel trochaic bisyllables as more cohesive units than novel iambic bisyllables, regardless of familiarity and novelty.

These results highlight that while young infants are sensitive to prosodic markings in the native language (as discussed in section 2.5.3), they may not acquire or use information about phonotactic patterns in their native language to detect dissimilarities between prosodic versions of two languages, until much older than 4 months of age. The studies reviewed here provide support for the notion that English-language infants come to adopt a trochaic metrical strategy for word segmentation by the end of the first year (Jusczyk, Cutler et al., 1993), and that the ‘trochaic bias’ has its roots in early linguistic experience.

2.6.2 Phrase Boundaries and Word Boundaries

The problem with word boundaries is that in most spoken language, few cues are available to signal where they are located (Cutler, 1996). Christophe, Dupoux, Bertoncini, and Mehler (1994) investigated whether French newborn infants demonstrate sensitivity to word boundaries and bisyllabic stimuli extracted from
naturally produced French utterances. The HAS procedure was used to assess whether 3-day-old infants can discriminate bisyllabic stimuli that either contain a word boundary or do not contain a word boundary. During the training phase, infants were presented with utterances of /mati/ from one word boundary condition, and were then tested on their discrimination of /mati/ from the other word boundary condition. The newborn infants showed discrimination of the cues rather than actually segmenting the words.

In related work, Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward and Piwotz (1992) tested whether 6-month-old and 10-month-old infants prefer to listen to speech that has pauses inserted naturally at major phrasal boundaries, or whether they show preference for speech segmented at an artificial location within a sentence. From a series of experiments, they found that 9-month-old infants are sensitive to the acoustic markers of major phrasal units in spontaneous speech, storybook samples with the intonational markings of natural IDS (e.g., long subject-noun phrases), and low pass filtered speech stimuli. In each instance, infants displayed a preference for samples segmented at a major phrasal boundary as opposed to samples where pauses were inserted in other locations within the subject-noun and predicate-verb phrases. The results with the low-pass filtered stimuli suggest infants are responding to information available in the prosody of the utterances. In contrast, younger 6-month-olds tested on the spontaneous speech and storybook samples did not show significant preference for samples segmented at coincident phrase boundaries. These results have interesting implications for understanding the developmental change in sensitivity to the acoustic correlates of major phrasal units that are available in the prosody of English utterances. Three-day-old infants may discriminate differences of this type, but between 6 and 9 months infants use this information and prefer what is most common in their environment.

2.6.3 Word Segmentation

During the second half of the first year, infants have been shown to be perceptive of the features of native language sound structure that are potentially useful in word segmentation. For example, Friederici and Wessels (1993) demonstrated that between
6 and 9 months of age, infants develop sensitivity to phonotactic properties that signal word boundaries in the native language. Their results demonstrate that 9-month-old Dutch infants listen longer to linguistically simple speech samples whose word boundary structure is legal in Dutch, than they do to speech samples where the middle syllable contained a phonotactically illegal word offset. In addition, the infants showed no preferential pattern for legal over illegal structures when the stimulus material was low-pass filtered, strongly suggesting that the observed preferences for legal over illegal phonotactic word boundary structures in the 9-month-olds are indeed due to phonotactic cues rather than prosodic cues. Similarly, Jusczyk and Aslin (1995) found that English-learning segment monosyllabic words from fluent speech. They demonstrated that 7½-month-olds familiarised with full vowel monosyllabic target words either in isolation or in a sentential context, subsequently listened longer to passages with these words than to ones without them.

Recently Jusczyk, Goodman and Baumann (1999) repeated and extended these findings by showing that 7½-month-olds can segment bisyllabic words with strong-weak stress patterns from fluent speech after being first familiarised with the target words in isolation. Furthermore, the infants responded to the strong-weak words as a whole rather than just to salient components of these words, as evidenced by the finding that when only strong syllables were used during familiarisation, but strong-weak words were presented during testing, there were no significant differences between the listening times for the stimuli. On the other hand, success at word segmentation was limited to segmenting strong-weak words; the infants familiarised with weak-strong words did not listen significantly longer to passages containing these words in the testing phase. Moreover, infants tended to link the final syllable of weak-strong words to weakly stressed monosyllabic words, indicating that 7½-month-old infants use strong syllables to signal word boundaries in fluent speech. In comparison, older infants (10½-month-olds) familiarised with weak-strong words listened significantly longer to passages containing these words and responded to the whole weak-strong words in fluent speech not just to the strong syllables of these words, even with misleading boundary cues.
Overall, the picture that emerges from studies on word segmentation, word stress, and native language phonotactics, is that the path to word segmentation is a gradual process that develops across the first year of life. It seems that infants have a perceptual sensitivity to the predominant word stress pattern of the native language that guides them to detect words in fluent speech (even though they may not know the meaning of items they are familiarised with) by the end of the first year of life.

2.6.4 Statistical Learning

Saffran and her colleagues have been interested in whether young language learners use statistical information to discover word boundaries in running speech (Saffran, 2001; Saffran, 2002; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). One consistent cue to when a series of syllables forms a word is that syllables within words usually have higher transitional probabilities than syllables spanning words. For example in the language learners’ experience, the probability that ty will follow ba in the phrase ‘pretty baby’ is much higher than the likelihood that ba will follow ty, because many words other than baby can follow pretty.

Saffran, Aslin et al. (1996) were the first to investigate whether infants could use transitional probabilities to distinguish words in an artificial language. In their study, 8-month-old infants listened to two minutes of continuous speech consisting of four trisyllabic nonsense words (bidaku, padoti, golabu and tupiro) randomly concatenated with no spaces between the words. Infants were then tested to see whether they would discriminate two of the familiarised ‘words’ (e.g., bidaku and golabu) from two new nonwords (e.g., kupado and bubida). Words and nonwords were derived from the same syllable set, so their component syllables were equally familiar and differed only in terms of the transitional probabilities between syllable pairs, for example, .33 for kupado and 1 for bubida. Thus successful discrimination of ‘words’ and ‘nonwords’ would demonstrate infants’ sensitivity to such probabilities. Infants showed differential attention to the familiar (‘word’) and unfamiliar (‘nonword’) syllable patterns.

---

15 Transitional probability is ‘the conditional probability of Y given X calculated by normalising the co-occurrence frequency of X and Y by the frequency of X’ (Gomez & Gerken, 2000). This statistic provides information for word segmentation on the basis of low predictability of a syllable sequence at word boundaries.
suggesting remarkably sophisticated learning abilities for linguistic stimuli and an ability to discover new word forms when there are no word boundary cues.

Further studies have demonstrated that infants are also sensitive to statistical probabilities in sequences of tones (Saffran & Gieperntrog, 2001; Saffran, Johnson, Aslin, & Newport, 1999), suggesting a general rather than speech-specific learning mechanism; the legal ordering of words in sentences (Gomez & Gerken, 1999); grammatical strings in a new vocabulary (Gomez & Gerken, 1999); and native language perceptual categories (Maye, Werker, & Gerken, 2002). In the latter study Maye et al. (2002) familiarised 6- to 8-month-old infants with a continuum of eight stimuli ranging from voiced unaspirated to voiceless unaspirated stops. By manipulating the distributional information in the input they investigated whether differential exposure would impact upon infants’ ability to discriminate voiced from voiceless stops. All infants were presented with at least four exemplars of each stimulus along the continuum. Half of the infants heard more instances of stimuli close to the voiced vs. voiceless category boundary (there is no such boundary in English), and the other half heard instances of stimuli all along the continuum. After only 2.4 minutes of exposure to the stimuli, infants who were familiarised with stimuli at the category boundaries of voiced and voiceless were successfully able to discriminate the series at the boundaries. The other group treated all stimuli in the series as belonging to a single category. This suggests that infants’ acquisition of native language phonetic categories may be moulded by statistical learning principles.

The growing body of evidence concerning infants’ statistical learning abilities for language and other domains of knowledge, highlights infants’ sensitivity to word boundaries (as discussed in sections 2.6.2 and 2.6.3), and their ability to integrate new linguistic input from (artificial) language-learning tasks with their native language knowledge. Hence these studies permit a better understanding of the involvement of learning in language acquisition.

---

16 Pegg and Werker (1997) demonstrated that 6- to 8-month-old infants can discriminate voiced from voiceless unaspirated stops, whereas 10- to 12-month-old infants have greater difficulty discriminating this contrast.
2.7 Infants’ Perception of Segments

Newborns discriminate various consonant contrasts and do so irrespective of whether they are native or nonnative. Between 6 and 9 months of age infants increase attention to detail in stress, word segmentation, and phonotactic patterns of the ambient language environment, but it is also during this time that infants increase attention to segmental properties of their native language. For example, Jusczyk, Friederici, Wessels, Svenkerud and Jusczyk (1993) explored when infants learn about the native language phonetic segments and constraints on their ordering. Infants listened to lists of low frequency abstract words from Dutch and English, which have similar prosodic organisation but vastly different phonetic and phonotactic organisation. At 9 months of age but not at 6 months, American infants preferred to listen to the English word lists; Dutch 9-month-olds listened significantly longer to the Dutch lists. Evidently these preferences are due to increasing familiarity with the sound patterns of the native language. An additional experiment was conducted using low-pass filtered versions of the word lists to investigate whether infants responded in this manner because they had learned about phonetic and phonotactic properties of native language words. When the lists were low-pass filtered, infants showed no preference for the native over the nonnative language. Therefore, 9-month-old infants appear to be sensitive to the phonetic and phonotactic properties of the native language, and they use this information for native language recognition. It is no coincidence that this sensitivity to native language sound patterns occurs during the period in which infants are moving from language-general to language-specific perceptual orientation, as indexed by a decline in sensitivity to phonetic contrasts that are not present in or used contrastively in the native language (Pegg & Werker, 1997; Werker & Lalonde, 1988; Werker & Tees, 1984a).

Infants’ attunement to native language segments is evident from their discrimination of native vs. nonnative contrasts, and is the focus of the following sections. The evidence for (and against) a perceptual reorganisation of consonants and vowels is investigated using the ontogenetic and differential language methods (Burnham & Sekiyama, in press). There has been very little research on infants’ perception of tone but the research that exists will be discussed later in chapter 3.
2.7.1 Development of Consonant Perception

Werker and colleagues were the first to find evidence for perceptual reorganisation as a product of native language experience. In a series of studies using the conditioned head-turn procedure, Werker and colleagues (Werker et al., 1981; Werker & Tees, 1983; Werker & Tees, 1984a) took an ontogenetic approach by testing English language environment infants at several ages, on several nonnative consonant contrasts—the Hindi retroflex /Da/ vs. dental /da/ stop contrast the Hindi voiceless aspirated breathy voiced contrast, /h-dh/ and the Salish (Nthlakampx) glottalised velar vs. uvelar ejective stop contrast /k-q/. Werker and colleagues found that English language environment infants at 6 to 8 months can differentiate these nonnative contrasts (Werker et al., 1981) even though 10- to 12-month-old infants (Werker & Tees, 1984a), 4-year-old children (Werker & Tees, 1983), and English speaking adults (Werker et al., 1981) find discrimination more difficult. The progressive developmental decline between 6 to 8, 8 to 10 and 10 to 12 months in the ability to discriminate the Hindi and Salish contrasts was also was observed in a longitudinal design (Werker & Tees, 1984a). In contrast to this pattern, the decline was not observed when three Salish-learning and three Hindi-learning 11- to 12-month-old infants (differential experience method) for whom the test contrasts are native, were tested for their discrimination, thus explicating the language-specific nature of the perceptual reorganisation (Werker & Tees, 1984a). The robust character of the pattern of decline across the first year of life has also been shown using a habituation-dishabituation procedure for English infants tested on the Salish contrast (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995), and for Japanese infants tested on the English /r-l/ distinction, a contrast which is especially difficult for Japanese speaking adults to discriminate (Tsushima et al., 1994).

Best (1995a) and Best et al. (1990) have also reported poor discrimination by 10- to 12-month-old English-learning infants for Zulu consonant contrasts, including a lateral fricative voicing contrast, a velar voiceless aspirated vs. ejective stop contrast, and a plosive vs. explosive bilabial stop contrast whereas English adults’ discrimination of the fricative voicing distinction was very good indicating that infants’ ability to perceive nonnative contrasts may decline across the first year of life, even when language-specific speech perception is not evident in adults. In contrast, several studies
have reported there to be no attenuation in the ability to perceive certain nonnative consonant contrasts in the first year of life. Best et al. (1988) employed a visual habituation procedure to test English-speaking adults, and English-learning infants between 6 and 14 months of age, and showed that all age groups could successfully discriminate the nonnative Zulu click (apical/lateral) consonant contrasts. Best and colleagues suggest that the Zulu clicks were not discriminated in a language-specific manner because they sound unlike any English phonological category, and cannot be assimilated within the English phonological space, (see Best’s Perceptual Assimilation Model, 2.9.1.2), and in fact were not even reported to be perceived as speech by the English-speaking adults. Presumably, the same is happening with adults’ discrimination of the fricative voicing distinction in Best (1995a). Similarly, in a study with 10-to 12-month-old English language environment infants and English speaking adults (ontogenetic method), it was found that both infants and adults were able to discriminate the Ethiopian ejective /p̩t̪/ contrast in which each phone of the contrast could be assimilated to two different native language categories (TC) (Best, 1991; Best et al., 1990 and see section 2.8.1.2)

Recently, Polka, Colantonio, and Sundara (2001) have found a new developmental pattern for cross-language consonant perception by including a control group of native listeners. They tested differentiation of the English stop-fricative contrast /d-ð/ (used contrastively in English but not French) by English- and Canadian French-speaking adults and 6- to 8- and 10- to 12-month-old English- and French-learning infants using the CHT. On the basis of the results obtained by Werker and colleagues (for a review see Werker & Tees, 1999), Polka et al. (2001) anticipated poorer discrimination of the contrast by French adults and French 10- to 12-month-old infants compared to their English counterparts, but no difference between French- and English- learning 6- to 8-month-olds and English language infants and adults. As expected, the French adults were consistently less accurate at discriminating the contrast than the English adults. Again as expected, the 6- to 8-month-olds showed no effect of language experience, with both the English- and French-learning infants successfully able to discriminate the contrast. However contrary to expectations, there was no decline in perceptual ability for the contrast between 6 to 8 and 10 to 12 months for the French-learning infants, either from analyses of the proportion of infants reaching a preset performance criterion
or in A-prime scores. Moreover, there were divergent developmental patterns observed in English and French subjects. For the French listeners, there were no significant changes in discrimination of the nonnative contrast as a function of increasing age. However for the native English listeners, perceptual discrimination of the native /d-ð/ contrast improved between 10 and 12 months of age and adulthood, thus suggesting that language experience has a facilitative effect on the perception of this contrast in that there is a significant age-related increase in differentiation of the contrast in the English but not the French language groups. These findings challenge the long established view that language experience serves to only prevent a developmental decline in perceptual discrimination of some contrasts, that is, that perceptual capabilities are maintained by appropriate experience. Rather these data suggest that language experience also serves a facilitative role in phonetic perception after 12 months of age. English language adults were better at /d-ð/ than their 10- to 12-month-old compatriots and also better than French adults, while for French language subjects there were no differences between 6- to 8- and 10- to 12-month-old infants and adults.

The experimental design of Polka et al. (2001) incorporates both the ontogenetic and differential experience methods, and as such provides a comprehensive picture of perceptual development for this contrast.

2.7.1.1 Development of Consonant Perception and Age-related Advances in Cognitive Abilities

Infants’ attenuation of phonetic sensitivity occurs not only for nonnative speech contrasts. It appears that the cognitive demands of an additional task can impair infants’ ability to make fine phonetic discriminations of native language syllable pairs. In this section the interaction of perceptual reorganisation and increasing cognitive sophistication are considered.

Stager and Werker (1997) provide evidence that 14-month-old infants ignore the place of articulation difference between [bi] and [di] when learning new object-word pairings, even though they can perceive this distinction in a standard minimal pair phonetic discrimination task (Stager & Werker, 1997; Werker & Stager, 2000), and
even though they can perform the word learning task when phonoetically more distinct items (e.g., 'lit' and 'neem') are used (Werker et al., 1998). Subsequent research has shown that at 14 months, infants also fail to discriminate phonetic distinctions in word learning tasks when the phonetic distinction is embedded in CVC word forms, voicing contrasts, or is a place of articulation and a voicing distinction difference (Pater, Stager, & Werker, in press). One possible explanation for this is that word learning adds a cognitive load that interferes with infants' attention to phonetic detail (Stager & Werker, 1997). However, by 17 months and in concert with increases in vocabulary size, infants can associate all phonetic differences with a change in word meaning (Werker et al., 2002).

2.7.2 Development of Vowel Perception

Kuhl (1991) has shown that there are language-specific influences on the internal structure of vowel categories by 6 months of age in the form of a perceptual magnet effect. These results have given rise to Native Language Magnet Theory (see section 2.8.1.3). Kuhl (1991) found that discrimination by 6-month-old infants is superior when a poor exemplar of a vowel prototype is used as a background stimulus and a good exemplar of the vowel category is the target, than when the reverse is used. Kuhl suggests that this pattern of discrimination occurs because the more prototypical vowel category member acts as a magnet, attracting other more peripheral members toward them and rendering them less discriminable (Grieser & Kuhl, 1989; Kuhl, 1991). This magnet effect appears to be a product of language experience. Kuhl et al. (1992) tested Swedish and American infants with variations on two vowel prototypes, the American English /i/ as in 'peep' and the Swedish /y/ as in 'fye'. The Swedish-learning 6-month-olds showed the magnet effect for the Swedish but not the English vowels, and the English 6-month-old infants demonstrated a magnet effect for English but not the Swedish vowels.

A decline in perceptual discrimination in the first year has also been shown in English-learning infants for several nonnative vowel contrasts, including the Norwegian /y-u/ (Best et al., 1997) and the German /u-y/ and /u-Y/ (Polka & Werker, 1994) contrasts. In the latter study, Polka and Werker (1994) tested 6- to 8- and 10- to 12-month-old
English-learning infants with a change from /u/ to /y/ (or from /y/ to /u/) or a change from /u/ to /Y/ (or a change from /Y/ to /u/). Polka and Werker found an asymmetry in vowel perception: Both 6- to 8- and 10- to 12-month-olds performed better when the direction of change was from /y/ to /u/ rather than /u/ to /y/; and /Y/ to /u/ rather than from /u/ to Y/. These perceptual asymmetries were in accord with previously observed infant and adult within-category discrimination (Kuhl, 1991; Kuhl et al., 1992), in that the more native-like vowel operated as the perceptual magnet.

Despite these results, it appears that infant vowel perception is not only influenced by the native language environment. Polka and Bohn (1996) tested English- and German-learning infants and English- and German-speaking adults for their discrimination of the English /æ-ɛ/ contrast, which is not phonemic in German, and the German /u-y/ contrast, which is not phonemic in English. The infants of both language backgrounds showed good discrimination of both contrasts, and adult perception of nonnative contrasts was also very good. The results show that the absence of experience with a language does not automatically lead to a complete loss of perceptual capabilities for sounds in those languages; some nonnative contrasts can still be discriminated. Furthermore, Polka and Bohn (1996) found consistent perceptual asymmetries for vowel perception irrespective of language background. Both English and German infants at both ages showed identical asymmetries: for the German contrast the /y/ to /u/ change was discriminated better than the /u/ to /y/ change, and for the English contrast, the æ-ɛ change was discriminated better than the ɛ-æ change. The finding that the /æ/ vowel acted as a reference vowel for both English- and German-learning infants is contrary to a perceptual magnet/native language familiarity view. Rather there appears to be a language-universal perceptual bias. Polka and Bohn propose that infants respond differently to vowels that occupy different positions in the articulatory/acoustic (F1-F2) vowel space (see vowel triangle in section 1.2.2.2). Specifically, their plot of F1-F2 frequencies for contrasts showing asymmetries in vowel discrimination shown in Figure 2.1, points to the more peripheral vowel being the perceptual anchor or referent, with contrasts easier to discriminate if the vowel changes in the direction of a more central to a more peripheral or corner vowel.
Figure 2.1. Plot of F1/F2 frequencies showing asymmetries in infants’ vowel discrimination. Arrows point to the reference vowel of the contrast. Vowel changes from peripheral to reference were easier to discriminate than vowel changes from reference to peripheral. Figure originally published in Polka & Bohn, 1996. Permission to reproduce this figure was obtained from the first author.

To test this peripherality hypothesis Bohn and Polka (2001) examined the role of target spectral, dynamic spectral, and duration cues, in German language environment infants’ discrimination of three German language vowel contrasts, /i/-/e/, /e/-/ɛ/, and /o/-/u/ in a ‘/d/ + vowel + /t/’ carrier syllable using the conditioned head-turn procedure. Bohn and Polka tested infants until 20 were successful in discriminating the full cue condition, 10 in each direction of vowel change. In support of the peripherality hypothesis significantly fewer infants had to be tested to reach the pre-set performance criterion when the vowel change was in the direction of a less to a more peripheral vowel, than when the direction of change was from a more to a less peripheral vowel. Polka and Bohn (2003) suggest that the perceptual salience and stability of peripheral vowels may facilitate the formation of language-specific vowel categories in three ways. Firstly, language specific vowel perception may emerge developmentally from the periphery of vowel space to the centre. Secondly, it may serve to guide infants’ production of vowels; infants first produce central vowels and vowel production expands
Speech perception during the 1st year

developmentally from centre to periphery (see Oller, 2000). Thirdly, vowel asymmetries may assist infants perceptually to distinguish function and content words, the former containing more centralised vowels and the latter containing more peripheral vowels.

The results of Polka and Werker (1994) and Bohn and Polka (2001) point to two hypotheses for nonnative vowel discrimination, native language familiarity, and perceptual anchor. It appears that both the ambient language environment and phonetic variables play important roles in vowel discrimination.

2.7.3 From Language-general to Language-specific Speech Perception

In this section a range of studies using the ontogenetic method, the differential experience method, and some using both, have been reviewed. The amalgamated evidence suggests that infants’ initial auditory preferences are reorganised into native language phonemic categories over the first year of life. From birth to 6 months infants are well prepared to perceive any language-general contrasts from any phonetic categories. From 6 to 12 months of age the native language increasingly directs infant speech perception - there is a shift from broad discrimination abilities to more adult-like language-specific perceptual ability. With this, comes an attenuation of sensitivity to nonnative contrasts\(^{17}\). One exception to this pattern is the development of salient vowel categories. Kuhl (1991) has provided evidence that the developmental change for vowels occurs much earlier – by 6 months of age.

The evidence reviewed thus far in this chapter, supports the view that suprasegments and segments are similarly affected by the native language. Infants’ sensitivity to infant-directed speech, native language prosody, and intonation patterns, seems to precede and augment a rapid reorganisation of language-specific knowledge in the second half of the first year of life with regard to phonotactic patterns, rhythm, permissible word boundaries, and phonetic contrasts.

\(^{17}\) A developmental decline in infants’ discrimination of nonnative contrasts is consistently reported if adults’ have a similar difficulty with the nonnative contrast. The results of Polka et al. (2001) is an exception.
2.8 Theories and Models of Speech Perception

It has been shown that language-general sensitivities in newborns undergo a developmental change and their linguistic abilities become more language specific by the end of the first year of life. Werker (1994) identifies three distinct periods in cross-language speech perception research that illustrate how key ideas about this developmental change in infancy have progressed.

2.8.1 The Three Periods of Infant Speech Perception Research

The first period of infant speech perception research suggested that the perceptual capacity for discriminating phonetic contrasts is ‘lost’ around 10 to 12 months of age due to a lack of experience with nonnative speech sounds. This interpretation was modified by two findings. The first was that older infants (Best et al., 1988) and children (Burnham et al., 1991) continue to discriminate some nonnative contrasts, despite a lack of experience with these contrast. The second was that infants have a decreased ability to discriminate allophonic contrasts that are apparent in the ambient language environment but do not have functional phonemic status (Pegg & Werker, 1997).

In the second period of research, it became apparent that the ‘loss’ account was inaccurate for two reasons. Firstly, infants showed attenuated perception of some nonnative contrasts, whereas adults did not (Best, 1995a; Best et al., 1990; Polka & Bohn, 1996) and secondly, it became apparent that adults could discriminate difficult nonnative contrasts when tested with sufficiently sensitive procedures (Pisoni, Aslin, Perey, & Hennessy, 1982), at ISIs of 500ms (Werker & Logan, 1985; Werker & Tees, 1984b) and/or if given extensive training (Tees & Werker, 1984). Thus it became clear that there was no sensorineural loss associated with discrimination of nonnative speech contrasts.

As an alternative, Werker and Tees (1984a) preferred to characterise the attenuation of nonnative speech perception as perceptual reorganisation of perceptual biases, rather than as an absolute loss. Several theoretical frameworks have been proposed to account for the initial sensitivities shown in early infancy and the resultant age-related changes
in speech processing which occur during the first year of life. These theoretically based predictions mark the third wave in infant speech perception research. For example, Burnham (1986) hypothesised an early (6-12 months) decline in sensitivity to nonnative contrasts that are not psychoacoustically salient and are rare across the world's languages (fragile contrasts). In contrast, the decrease in sensitivity to more robust contrasts (prevalent, more psychoacoustically salient contrasts) is reasoned to occur later in development even though they serve no function in the native language phonological system. Best et al. (1988) take a different stand. They argue that phonological status alone determines the likelihood that a contrast is discriminable to a nonnative listener. For example, nonnative contrasts that are highly similar to a native language phoneme are the most difficult to distinguish from that phoneme. At the other extreme are nonassimilable contrasts, including phones that do not show any phonological resemblance to contrasts from the native language, for example Zulu clicks. Nonassimilable phones are predicted to be the most easily discriminable from native language phonemes.

Current theorists of speech perception development take as their starting point the notion of 'perceptual reorganisation' and attenuation of nonnative contrast discrimination rather than total loss. According to Pisoni, Lively, and Logan (1994) and Vihman (1996), theories of speech perception should both characterise the perceptual abilities of mature listeners, and provide some principled account of genetic-environment interaction. That is, theories should explain how speech perception abilities develop and how the linguistic environment selectively shapes them. The theories outlined here deal with how the ambient language environment shapes developmental changes in speech perception, but differ in their explanation of the transition from processing language-general to language-specific speech sounds.

2.8.1.1 Universal, Attunement, and Perceptual Learning Theories

Aslin and Pisoni (1980) proposed a number of models covering the way in which postnatal experience might interact with genetic factors in the development of speech perception: Universal theory, Attunement theory, Perceptual Learning theory, and Maturational theory. Aslin and Pisoni's models (1980) shown below in Figure 2.2 are
a heuristic framework in which processes of development (attunement, loss, attenuation etc.) can be considered.

![Diagram showing relative discriminability across prenatal and postnatal ages with phases of maintenance, loss, facilitation, and perception learning theory](image)

*Figure 2.2. Models of perceptual development: Universal Theory, Attunement Theory and Perceptual Learning Theory, originally published in Aslin and Pisoni (1980). Permission to reproduce this figure was obtained from the first author.*

**Universal Theory** assumes that infants come fully equipped with broadly tuned perceptual and attentional mechanisms, allowing them to discriminate all possible phonetic contrasts that might appear in any natural language. The theory claims that experience is the key to maintaining the ability to discriminate phonetically relevant distinctions belonging to the language-learning environment. In the absence of early experience, contrasts that are not phonetically distinctive are subject to ‘loss’ (or attenuation). This loss is either neural, attentional, or a combination of both. Universal theory makes a number of predictions concerning the reacquisition of ‘lost’ abilities. For example, a lack of exposure in development to some contrasts may produce ‘loss’ or attenuation that cannot be overcome by later training, thus suggesting a critical period for speech perception development (Pisoni et al., 1994). Indeed, as to whether attenuation occurs with age for phonemically irrelevant VOT contrasts, Pisoni et al. (1982) suggest that the basic auditory ability to detect differences in phonemically
irrelevant VOT contrasts is not ‘lost’ as a result of phonological deprivation and can be ‘reacquired’ following minimal training.

Second, according to the *Attunement Theory* position, all infants are capable of discriminating at least some phonetic contrasts in the world’s languages but this perceptual ability may only be partially developed at birth. These ‘partially’ developed discriminatory abilities will only become fully developed if early experience or exposure to the appropriate speech sounds aligns or sharpens them. Attunement theory states that phonetically relevant contrasts in the native-language environment become finely tuned with experience, while phonetically irrelevant contrasts remain broadly tuned, or become attenuated in the absence of early experience.

In contrast, *Perceptual Learning Theory* posits that a perceptual ability may be absent at birth and that specific early experience with phonetically relevant contrasts in the ambient language environment leads to discrimination abilities. The rate of development is dependent on the psychoacoustic distinctiveness of the contrast compared to other phonetic contrasts, infant attention, and the importance of the contrast in the early perceptual environment. The discrimination of phonetically irrelevant contrasts will never be superior to phonetically relevant contrasts because specific early experience with the contrast is a prerequisite for discrimination of phonetically irrelevant contrasts.

Finally, *Maturational Theory* holds that discriminatory abilities unfold developmentally according to a genetically determined maturational schedule. In line with this theory, all possible phonetic contrasts are initially discriminated equally. The difference is that the initial ability to discriminate any given contrast may advance, decline, or be maintained at different ages in development, due to maturational factors and irrespective of early experience. Pisoni et al. (1994), suggest that a combination of these theories probably provides the best description of the development of speech perception as no single theory will account for the development of all phonetic contrasts. Of these theories, Attunement theory has been cited as the most supported position (see Burnham, 1986). Attunement theory states that facilitation occurs for partially developed speech capacities if there is necessary and sufficient experience.
with the contrast in the language-learning environment. Attunement theory also accounts for 'loss' or attenuation of perceptual abilities due to limited exposure with particular contrasts.

2.8.1.2 Perceptual Assimilation Model

Best has proposed the Perceptual Assimilation Model (PAM) to account for variability in discrimination of nonnative contrasts (Best, 1994). This direct-realist view of cross-language speech perception is based on the premise that there is no special-purpose built-in speech processing modules. Rather, the direct-realist perspective assumes that infants pick up those distal articulatory gestures that are used in their native language from the information in the native language environment (Best, 1995a).

The direct realist approach presumes that infants initially perceive only non-linguistic information in speech (Best, 1995a). The infant is able to detect the range of vocal events (simple gestures) produced by native and nonnative speakers, but these gestures initially have no linguistic relevance to them as they are simply part of the universal phonetic domain. The experiential changes in speech perception that occur during infancy come about via increased attention to the specific acoustic information in the wave-form that reflects specific native language properties of vocal production. The developmental shift to a linguistic focus results from perceptual attunement to language-specific, higher-order gestural invariants. Mature perceivers who detect these higher order invariants in native language phonology are actually gathering reduced information from the speech environment than they detected prior to recognizing the patterns in native speech. Importantly, as languages differ in their assembly of simple gestures and gestural constellations, native-language experience draws attention to the selection of simple gestural settings relevant to native language phonology and constrains the perception of nonnative contrasts (Best & Strange, 1992).

Best (1995a) states that the way nonnative phones are discriminated is dependent on how they are perceived in terms of their similarities (and dissimilarities) to native phonemes. Best et al. (1988) outline four patterns by which the two members of a given nonnative contrast could be perceptually assimilated to native phonemes:
The two members of the contrast may be gesturally similar to two different categories in the native phonology, thereby perceptually assimilated to Two Categories (TC type).

1. Both nonnative phones may be assimilated into a single category, equally well (or poorly), in which case they are both similar/discrepant to native exemplars of a Single Category (SC type).

2. Both may be assimilated into a single category but unequally, that is, one exemplar may be more similar to a native phoneme than the other, thus showing a Category Goodness difference (CG type).

3. The nonnative sounds are very different from the gestural properties of any native categories, and are therefore Nonassimilable (NA type).

There are also additional assimilation patterns (Best, 1995a). Firstly, both nonnative sounds may fall within native language phonetic space but outside of a particular native category. These contrasts are both uncategorisable (UU type) and vary in their range of discriminability. Discrimination is dependent on the proximity of the contrasts to one another and to other native phonetic categories. Finally, discrimination is expected to be very good for nonnative contrasts where one member can be assimilated to a native-language category, but the other falls outside the native categories.

As an extension of the PAM, Best (1994) proffers a number of different hypotheses about the nature of perceptual language-specific reorganisation for older infants. The strong phonological hypothesis suggests that the perceptual shift at the end of the first year of life reflects the infant’s stage-like emergence into the adult native phonological system. Both infants and adults show comparable discrimination performance (i.e., excellent discrimination of TC contrasts, very poor discrimination of SC types). The strong phonological hypothesis assumes that the infant can recognise the linguistic function of phonemic contrasts and other phonological rules. The phonemic contrast hypothesis entails the notion that older infants organise perception according to phonemic contrasts. Thus they can clearly discriminate between-category contrasts, but
Speech perception during the 1st year

discriminate within-category contrasts less successfully. According to this view, TC and NA contrasts would be easy to discriminate but CG and SC types would be difficult to discriminate because discrimination relies upon a refined understanding of within-category contrasts. The final hypothesis is the category recognition hypothesis, which states that older infants organise speech perception by recognising whether gestural patterns are similar or different to one another, and focus on the recognition of the patterns of gestural coordination within native language categories. The infant comes to recognise segmental patterns and larger syllable and word units as well. This hypothesis predicts that older infants will have the most difficulty discriminating the unfamiliar gestural patterns of SC contrasts and also TC phones whose gestural patterns deviate from the most similar native phones. Nonassimilable phones such as the Zulu clicks (Best et al., 1988) should be recognised and discriminated as nonspeech sounds by nonnative listeners, because in these contrasts there would be no gestural similarities to any native language phones.

To summarise, the PAM makes a number of predictions about how listeners will assimilate nonnative phones with respect to native-language phonological categories. The model assumes that the perceptual primitives for speech perception are articulatory gestures that are directly picked up by listeners from the environment and suggests three hypotheses for how this might occur in older infants.

2.8.1.3 Native Language Magnet Theory

The Perceptual Magnet Effect (PME) for vowel perception shows that simply being exposed to a particular language results in a distortion of the perceived distance between speech stimuli by altering phonetic perception (Kuhl & Iverson, 1995). The result is that perceptual categories are formed that mirror phonological categories of the ambient language environment.

Kuhl (1991) shows that prototypes, the most representative instances of phonetic categories, function like perceptual magnets for other sounds belonging to that category. Thus, exemplars of a category which are close to the category prototype in phonetic space are drawn toward it, making it difficult to hear the difference between
the prototype and the surrounding stimuli (Kuhl, 1995). Developmental studies have shown that the PME is evident by 6 months of age for native language vowel categories (Kuhl, 1991). Furthermore, the decline in the discrimination of nonnative vowel contrasts occurs much earlier for vowels (4 to 6 months) than has been found for nonnative consonant discrimination (Polka & Werker, 1994).

Recent work by Kuhl (1995) with adult listeners has attempted to identify the perceptual mechanisms underlying the PME. The results show that language experience results in the formation of perceptual maps that specify the distance between speech sounds in acoustic space. These maps of acoustic space are defined in terms of the native language and show a clear distinction between speech sounds of different categories, but show strong internal category cohesion. For example, American listeners tested with the consonants /r/ and /l/ showed perceptual clusters around both the consonant prototypes. Contrastively, native Japanese listeners showed a very different perceptual map for /r/ and /l/, because /r/ and /l/ are not distinctive phonetic categories in the Japanese language.

Kuhl has proposed a three-step theory of speech development called the Native Language Magnet Theory (Kuhl & Iverson, 1995). The theory accounts for infant’s initial sensitivities and the changes that occur during the first year of life brought about by language experience. At birth, infants perceptually partition the sound stream into innately specified language-general categories (categorisation) that do not depend on linguistic experience (Phase 1). The category boundaries are due to basic auditory perceptual mechanisms and not a specific language module. By 6 months, infants no longer show the innate boundaries. Rather, from experiencing consonant and vowel sounds from the ambient language, infants develop language-specific boundaries and begin to exhibit these magnet effects in Phase 2. In Phase 3, nonnative phonetic boundaries are erased and the infants’ perceptual system is reorganised and dominated by language-specific phonetic categories or ‘magnets’.

By positing prototype representations of phoneme categories, Kuhl has attempted to explain why it is particularly difficult for the mature listener to perceive nonnative contrasts. If listeners do actually form prototypes, any nonnative sounds would be
interpreted in terms of the native-language prototype system. Nonnative contrasts with similar acoustic weighting systems would be attracted to the prototype and be identified as a native phoneme (Goodman, Lee, & DeGroot, 1994).

2.8.1.4 Word Recognition and Phonetic Structure Acquisition Model

The Word Recognition and Phonetic Structure Acquisition (WRAPSA) model accounts for how speech perception capacities evolve to support word recognition in continuous speech. It describes the initial state of infant speech perception capacities and how these capacities develop and change as the infant becomes more familiar with the native language (Jusczyk, 1994, 1997).

Processing at the first stage involves a preliminary analysis of the signal. Acoustic speech signals pass through the peripheral auditory system that extracts syllable-sized units from the array of spectral and temporal properties present in the signal. The features extracted at this preliminary stage of auditory analysis reflect the inherent organisation of the human auditory system. The acoustic description that emerges is neutral with respect to the language spoken, and provides a great deal of fine-grained information that can be used to discriminate different sounds from each other. These innate acoustic processes represent the sensory boundaries for the processing and classifying of speech and nonspeech sounds, they equip the infant for speech perception, and make it possible for adults to acquire a second language (Flege, Bohn, & Jang, 1997).

It is possible that a number of descriptive patterns are extracted from acoustic input at any given moment. During early infancy the patterns of extraction are very general and apply to both speech and nonspeech input. These patterns provide the infant with a preliminary categorisation of input. Information about the properties of sounds to which infants are exposed and how these are distributed in the input is critical to how the weighting scheme is shaped. With experience, the infant learns the native-language interpretative schema. Learning the interpretative schema is a matter of becoming sensitive to the features that signal meaningful distinctions in the native language and weighting the description-selection process to provide an accurate decoding of speech.
The resulting weighted representations are matched against scored lexical items. Jusczyk (1994) states that this is the age when there is a decline in sensitivity to some nonnative speech contrasts in language and is the point where an interpretative scheme develops and shifts the infant’s attention away from foreign language contrasts. The capabilities underlying infant’s speech perception allow the infant to learn about the distribution of the sound properties in the native language. Furthermore, Jusczyk argues that in relation to Zulu clicks, infants treat these sounds as nonspeech because they bear no resemblance to any sounds in the English language.

Finally, after a description has been selected it can be compared to stored lexical representations in long-term memory, so that its meaning can be accessed. Initially, representations of words are global. As the infant stores a description of the sound structure of lexical items and the features of the spoken word that distinguish it from other words, a prototype of the word forms in the lexicon. The weighting scheme forms, refines, and reorganises the prototypes so that a ‘matching’ process to the prototype operates in spoken word recognition (Jusczyk, 1997).

In summary, WRAPSA hypothesises that when memory traces for words are stored in the lexicon, the overall properties derived from perceptual representations are preserved.

2.8.1.5 An Epigenetic Model of Speech Perception

Werker and Tees (1999) propose an epigenetic model to account for the initial state of infant speech perception and the subsequent changes with age. In relation to sensitivities in newborns, Werker and Tees acknowledge that genetically inherited mechanisms have evolved to allow experience-expectant brain development. They argue that experience-expectant changes in relation to speech perception, tend to occur relatively early in development and result in relatively stable and permanent neural structures (Werker & Tees, 1992). Experience-expectant interactive changes shape the speech perception biases of the prenatal infant. For example, the mother’s voice appears to be audible to the neonate in the last trimester of gestation (for example,
Moon, 1993) and neural mechanisms becomes organised to respond preferentially to human sounds and to process speech (Werker & Tees, 1992)\(^\text{18}\).

Werker and Tees (1999) assert that perceptual tuning to native-language properties during the first postnatal year involves experiential influences and changing brain structure. Specifically they argue that auditory speech input and self-vocalisations impact on the neural organisation that may play a role in the preferential sensitivity to native-language phonetic information and influence the native language storage database, that is, the acquired information about native language phonetic, prosodic, rhythmic, and phonotactic patterns. Thus, infants use this store of ‘probabilistic’ information to learn about the characteristics of the native language and ignore nonnative characteristics. However in order to coordinate the different sources of information for the goal of spoken word recognition, sensitivity to probabilistic information is necessary, but insufficient. Lalonde and Werker (1995) found that infants manifest the ability to coordinate two or more sources of information by 9 months of age, and that this may influence the decline in sensitivity to nonnative contrasts by the end of the first year of life. Thus the capacity to ‘ignore’ nonnative consonant contrasts seems to be related to emergent cognitive abilities. On the basis of the sudden emergence of coordinative abilities for speech perception and other domains, Werker and Tees (1999) propose that an epigenetic advance in the prefrontal cortex might allow the infant to coordinate information in many domains, including functional reorganisation of speech perception capabilities.

This model successfully accounts for both language-general sensitivities at birth and the changes that occur during infancy. However, unlike the WRAPSA model, this account does not detail the pattern of spoken word recognition and word learning in infancy. The PAUSE:mir model, outlined next, attempts to address these issues.

\(^{18}\) Kuhl and Miller (1978) hold a different view, suggesting that it is languages that have evolved to take advantage of the perceptual sensitivities of the human auditory system, rather than the other way around.
2.8.1.6 PAUSE: mir

This hypothesis states that to be expert language users, infants engage in the computationally intensive task of learning the associations between word forms and objects (Werker, 2003). Given the cognitive demands that the infant must bring to such a task, errors may be made in the details acquired about the object and/or the word form (Stager & Werker, 1997). Initially, word-object pairings are stored as individual units, but gradually the infant’s knowledge generalises to new word clusters in the multi-dimensional representational space in accordance with language-specific phonetic categories. With enough word-object pairings, the words begin to become organised into language-specific categories and higher order regularities become apparent to the infant.

Werker and Curtin (2003) suggest that as the higher-order regularities in the input become strengthened, they come to function as phonemes. These phonemes then steer subsequent word learning, thereby easing the cognitive demands of learning new word-object associations.

2.8.1.7 Summary

It is clear that these models take very different approaches to the development of consonant perception, vowel perception, and word recognition during the first year of life. Although comprehensive in their own right, none of the models outlined here consider the development of lexical tone perception during infancy. Consequently it is unknown whether the patterns of development for consonant and vowel perception equally explain the developmental course of lexical tone perception, or whether a new framework is essential. In other words, which model can best accommodate tone?

The next chapter, chapter 3, concerns lexical tone - what it is, and how tone and non-tone language speakers perceive linguistic tone and non-linguistic pitch. Furthermore the chapter will review what little is known about the development of lexical tone perception. Later, in chapter 4, the experiments on infants’ lexical tone and nonspeech tone (pitch) discrimination are introduced.
CHAPTER 3
Lexical Tone

‘Every language has tones but not every language uses its tones for the same purpose’ (Douglas M. Beach, 1924)
This chapter begins by providing specific detail about lexical tone, its features, and how it is perceived, and then introduces the tone systems of Thai, Cantonese and Mandarin. The focus tone languages in this chapter are Thai, Cantonese and Mandarin (although some others are mentioned) because they each have a different number of tones and are the languages of interest in the experiments. The final sections of the chapter are devoted to a discussion of infants’ ability to perceive pitch variations in speech and music, and the need for research into infants’ discrimination of lexical tone.

3.1 What is Lexical Tone?

According to the purposes for which they are used, tones may be called semantic, syntactic and emotional. Semantic tones may be defined as tones which are used to differentiate fundamental ideas denoted by words and phrases. Syntactical tones are tones which are used to show the relation of one idea to another, either expressed or unexpressed. Emotional tones are tones which show the effect upon the speakers emotions of something which has just been thought or felt ... tone languages have all three kinds of tone: Semantic, syntactical, and emotional tones (Beach, 1924).

When Beach (1924) uses the term semantic tone to distinguish meaning at the word level, he is describing lexical tone. Lexical tone refers to a particular way in which pitch (and associated parameters, as specified below) are utilised in language to carry lexical information. A tone language then, is a language that utilises F0 (perceived as pitch) and associated variations plus the usual segmental variations of consonants and vowels to contrast the meaning of words. Over half the world’s population speak a tone language - the traditional tone languages of Western Africa, (e.g., Yoruba and Sesotho), the Americas, and Asia (e.g., Cantonese, Mandarin, Thai, Vietnamese, Taiwanese, Burmese) as well as the marginal tone or pitch accent languages of Europe (Swedish, Norwegian, Latvian) and Japan19.

---

19 Pitch accents arise from the relationship between adjacent strong and weak syllables in words, whereas lexical tone is carried on the vowel of a syllable McCawley (1978).
Just as a distinction can be made between phones and phonemes (see section 1.1), tones and tonemes can also be distinguished. Tone refers to acoustic and phonetic level variations in pitch, voice quality, duration, vowel length and height (see 1.3.2.2 for a full description of the acoustic features of tone), which can be used to distinguish meaning in the world’s languages. A toneme is a group of tone variations that are produced and perceived as a single tone class. Thus, again analogous to phonemes, there may be allotones (distinct tones) that make up a particular toneme in a particular language. While tones and tonemes in this sense are logically similar to phones and phonemes, in practice the analogy may not strictly hold (see earlier discussion of phones and phonemes 1.1).

A tone language can be distinguished from non-tone languages such as English and French that contain pitch and associated variations due to intonation. As discussed in section 1.3.3.2, intonational variations are determined by the structure of larger syntactic or semantic distinctions at the phrase or sentence level but unlike tones they do not convey linguistic information at the word level (Anderson, 1978; Gandour & Harshman, 1978). Like intonation and stress, tones can be considered to be a suprasegmental feature of speech because they contribute to the surface intonation form and their F0 contours are argued by Xu and Wang (1997) to be intonation primitives. This view is in competition with the commonly held view that tones are influenced by intonation, discussed later in section 3.4.2.1. However tones also serve to contrast lexical meaning and therefore also function like segments (Clumeck, 1980, see also section 3.3 and 3.6.4.1). In support of the segmental function of tone, tone is mainly carried on the vowel of a syllable although some involvement of surrounding consonants may also be implicated.

### 3.1.1 What Lexical Tone is Not

To understand fully what lexical tone is, it is important to know what it is not. Firstly in physical terms, F0 can vary continuously across a wide range of frequencies, depending on a number of features. Neither emotional expressive factors of pitch that are unrelated to other linguistic aspects of an utterance, nor features of individual voice range form part of linguistic descriptions of tone, despite the fact that F0 plays an
important role in these tone variations. Secondly, although tone and intonation are suprasegmental sisters that are related because they have pitch variations as their primary property, lexical tone is not intonation. Intonation refers to the pitch variations determined by the structure of larger syntactic units and is carried usually over polysyllabic utterances, whereas tone refers to pitch variations determined by the meanings of individual words and is carried on a single syllable.

3.2 Tonogenesis

The term *tonogenesis* refers to the development of tone within tone languages. For example, a language originally with two tones, may split and become a four tone language. While tonogenesis does not directly refer to the origin of tone as a linguistic tool, discussing tonogenesis may shed some light on the origins of tone languages (Henderson, 1981).

The best-documented evidence for tonogenesis is from the development of contrastive tones on vowels due to a loss of voicing distinction on pre-vocalic obstruents (corresponding to around 600 A.D). This is the case in Chinese dialects, where there has been a merger between voiced and voiceless consonants. For example in the Canton dialect of Chinese (also known as Cantonese) the voiced/voiceless pairs /p:/b/, /t:/d/ and /k:/g/ merged, and were replaced by phonologically distinctive pitch differences, resulting in a two-way split in the tone system and thus evolution from three phonemic tones to six (Henderson, 1981). When such a development as this occurs, as it has in many other South East Asian and African tone languages, a relatively lower tone is substituted on vowels following voicing of the preceding consonant, and a relatively higher tone after a previously voiceless or voiceless aspirated consonant. The apparent synchronic correlation between consonant type and F0 (Haudricourt, 1972; Hombert, Ohala, & Ewan, 1979) has at least indirect support from acoustic phonetic research of the non-tone language English. Both House and Fairbanks (1953) and Lehiste and Peterson (1961) demonstrated that the F0 of

---

20 Some tone languages such as Mak and Tung (members of the Tung-Sui group of languages) have undergone a three-way split in the tone system through the three-way merger of glottalised, aspirated and voiced sonorants. The three-way split in the tone system is beyond the scope of this section. Interested readers should refer to Haudricourt (1972).
Lexical Tone

Phonation immediately following the release of a voiceless consonant is higher than the F0 after a voiced consonant. The phonetic explanation for this intrinsic pitch effect is that the air flowing through the open glottis for the voiceless consonant briefly perturbs the vibration rate of the vocal folds upward, while for voiced consonants the air flow is impeded and the closed position of the glottis may provide sufficient force to keep the vibration rate at a steady level, thus allowing a drop in F0 during and after the production of the consonant closure or constriction (Abramson, 1975b).

In a more focused study of intrinsic pitch in a tone language, Gandour (1974) showed that in Thai, the perturbation due to prevocalic consonants is present but minimal (30ms of a voiceless consonant and 50ms of a voiced consonant were affected by the preceding consonant). The functional significance of this is that the tones remain maximally distinct. With respect to tonogenesis, Abramson (1975b)21 suggests that in the early development of Thai, adjustments of F0 due to intrinsic pitch were heard as pitch perturbations that were gradually enhanced in speech until they attained phonemic status. The specificity of this effect is shown by the fact that consonant voicing affects the pitch of the following vowel, whereas there is little evidence of consonant voicing affecting the pitch of a preceding vowel, or of tone affecting consonants.

There are also strong indications that laryngeal features such as aspiration, glottalisation, and breathy or creaky voice quality have played a role in tone splitting. For example in Thai, what are assumed to have once been voiced stops have lost their voicing and become aspirated, giving rise to a new set of tones (Henderson, 1981). Specifically, the voiced consonants of Thai lower subsequent F0 more than do voiced obstruents. The seeds of this phenomenon can be found in Hindi, a non-tone language, in which the onset F0 of a vowel is markedly lower following a breathy voiced consonant than it is after any other consonant type. There is also evidence that implosives produce tone lowering less than do voiced stops, and some linguists believe that implosives are actually tone raisers because the rapid lowering of the larynx during the articulation of implosives can generate a high rate of glottal air flow that has the potential to raise the F0 above its average level Hombert et al. (1979).

---

21 Interestingly, there is still no conclusive evidence as to whether Thai in its original form (i.e., Proto Tai), was or was not a tone language, although most linguists believe that there were at least two tones Henderson (1981).
The evidence suggests that the two-way splitting of tone languages originated from the effect of prevocalic stop consonants, and to a lesser extent from voice quality and vowel height. Regardless of changes that have occurred to tone languages across centuries, an important point is made by Henderson (1981) - that following the split of a tone system, tones reorganise themselves in tone space such as to maintain maximal perceptual distance between each other. In the next section the tones of Thai, Cantonese and Mandarin are described in terms of their F0 and the importance of these tone languages to the current study is outlined.

3.3 Tone Space in Thai, Cantonese and Mandarin

In this thesis, infants from families who speak a tone language, either Cantonese or Mandarin, and infants from families who speak a non-tone language, English, are tested on a lexical tone and a nonspeech tone discrimination task. The stimuli for the lexical tone discrimination task are Thai tone contrasts, and the nonspeech stimuli are synthetic analogues of these. As Thai tones and Cantonese/Mandarin language-learning infants are involved in this study, the tone structure of Thai, Cantonese, and Mandarin are described below in sections 3.3.1 and 3.3.2 respectively. Tone space is defined by Abramson (1986) as ‘the set of articulatory and auditory dimensions by which the speaker is constrained in production and perception’ (p. 105). Tone space is thought to be analogous to that of vowel space (see section 1.2.2 2 for a description of vowel space, i.e., vowel quadrilaterals).

3.3.1 The Nature of Thai Tones

Standard Thai, also known as Central Thai or Siamese, is the official national language of Thailand and the dialect of the Central Region including the capital of Thailand, Bangkok. Standard Thai is spoken in educational, commercial, and political settings, despite other dialects of Thai being more widely spoken in other regions of the country (Abramson, 1962).

Standard Thai has five tones, conventionally labelled as high, falling, mid, rising and low and every syllable of lexical items in Thai is characterised by one of these tones as
well as by consonant and vowel features (Abramson, 1975c). An example of the five-way lexical tone distinction in Thai is shown in Table 3.1. The fundamental frequency contours of the Thai tones over time on the syllable [kha] are shown in Figure 3.1 (to hear these tones, refer to Appendix CD1 on the CD supplement accompanying this thesis). The mid, high and low tones have relatively stable pitch contours and are known as level or static tones. Of the static tones, the mid and the low each have a narrow range continuous fall and are difficult to distinguish from one another, while the high tone has a narrow range continuous rise with a slight fall at the end (Erickson, 1974; Luksaneeyanawin, 1984). The rising and falling tones are known as contour or dynamic tones because they involve more rapid changes in fundamental frequency movement and audibly rise and fall (Abramson, 1978b). For the dynamic tones, the fall is a wide range continuous fall and the rise is a wide range continuous rise (Luksaneeyanawin, 1984).

Table 3.1

*The Five Lexical Tones of Standard Thai and an Example of a Five-way Lexical Contrast on the Syllable /kʰə/.*

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Lexical item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid</td>
<td>kʰā</td>
<td>‘To be stuck’</td>
</tr>
<tr>
<td>Low</td>
<td>kʰà</td>
<td>‘Galangal’</td>
</tr>
<tr>
<td>Falling</td>
<td>kʰâ</td>
<td>‘To kill’</td>
</tr>
<tr>
<td>High</td>
<td>kʰǎ</td>
<td>‘To trade’</td>
</tr>
<tr>
<td>Rising</td>
<td>kʰǎ</td>
<td>‘Leg’</td>
</tr>
</tbody>
</table>
Figure 3.1. Time-normalised fundamental frequency contours of the five Thai tones (spoken by a male speaker of Thai).

3.3.2 The Nature of Chinese Tone: Cantonese and Mandarin

Cantonese is a tone language belonging to the Yue family of dialects distributed across Southeast China. It is the official language of Hong Kong, China, with over 90% of Hong Kong’s population being native Cantonese speakers. Cantonese is the medium of instruction in Hong Kong schools, and is used widely in television and radio news broadcasts, movies, novels, theatre, popular culture and mass advertising. Cantonese is spoken along with Mandarin in Guangzhou\(^{22}\), China, the capital of the southern coastal province of Guangdong, China, and in the neighbouring cities of Nanning and Wushou of Gungxi, China. Cantonese is also spoken in Macau\(^{23}\), formerly a Portuguese colony that now belongs to China. The total population of speakers of Cantonese and other Yue dialects in the world is estimated to exceed 40 million.

In mainland China however, Standard Mandarin or Putonghua is unchallenged as the official language and even in Guangzhou, Mandarin is taught in schools. The Chinese

\(^{22}\) Guangzhou was formerly Canton.

\(^{23}\) Macau is sometimes written as Macao.
government authorities discourage people from speaking Cantonese, however Cantonese remains the predominant language of business in southeast China, as Guangdong has built close economic, cultural and linguistic ties with Hong Kong.

It is important to note that Chinese dialects do not function as dialects do in English; rather they are mutually unintelligible and are as different as Italian and French. In contrast, for example, Australian and American English speakers can communicate easily with one another. Thus some linguists refer to the Chinese dialects as Chinese languages. In terms of written Chinese, Cantonese and Mandarin share the same ancient logographic system of writing\(^{24}\), but the reader can understand the characters as either Cantonese or Mandarin.

Cantonese (Hong Kong dialect) has at least six tones\(^{25}\), and Mandarin has four tones, (Bauer & Bendedict, 1997). These tone systems are discussed in the following section.

3.3.2.1 Description of Cantonese Tones

There is some disagreement over the number of tones in Cantonese. So (1996) states that there are a total of nine tones in Hong Kong Cantonese, six contrastive tones with three allotones of the three level tones, High Level tone (HL), the Mid Level tone (ML), and the Low Level tone (LL), that mainly differ from each other in overall pitch height. The three contour tones are the High Rise tone (HR), the Low Rise tone (LR) and the Low Fall tone (LF). The HR and LR tones have rising contours but rise to different levels. The LF tone is a falling tone with a falling contour. The three allotones or entering tones, are not contrastive tones: The High Entering (H) tone is an allotone of HL, Mid Entering (M) an allotone of ML and Low Entering (L) is an allotone of LL. These allotones have heights similar to their respective level tones and differ only in tone length (So, 1996).

\(^{24}\) Note that there are slight differences in written Cantonese and Mandarin, but a discussion of these discrepancies is not relevant to the present thesis. Interested readers should refer to Bauer and Bendedict (1997).

\(^{25}\) The issue of how many tones are present in Cantonese is discussed in 3.3.2.1.
In another view on this matter Bauer and Benedict (1997) claim that the Cantonese tone system includes six basic tones: Three level tones, High Level, Mid Level and Mid-Low level\textsuperscript{26}, and three contour tones, High Rising, Mid-Low Rising\textsuperscript{27}, Mid-Low Falling\textsuperscript{28}, with some speakers having an additional contrastive tone or a variant of one of the basic tones, High Falling. Nevertheless, both Bauer and Benedict (1997) and So (1996) agree that two variants of the high basic tone, High Level and High Falling, are found in Guangzhou Cantonese whereas most Hong Kong speakers now lack the High Falling tone, or do not distinguish between the High Level and High Falling tones in the same way as do Guangzhou speakers. So (1996) posits that the changes in the Hong Kong tone system are due to the separation of Hong Kong (under British rule) from Guangzhou (China) for the majority of the latter part of the 20\textsuperscript{th} Century. As a result there were distinct political, social and linguistic influences affecting the Cantonese spoken in each city, and accordingly, the characteristics of Hong Kong Cantonese and Guangzhou Cantonese have diverged. With the handover of Hong Kong to China in 1997, the political, social and linguistic influences have changed dramatically and it will be interesting to observe whether the tone systems of Guangzhou Cantonese and Hong Kong Cantonese become more similar as a result of the merger.

Bauer and Benedict (1997) give an acoustic description of the Cantonese tone contours and depict each tone contour with tone values of their onset and ending points from 1 to 5\textsuperscript{29}. These tone values are known as Chao tones. Since there is consensus between Bauer and Benedict (1997) and So (1996) of at least six contrastive tones in Hong Kong Cantonese, only these six will be described here.

The High Level tone (55 in Chao tone values) starts at a high register and remains reasonably level with a possible slight fall toward the end. The Mid Level tone (33)

\textsuperscript{26} This is equivalent to the Low Level tone described by So (1996), the name difference arising from different descriptions of the movement of the Cantonese tones.

\textsuperscript{27} This is equivalent to the Low Rise tone described by So (1996), the name difference arising from different descriptions of the movement of the Cantonese tones.

\textsuperscript{28} This is equivalent to the Low Fall tone described by So (1996), the name difference arising from different descriptions of the movement of the Cantonese tones.

\textsuperscript{29} The Chao tone values depict tones in terms of the speaker’s pitch range and this system is recognised by the International Phonetic Association to represent graphically the tone contours of any language: 5 (High), 4 (Mid-High), 3 (Mid), 2 (Mid-Low), 1 (Low).
occupies the mid register and remains reasonably level with a possible slight fall toward the end, 3. The *Mid-Low Level* tone (22) is spoken at a low register. It begins at the mid-low point of 2 and drops slightly at the end. The *High Rising* tone contour (25) rises from mid-low to high or begins by falling from mid to mid-low and then rising to high. The *Mid-Low Rising* tone contour (23) starts at approximately the same place as the High Rising tone but only rises to a mid level. The *Mid-Low Falling* tone contour (21) starts at a mid-low register and falls to the lowest, 1. This tone falls to a register lower than the Mid-Low Level tone. Table 3.2 illustrates the possible six-way tone contrast that occurs in Cantonese. Figure 3.2 presents the fundamental frequency contours of the Cantonese lexical tones carried on the syllable /fan/, spoken by a female speaker of Cantonese (to hear these Cantonese tones refer to Appendix CD2, CD supplement).

Table 3.2

*The Six Lexical Tones of Cantonese and an Example of a Six-way Lexical Contrast on the Syllable /fan/. Numbers refer to the Chao values for that tone.*

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Lexical item</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>fan55</td>
<td>‘Divide’</td>
</tr>
<tr>
<td>Low-mid/high-rising</td>
<td>fan25</td>
<td>‘Powder’</td>
</tr>
<tr>
<td>Mid</td>
<td>fan33</td>
<td>‘Lecture’</td>
</tr>
<tr>
<td>Low-mid/low-falling</td>
<td>fan21</td>
<td>‘Grave’</td>
</tr>
<tr>
<td>Low-mid/mid-rising</td>
<td>fan23</td>
<td>‘Angry’</td>
</tr>
</tbody>
</table>
Figure 3.2. Time-normalised fundamental frequency contours of the six Cantonese tones (spoken by a female speaker of Cantonese).

Descriptions of the Cantonese tones customarily divide them into two categories, according to the kind of syllable on which they occur (Bauer & Bendedict, 1997). It is generally the case that the six contrastive tones may occur on all open syllables and on closed syllables which end with the nasal consonants /m, n, ŋ/ or one of the semivowels /w, j, y/. On the other hand, syllables carrying the entering tones are closed with final stop consonants /p, t, k/ (Bauer & Bendedict, 1997; So, 1996).

3.3.2.2 Description of Mandarin Tones

There are four tones in Mandarin Chinese, high-level, mid-rising, mid-falling-rising, and high-falling characterised by their F0 contours. The High-Level tone (tone 1, 55)\(^3\) has a flat F0 pattern, Mid-Rising (tone 2, 35) has a rising F0 pattern, Mid-Falling-Rising (tone 3, 214) has a falling-rising pattern, and High-Falling (tone 4, 51) has a falling F0 pattern. As an example of the maximum four-way lexical reference of tones in Mandarin, the same syllable /ma/ can mean different lexical items for each of the tones.

---

\(^3\) Tone number and Chao values stated in brackets.
as depicted in Table 3.3 below. The fundamental frequency contours of these 4 tones over time are shown in Figure 3.3 (to hear the tones refer to Appendix CD3 on the CD supplement). There is also a ‘fifth tone’, known as the inherent neutral tone, whose pitch value varies dependent on its preceding full tone. Thus the ‘fifth tone’ occurs as a result of tone sandhi, a phenomenon that occurs in Mandarin (as in Cantonese) and is outlined below.

Table 3.3

*The Four Lexical Tones of Mandarin and an Example of a Four-way Lexical Contrast on the Syllable /ma/. Numbers refer to the Chao values for that tone.*

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Lexical item</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-level</td>
<td>ma55</td>
<td>‘Mother’</td>
</tr>
<tr>
<td>Mid-rising</td>
<td>ma35</td>
<td>‘Hemp’</td>
</tr>
<tr>
<td>Mid-falling-rising</td>
<td>ma214</td>
<td>‘Horse’</td>
</tr>
<tr>
<td>High-falling</td>
<td>ma51</td>
<td>‘Reproach’</td>
</tr>
</tbody>
</table>
Figure 3.3. Time-normalised fundamental frequency contours of the four Mandarin tones (spoken by a female speaker of Mandarin).

3.3.3 Tone Sandhi

As part of their tone system, Cantonese and Mandarin exhibit a distinctive pattern of tone sandhi. For example tone sandhi is noted to be responsible for changing tone 1 in Cantonese from a High Level tone to a High Falling tone in connected speech, possibly due to intonation effects. Thus among speakers who distinguish between the High Falling and High Level tones, in a High Falling-High Falling-High Level sequence, the second High Falling tone changes to a High Level tone when there are two neighbouring High Falling tones. When there is an alternating sequence of High Falling and High Level tones, the first High Falling tone is influenced by the High Level tone and is changed to a High Level tone. In contrast, Thai does not have any sandhi rules that operate on its five lexical tones although there are some effects of intonation on tone (see section 3.4.2.1). In addition to this difference between Cantonese and Thai, the shape of these lexical tones also differs between the languages. Tone level rather than tone contours are more important in Cantonese whereas tone contours are more important in Thai than tone level (Gandour, 1983).

---

31 For interest, tone sandhi is also present in Taiwanese and Sanskrit.
As in Cantonese, in connected speech the underlying full tones of Mandarin may change their tone category modified under the influence of their tone phonetic environment. For example, the mid-falling-rising tone in Mandarin changes into a rising tone when followed by another low tone. When that happens the new rising tone (35) is perceptually indistinguishable from the lexical mid-rising tone (35) (Tsan, 2001; Xu, 1994). The reason for tone sandhi occurring on Tone 3 (T3) could be physiological. Studies of vocal physiology have highlighted the difficulty speakers of a tone language have in producing two consecutive dipping T3s. However, there is no explanation as to why T3 simplifies to a 35-like rising tone and not to 53 or 51, as Tone 1 and Tone 4 are considered to be even easier to articulate than T2 (Tsan, 2001).

(Xu, 1994) conducted a systematic comparison between tone variations modified by different phonetic environments, namely compatible and conflicting environments, and found that the degree a tone changed from its canonical form due to coarticulation varied depending on the nature of the tone context. In an environment in which adjacent tone units are compatible or similar along a phonetic dimension (e.g., a three word phrase with a falling, rising, falling tone pattern), the deviation was relatively small compared to a context where adjacent tone values are conflicting (e.g., a three word phrase with a high, rising, low pattern), even to the extent of changing the direction of a contour tone.

The phenomenon of tone downdrift or declination is also present in Mandarin and is realised in a high level-low-high level tone sequence, when the second high level is lower than the first because the low both lowers the second high level and raises the first high level (Xu, 1994). A clear understanding of tone downdrift relies on knowledge of interaction effects between tones and intonation (a topic of the next section), and thus the issue of tone downdrift is revisited in section 3.5.2.1.

### 3.4 The Perception of Tone

This section begins with a discussion of the perceptual dimensions of tone in Thai, Cantonese, and Mandarin, and which tone cues tone and non-tone language speakers
Lexical Tone

use to perceptually identify and discriminate tones. This is followed by a discussion of two issues in the perception of tone, the interplay between tone and intonation in tone languages and how this affects tone perception, and coarticulation between tones.

3.4.1 Perceptual Dimensions of Tone

Each monosyllabic word in Thai is minimally differentiated by tone but the phonology is such that the case of maximum five-way differentiation can only occur on syllables ending in long vowels, diphthongs, or nasals (Abramson, 1975c). In perceptual experiments with synthetic speech, Abramson (1962) demonstrated the sufficiency of the tone F0 for high intelligibility in the labelling of monosyllabic words. In fact, while the absolute levels of fundamental frequency carry information used to correctly identify static tones, correct perceptual identification of static tones improves with slow movement in the F0 of the synthetic tones (Abramson, 1978b), and rapid F0 movement is a sufficient cue for the identification of dynamic tones. Furthermore, when Thai speakers are presented with a synthetic continuum of level fundamental frequencies presented in isolation, perceptually they can divide this into the high, mid and low tones fairly well. However, contrary to this finding Abramson (1978b) found that native Thai speakers have difficulty dividing the synthetic tone continuum due to the lack of change in the fundamental frequency of the synthesised tones. Similarly, this is the case with natural speech, where there must be some auditory accommodation to the pitch range of the speaker (Abramson, 1975c), as well as to the speaker's tone space (Abramson, 1986). For example, Abramson (1986) found that some Thai speakers provided only minimal cues for the distinction between mid and low tones spoken in isolation and without access to the tone range of each speaker, the listeners had difficulty identifying the tones.

Over and above the F0 information in the speech signal, there are several important perceptual dimensions of the pitch component of tone. Early work by Liang (1963) showed that in Mandarin 94.6% correct tone recognition can be achieved in the absence of F0 information when speech is high-pass filtered at 300Hz. Thus pitch is perceived even in the absence of F0 information, and tone recognition can occur on the basis of such residue pitch that is associated with higher harmonics (Schouten, 1940). What is
more, Liang also found 64% correct tone recognition in whispered speech, which contains neither fundamental frequency nor spectral envelope cues. The whispered speech results indicated that the temporal waveform envelope (duration cues) also contained information for tone recognition. More recently Fu, Zeng, Shannon, and Soli (1998) systematically studied the trade-off between the spectral and temporal envelope cues in Mandarin for recognition of consonants, vowels, and tones, in words and sentences. They found a high level of recognition with only the temporal cues, and tone recognition at the word level was found to play a significant role in sentence recognition. The results show that beyond F0, other acoustic cues such as duration, can contribute to tone recognition. However it is important to note that the relative role of duration in tone recognition may differ over languages. Abramson (1972) found that tone contrasts were not well preserved in whispered Thai, so it appears that duration may be a more important cue in Mandarin than in Thai.

In a cross-language study of tone perception, Gandour and Harshman (1978) reported five perceptual dimensions used by speakers of the tone languages Yoruba and Thai, compared with the non-tone language English. These are average pitch of tone, direction of tone, tone length, tone contour, and slope of contour. Perceptual data for these dimensions showed reliable differences between the tone language speakers and the non-tone language speakers, but no particular dimensions separated the tone language speakers of Yoruba and Thai. Specifically, all three groups attended to the average pitch and tone length dimensions, while only the Thai and Yoruba speakers used the direction and slope dimensions. Moreover Thai speakers’ relative emphasis of the length dimension compared to average pitch, direction, and slope dimensions, was greater than for the Yoruba and English speakers. On the basis of these results Gandour and Harshman (1978) argue that average pitch and length are either linguistic-phonetic or non-linguistic auditory dimensions, whereas direction and slope are specifically linguistic-phonetic dimensions of tone that are primarily used by tone language speakers. Thus while the length dimension represented a perceptual difference between stimulus tones for all subjects, its use by Thai speakers may also reflect differences in vowel length under the various F0 contours. For example, vowels are typically produced with longer duration under rising tones than falling tones (Abramson, 1962).
In a subsequent study Gandour (1979) investigated the perceptual dimensions of tone with 114 native Thai speakers. Of these Thai speakers, 38 spoke only Central Thai, and 76 spoke the Thai dialect of their local province but had at least 10 years experience with the Central Thai dialect. Pitch patterns were manifested on a synthetic speech-like syllable\textsuperscript{32} phonetically approximating [wa]. Participants were instructed to judge the similarity of the pitch patterns of paired stimuli presented in a carrier sentence (to facilitate speech mode processing) and to ignore any other differences between the stimuli. Multi-dimensional scaling analyses indicated that four dimensions underlie native Thai speakers’ perceived dissimilarity between the stimulus tones. In order of perceptual salience these were average pitch and three dimensions that serve a linguistic function in Thai, length of vowel duration in Thai (i.e., long or short), direction of pitch movement, and slope, (i.e., level or contour). Unexpectedly, there was little difference in perceptual judgements across dialect subgroups, but Gandour (1979) suggested that the dominant influence of Central Thai as Thailand’s official language might have ameliorated any dialect differences in the use of tone perception dimensions.

To investigate the role of language background on tone perception further, Gandour (1983) tested tone language participants who were speakers of Cantonese, Mandarin, Taiwanese (all part of the Chinese language group) or Thai, plus a control group of English speakers. On the basis of a paired comparison similarity task using 19 tones reflecting pitch distinctions found in East Asian languages superimposed on the [wa] syllable, two dimensions were identified. The first was primarily a height dimension, and the second primarily a direction dimension. Relative attention to these two dimensions was related to language background. In accord with the findings of Gandour and Harshman (1978), all English-speaking subjects loaded higher on the height than the direction dimension. Within tone speakers, Cantonese-speaking participants paid significantly more attention to height than did the Mandarin and Taiwanese groups, and Thai listeners attended more to the direction dimension than did the Chinese listeners.

\textsuperscript{32} The fundamental frequency values associated with the stimulus tones approximated the real F0 measurements of the Thai tones.
Somewhat surprisingly, a slope dimension did not emerge from the analyses, even though dimensions relating to level versus contour characteristics of tones were found in Gandour’s previous investigation (Gandour, 1979) using smaller stimulus sets. This failure may be due to the phonological relationship between level and contour tones in the four tone languages. For example, the three Chinese languages possess tone sandhi rules while Thai does not. Thus the Thai and Chinese language groups may have attached different degrees of importance to height and direction characteristics, as shown in the hierarchical clustering analyses, and with the greater number of Chinese listeners, a level vs. contour dimension may have been washed out.

Overall, these results converge upon pitch height as a fundamental perceptual dimension of tone used by both tone and non-tone language speakers. Direction (and possibly length and slope) serves to distinguish between tone and non-tone language speakers, and to some extent between speakers of different tone languages. Thus pitch height may be thought to be a general auditory dimension of tone perception while duration, length, and slope may be considered to be linguistic dimensions of tone used more by speakers of a tone language. In section 3.5 on the acquisition of tone, these perceptual dimensions will be discussed in relation to the development of tone perception in Thai and Australian children.

Next, the interaction of tone with intonation and other tones (coarticulation) is outlined.

3.4.2 Interaction Effects and Tone Perception

It was noted above that tone and intonation are not the same (section 3.1.1), and also that the production of tone can change in particular articulatory contexts, that is, tone sandhi (section 3.4.2.2). Tone interacts with other tones, other segments, and with sentence intonation, and these interactions influence how tone is perceived. In the sections below, the perception of tone and intonation are compared and the issues of sentence declination and tone coarticulation are raised.
3.4.2.1 Lexical Tone versus Intonation

Despite the distinction made above, lexical tone and intonation interact. For example even though Thai is a true tone language in which every syllable in the morpheme stock bears a controlled local F0 pattern of the tones, global sentence intonation is still present and interacts with tone (Abramson & Svastikula, 1983). In intonation, as in tone, F0 levels and contours are the main phonetic foundations, although other features such as amplitude shifts do play a role. Nevertheless, despite the fact that the laryngeal and aerodynamic mechanisms responsible for controlling global intonation contours over time also simultaneously control local F0 patterns of tones, the communication of sentence intonation is unconstrained in the fluent tone language speaker (Abramson & Svastikula, 1983).

Declination (sometimes called downdrift or downstep) is a typical feature of the intonation pattern of a declarative sentence whereby F0 gradually reduces over a sentence. Anderson (1978) describes it as a "steadily falling (or rather monotonic nonincreasing series of pitch levels)" (p. 138; see also section 3.3.3 in relation to how declination changes relative tone levels in Cantonese and Mandarin).

Abramson and Svastikula (1983) investigated whether there is an interaction between declination and tone contours in simple three-word and complex (more than three-word) declarative Thai sentences by comparing the F0 peaks at three diagnostic parts in each sentence. Declination was more evident in complex declarative sentences than in simple sentences, and on average across the sample of sentence stimuli, all five tones kept their distinct contour despite the fact that due to intonation declination absolute F0 values of tones declined over the course of the utterance.

Luksaneeyanawin (1984) investigated how the systems of lexical tone and sentence intonation interplay to form the melody of Thai speech. Using cue cards Luksaneeyanawin elicited both grammatical utterances (e.g., statements, yes-no questions, unfinished sentences, and 'please continue'), and attitudinal utterances (e.g., emphatic/unemphatic, interested/bored, angry/concealing anger\textsuperscript{33}) from participants.

\textsuperscript{33} For the full list of attitudinal responses, refer to Luksaneeyanawin (1984).
Lexical Tone

Four intonation (tune) patterns were identified, namely, (a) Tune 1 - present in statements, citation forms, attitudinally unmarked and submissive sentences; (b) Tune 2 - present when asking questions and in unfinished statements, and used to convey the attitudes ‘disagreeable’, ‘disbelieving’ and ‘surprised’; Tune 3- the telephone ‘yes’ intonation, present when conveying ‘concealed anger’, ‘boredom’, and ‘authoritativeness’, and (d) Tune 4- the intonation for ‘emphatic’, ‘angry’, ‘agreeable’, ‘interested’, and ‘believing’ intonation.

The changes to tones under Tunes 1, 2 and 4 are most interesting. Phonetic analyses revealed that in Tune 1, recognition of static tones (low, mid and high) became confused because intensity was uniformly high across utterances, however the two dynamic tones (rising, falling) maintained their distinct pitch patterns. The F0 of the mid and low static tones became more similar in Tune 2 but remained distinct in terms of intensity or loudness. In Tune 4, the high and mid static tones and the dynamic falling tone have a rising-falling phonetic realisation, while the low tone has a rise at the end.

Luksaneeyanawin (1984) draws a similar conclusion to Abramson and Svastikula, (1983), that while the fundamental frequency of the tones may become very similar under some patterns of intonation resulting in some confusion (e.g., Tune 1), the phonetic features of the tones remain relatively distinct if there is access to the speaker’s pitch range, pitch height and pitch shape of surrounding tones. Thus intonation interacts with rather than interferes with the system of tone in Thai.

In Cantonese, Chan (1999) recorded simple declarative sentences spoken by several Cantonese speakers with a neutral expression. High F0 was generally found at the beginning of an utterance and lower F0 values at the end, indicating a falling intonation pattern. In addition, there was a progressive lowering of pitch over the sentence unit; tone contours became lower as the speaker reached the end of the sentence. Nevertheless the Cantonese tones were still distinct.
3.4.2.2  Lexical Tone and Coarticulation

Tone languages such as Thai (Gandour, Potisuk, Dechongkit, & Ponglorpisit, 1992), Mandarin (Xu, 1994) and Vietnamese (Han & Kim, 1974) are not immune to coarticulation, the phenomenon in which the phonetic properties of a given speech sound are altered by the phonetic properties of adjacent sounds. Coarticulation can be anticipatory, in which one speech sound is influenced by subsequent speech sounds, or perseverative, where a speech sound is influenced by preceding sounds (i.e., carry-over effects). Both anticipatory and perseverative effects are found in consonant and vowel sounds in languages all over the world, but the degree of effect has been found to be asymmetric. For example, Ladefoged (1972) found that adjacent vowels affect the articulation of consonants in English in an anticipatory manner, whereas in French, there are equal anticipatory and perseverative effects.

For tones, coarticulation means the influence of adjacent tones on each other. The few studies of coarticulation effects on tones have focused on the Asian tone languages, Vietnamese, Mandarin, and Thai, and have generally shown that coarticulatory effects are more often perseverative than anticipatory (Ladefoged, 1982).

The tones of Vietnamese show greater perseverative than anticipatory effects, with coarticulation affecting F0 contour and height (Han & Kim, 1974). In Mandarin, the four lexical tones also show changes in F0 height due to perseverative and anticipatory coarticulation, with no effect on contour (Shen, 1990). In Thai, Abramson (1975a) found variation in F0 contour and height of the mid tone onset in final sentence position to be a function of preceding tone type. He also found initial perturbation in F0 height and F0 contour for falling tones followed by a mid tone.

Gandour, Potisuk et al. (1992) took a different approach in their investigation of perseverative coarticulatory effects in Thai disyllabic utterances by using quantitative measures of F0 height and F0\textsuperscript{34}. Ten native male Thai speakers were recorded reading disyllabic noun-verb phrase pairs in which the tone carried on the first syllable varied, but the tone on the second syllable was constant across the two items of a pair. Five

\textsuperscript{34} Anticipatory coarticulatory effects were not investigated in this study.
native Thai listeners were required to identify the phrase they heard and only those identified correctly by at least four of the five listeners were analysed for tone coarticulation. Minimal perseverative coarticulation was found: There was no effect of initial syllable tone on second syllable F0 height. However, F0 slope of the mid tone was affected by the toneme of the preceding syllable; mid tone onset was higher and thus the slight F0 drop used in the Thai tones was greater when following a low tone and when tone slope was falling after a falling tone. These results differ from those of Abramson (1975a), possibly due to methodological issues. Abramson (1975a) used phrases embedded in sentence form whereas Gandour, Potisuk et al. (1992) used phrases produced in isolation rather than with sentential context. Gandour, Potisuk et al. (1992) also held vowel quality and the occurrence of plosives for the initial consonant of the final syllable constant, whereas these features were not as rigorously controlled by Abramson (1975a). So while there are definitely coarticulatory tone effects, there are issues of method that need to be addressed before definitive conclusions can be drawn.

3.4.2.3 Summary

It is clear that the relationship between tone and intonation is complex. The two most important points shown here are that there can be changes to tone F0 depending on intonation, and that despite F0 differences in tone due to intonation, native listeners show no perceptual confusions of the tones. In the running speech of tone languages, tone coarticulation causes some departure from the idealised contours (i.e., F0 when a tone is produced in isolation) of the lexical tones, but for the most part native listeners can correctly identify tones despite such coarticulation.

3.5 The Acquisition of Tone

The widespread use of tone in the world’s languages as evidenced by the large number of tone language speakers in many parts of Asia, Africa, and Central and South America, is indicative of how important the study of tone is. Most developmental studies of language have investigated the acquisition of phonological perception and production in children acquiring Indo-European languages (Clumeck, 1980). Consequently, the acquisition of tone has been relatively neglected despite the many
 Lexical Tone

tone languages in the world and the large proportion of the world’s population (over half) who speak tone languages. The term ‘language acquisition’ is typically used to refer to production, but as we have seen for segments and suprasegmentals, before they can talk, infants learn perceptually about language as well. Therefore both perception and production of tone are considered here.

Tone acquisition refers to both the phonetic and phonological development of tone. The phonetic development of tone refers to a child’s accuracy in producing tone relative to adults’ production of tone, while the phonological development of tone refers to the child’s deeper understanding of tone - an understanding of what pitch levels are used contrastively in the native language and which are irrelevant (Clumeck, 1980). But if recent research by Burnham, Davis et al. (2003) is anything to go by, children do not acquire tone phonologically (this issue is discussed later in section 3.6.4.3).

The study of the acquisition of tone by children learning a tone language as their first language raises a number of interesting questions about the development process such as:

1. As tone contrasts meaning, is the acquisition of tone similar to the acquisition of segments, or in so far as tone can also be considered to be suprasegmental, is tone acquisition similar to the acquisition of stress and intonation (Clumeck, 1980; Li & Thompson, 1978)?

2. When do children begin to use pitch in a lexically contrastive way?

3. Are some tones produced earlier than others?

4. At what age are tones fully acquired?

5. Does the developmental course of tone perception parallel that of tone production?
This section will deal with these questions with particular reference to the data available in the languages of interest - Thai, Cantonese, and Mandarin. The section is divided into two main parts, the first dealing with tone production and the second with tone perception.

3.5.1 Production

Tuaycharoen (1977) studied tone acquisition in a single Thai infant. She reports that from the time of production of the first words, the infant correctly produced the mid and low tones but also substituted these tones for the falling, high, and rising tones. Rising tone was first used at 14 months of age for word items, and use of the high and falling tones was evident by the end of 15 months of age. Full tone acquisition was evident by nearly two years of age, but the child had not yet mastered the production of diphthongs, triphthongs and initial consonant clusters.

Studies of tone acquisition in the Chinese language family similarly reveal that tonemes are acquired earlier and produced more accurately than are consonants or vowels. In a longitudinal study following children between 18 and 36 months, Li and Thompson (1977) investigated the spontaneous speech of Mandarin-language children. The results show that the Mandarin tone system is acquired relatively quickly, and earlier than consonants and vowels. The high and falling tones of Mandarin are acquired earlier and more easily than the rising and dipping tones, and tone sandhi rules are learned when the child begins to produce multi-word utterances. In contrast, segmental inaccuracies were frequent for production of fricatives, affricates, and the Mandarin liquid.

Data comparing tone with segmental acquisition in Cantonese comes from a case study by Tse (1977), who observed his son’s development. Perceptual discrimination of Cantonese tones began when the child was 10 months of age. Between 16 and 20 months, the single word stage, the child mastered production of tones in the following order, high-level tone, low-level tone, mid-level tone and high-rising tone. Production of the remaining tones, the low-rising and the low-falling tone, occurred at 21 months.

---

35 Interestingly, rising tone was used in nonsense syllable production but not for lexical items.
along with the emergence of two word utterances, however the child demonstrated some confusion between the high rising and the low rising tones.

In summary, data from the Thai and Chinese languages indicate that the tone system of a language is mastered at an earlier age than the segmental system, and that tone production is relatively accurate and robust compared to production of consonants and vowels.

3.5.2 Perception

The number of studies examining when children first perceive pitch as having a lexical function are even fewer than those on tone production. In his case study of a 10-month-old child, Tse (1977) investigated whether the tone or the segmental information for the word ‘light’ was more salient. The child’s responses were more “sure” when correct tone information was provided with incorrect segments than in the inverse situation. The tests were administered numerous times across a 2-month period, with similar results each time, and Tse (1977) concluded that tones are probably more salient to the child than segments. However due to the informal nature of the tests, and as only one child was tested, care should be taken in generalising from these results.

3.5.2.1 The Use of Tone Cues Over Age

Evidence that language experience influences tone discrimination comes from Burnham, Kirkwood, Luksaneeyanawin, and Pansottee (1992), who tested Thai and Australian speaking adults in a same/different discrimination of all possible combinations of the five Central Thai tones, plus two consonant contrasts. Thai speakers were significantly better at discriminating the contrasts but both Thai and English speakers discriminated dynamic-dynamic tone pairs more easily than either static-static or dynamic-static pairs. Burnham and colleagues suggested that the most salient and perceptually useful acoustic dimension appeared to be F0 onset, although they conducted no formal analyses.

In a subsequent study, Burnham and Francis (1997) investigated whether the differences found by Burnham et al. (1992) for Thai and English language adults are
also obtained for children (8-, 6-, and 4-year-olds) from these language environments. Burnham and Francis (1997) found significantly better discrimination of tone contrasts by Thai speakers compared to English speakers, and for both language groups, discrimination performance increased as a function of age. They also correlated the acoustic variables for each contrast with discrimination accuracy and found results consistent with the conclusion that Thai speakers used mean F0 at all ages, and that the use of the onset value of F0 increased over age. Thai speaking children also used cues such as F0 offset and variability in F0 in their perceptual judgements cues that Thai adults did not use. In contrast, young English speakers did not use mean F0, relying more on F0 onset and offset, whereas English-language adults used mostly mean frequency (and to a lesser extent F0 standard deviation) as a cue for discrimination.

These results support the view that exposure to a tone language across development fine-tunes a listener’s use of cues for lexical tone, and that over age speakers of a tone language integrate more cues in their perception of tone. Burnham and Francis’ (1997) data support Gandour’s (1983) findings that pitch height is an important perceptual dimension in adults’ tone discrimination, for both tone and non-tone language speakers. However, Burnham and Francis (1997) found that young listeners from both tone and non-tone language environments rely on a complex set of acoustic variables rather than a single dimension, and that only with extensive tone language experience does the young listener learn the optimal cues for tone perception.

3.6 Issues in Tone Perception

It was made clear in section 3.4.2 that in the face of F0 variation in tone due to intonation and coarticulation, native listeners show invariance in tone perception. In this section a number of additional issues relevant to tone perception are considered, specifically mode of processing in tone perception, categorical tone perception, the role of linguistic experience in tone perception, the relative salience and perceptual status of tones and phones (consonants and vowels), and the role of musical training in tone perception.
3.6.1 Mode of Processing in Tone Perception

Two very different areas of research, the influence of inter-stimulus intervals on tone discrimination, and hemispheric processing of tone and non-speech tone by native, non-native, and aphasic listeners, are reviewed later in sections 3.6.1.2. However both these domains of investigation reveal that linguistic experience influences how tone is processed, an issue developed further in section 3.6.2.1.

3.6.1.1 Evidence from Inter-Stimulus Interval Studies

Throughout chapter 2 it was shown that linguistic experience systematically biases speech perception abilities toward native language specific phonological distinctions, such that adults have more constrained speech perception abilities than infants from identical language backgrounds. It was further pointed out that this decline in speech perception abilities is not a sensorineural loss, for adults are able to perceive consonant contrasts that are irrelevant in their native language when the to-be-discriminated sounds are presented at an interstimulus interval (ISI) of 500ms but not 1500ms (Werker & Logan, 1985; Werker & Tees, 1984b). Werker and colleagues suggested that an ISI of 500ms allowed a language-general mode of perception in which phones were perceived directly, without any influence of specific linguistic experience. Thus, the non-native listeners were able to perceptually discriminate contrasts that were not phonologically relevant in their native language by operating at a phonetic level of perception. In contrast, an ISI of 1500ms encourages perceptual processing in a phonological language-specific mode. In this mode the perception of non-native speech contrasts is inhibited, and that of native contrasts is enhanced because incoming speech sounds are coded in terms of the native phonological categories.

In relation to mode of processing in tone perception, Cutler and Chen (1997) using a speeded same/different tone judgement task with a 250ms ISI, found that Cantonese speakers and Dutch (non-tone language) speakers had longer reaction times and made more errors when there was only a tone difference between the stimuli than when there was a consonant, vowel, consonant-vowel, consonant-tone, or vowel-tone difference. The results were interpreted as evidence that tone processing is auditory in nature rather than linguistic (phonemic), irrespective of whether or not the listener is familiar with
Lexical Tone

Nevertheless this conclusion could be considered premature, as Cutler and Chen (1997) only provided subjects with a 250ms ISI, forcing them to process the tone information in a very short period of time, at an acoustic level. Participants may have performed better and in accordance with expected language-specific patterns if they were also tested at longer ISIs that encouraged phonetic and phonemic modes of processing.

Burnham et al. (1992) and Burnham and Francis (1997) investigated the phonetic and phonological modes of perception with tone contrasts. They tested Thai-speaking and English-speaking adults on their AX (same-different) discrimination of Thai tone contrasts36, and found that English speakers’ discrimination was generally better with a 500ms than a 1500ms ISI, and that for Thai adults, discrimination was better at a 1500ms than a 500ms ISI. As well as testing adults Burnham and Francis (1997) tested Thai and English speaking children of 4, 6 and 8 years of age. Like their adult compatriots, Thai speaking children were better at discriminating the tone contrasts at a 1500ms ISI (phonological mode) while English speaking children were much more competent at discriminating tones at a 500ms ISI (phonetic mode). Therefore non-tone language speakers perceive tone better in a phonetic mode, whereas native tone language speakers perceive tone better in a phonological mode of processing. Future studies using a combination of two methods, reaction times to different feature combinations, and the ISIs (250ms, 500ms, 1500ms) used by Cutler and Chen (1997) and (Burnham et al. 1992) would be enlightening.

In conclusion, the results of the studies outlined in this section suggest that experience plays a role in tone discrimination, and that this role can be uncovered in ISI and task complexity manipulations. In general it can be concluded that tone language speakers process tone in a phonemic manner from an early age, whereas non-tone language speakers (even those with some tone language experience) do not. This implies that tone is phonemic (or tonemic) just like consonants and vowels.

36 Burnham et al. (1992) also tested the adults on the consonant contrasts [b]-[pʰ], finding phonetic processing by English speakers and phonemic processing by Thai speakers, but discrimination of the tone contrasts is of most interest here.
3.6.1.2 Evidence from Hemispheric Differences and Aphasics

Investigating hemispheric brain processing in tone perception is enlightening with respect to whether tone is processed linguistically by both tone and non-tone language speakers and by tone language speaking aphasics with language processing deficits. In a dichotic perception paradigm, Wang, Jongman, and Sereno (2001) found that Mandarin tones were predominately processed in the left hemisphere\textsuperscript{37} (a right ear advantage) by Mandarin speakers for whom the stimuli are lexically meaningful, whereas they were more bilaterally processed by American English speakers. In a similar vein, there is evidence that Thai (Gandour & Dardarananda, 1983; Gandour, Ponglorpisit et al., 1992), Mandarin (Naeser & Chan, 1980), and East Norwegian (pitch-accent) (Moen & Sundet, 1996) aphasics with left-hemisphere lesions have difficulty with the identification of their native lexical tones and pitch accents respectively.

In a cross-linguistic brain imaging study, Gandour, Wong, and Hutchins (1998) used Positron Emission Tomography (PET) to compare patterns of activated foci when native speakers of English and Thai perform same-different judgements of Thai words and nonspeech low-pass filtered equivalents. They assumed that any observed differences in brain activation patterns between the two tasks are likely to reflect differences in cognitive processing. When discriminating tones in Thai words the Thai speakers showed activation in the left hemisphere, namely the frontal operculum involving the language region Broca’s area, almost certainly because pitch variations in speech are perceived as phonologically significant at the lexical level in the Thai language. However, such activation was not observed for Thai listeners presented with non-linguistic pitch variations nor was it for English listeners’ discrimination of either non-linguistic pitch or tones. These results suggest that the English listeners implicitly focused on phonologically significant sounds, that is, consonants and vowels, before extracting the F0 patterns of the Thai tones. This differential outcome across language group for tone discrimination, and across tone and pitch tasks, strongly suggests that linguistically relevant properties are seminal in determining which neural mechanisms are engaged in the perception of pitch-related cues. Later work by Gandour et al.

\textsuperscript{37} The left hemisphere of the brain is more linguistically sophisticated than the right hemisphere which is more proficient at affective functions Wang et al. (2001).
(2000) provides converging evidence for activation in the vicinity of Broca’s area for Thai listeners’ perception of lexical tone, but not for homologous pitch patterns in a non-linguistic context. Interestingly in this study Gandour and colleagues also tested Chinese and English listeners for their discrimination of the tone and non-linguistic pitch pairs. Despite familiarity and experience with lexical tones, no significant activation was observed in Broca’s area when Chinese listeners’ heard Thai tones, compared with the pitch or resting baseline conditions. Thus the lateralization evident for tone appears to be language-specific, presumably because the Thai tones hold phonological significance for the Thai listeners but not to listeners who are experienced with another tone language, Chinese, or a non-tone language, English.

3.6.2 Experiential Influences on Tone Perception

Given the fact that tone language speakers process tone differently to non-tone language speakers, it is of interest to investigate how stimulus context and musical training may influence the perception of tone by tone and non-tone language speakers.

3.6.2.1 The Role of Linguistic Experience in Tone Perception

The results of many cross-language studies have shown that linguistic experience influences the way in which tone is perceived. In an earlier section, 3.6.1.1, Burnham and Francis’ cross-linguistic findings were presented showing that Thai adults and children best use a phonological mode of processing to discriminate Thai tone pairs. However Burnham and Francis (1997) also found that across age, and thus with increased tone language experience, a greater reliance on the phonological mode was evident. Specifically, the 6- and 8-year-old children and adults were better at discriminating in a phonological mode than younger 4-year-olds.

Recently, Wayland and Guion (2003) compared the ability of English speakers learning Thai as a second language (between one and five years experience), native Thai speakers, and monolingual English speakers with no experience learning Thai, on the
Lexical Tone

discrimination of the mid versus low Thai tone contrast in closed and open syllables. Results showed that linguistic experience did play a facilitative role in tone discrimination. The native Thai group had higher discrimination scores than both the Thai second language learners and the English speakers with no knowledge of Thai. Moreover, it was found that native Thai speakers and Thai second language (L2) learners performed equally well at discriminating the closed syllable tone contrast, suggesting that exposure to and experience with the Thai language can assist in the development of native-like discrimination among adult L2 learners of a tone language.

In section 3.6.2.1 it was reported that listener’s linguistic experience appears to affect the dimensions underlying the perception of tones (e.g., Gandour, 1983), and there is also evidence that experience affects the accuracy and reaction time of tone perception. For example, Lee, Vakoch, and Wurm (1996) investigated whether native Cantonese and Mandarin speakers were better at discriminating same-different lexical (word) and non-lexical (nonword) tone pairs from their own tone language than a foreign tone language. They also tested a control group of non-tone English speakers. Reaction time and error rate data showed that Cantonese speakers were noticeably better at discriminating the lexical and non-lexical pairs than Mandarin and English speakers (who performed similarly to each other). Furthermore, Cantonese speakers discriminated the lexical tone pairs better than the non-lexical pairs presumably because they are experienced with the lexical but not the non-lexical tones. Mandarin speakers were superior to Cantonese speakers at same-different judgements of the Mandarin lexical and non-lexical tone pairs, but unlike the Cantonese speakers, they performed equally well on the lexical and non-lexical tone pairs. Lee et al. (1996) suggest that Mandarin tones are easier to discriminate than Cantonese tones. This is supported firstly by less errors and shorter reaction times, and secondly because Mandarin speakers did not discriminate the Cantonese tone pairs more reliably than the English speakers, whereas Cantonese speakers discriminating Mandarin tones did.

In conclusion, the results of the above studies provide experimental evidence that experience with a tone language affects tone perception in two crucial ways. Firstly,

38 In an acoustic analysis of their stimuli, Wayland and Guion (2003) found that the mid and low tones were more distinct in the closed syllable stimuli than in the open syllable stimuli. Thus they predicted that the mid vs. low tone distinction would be easier to discriminate in closed than open syllables.
Lexical Tone

tone language speakers are superior to non-tone language speakers in their discrimination of lexical tone contrasts. Secondly, at least among speakers of the two most widely spoken Chinese dialects - Cantonese and Mandarin - the discrimination of native tone contrasts is better than discrimination of nonnative tone contrasts.

3.6.2.2 The Role of Context in Tone Perception

Given the influence of language background on tone perception, it is of interest to investigate the effect of removing linguistic context on non-tone language speakers' perception of tone. Burnham, Francis, Webster et al. (1996) investigated whether non-tone language speakers' inferior perception of lexical tones is evident when pitch variations of lexical tone are embedded in nonspeech contexts. English speaking and tone language speaking adults of Thai, Cantonese, and Swedish language backgrounds (Experiment 1), and Thai and English speaking children (Experiment 2), were tested for their discrimination of Thai tones in three contexts - speech, filtered speech, and violin sounds. It was found that English speaking children and adults discriminated tone contrasts in the two nonspeech contexts, musical and low-pass filtered speech, better than the same pitch variations presented as speech (lexical tone), thus suggesting that speakers of a non-tone language have attenuated ability for the discrimination of pitch in speech. The tone language speaking adults and children performed equally well on pitch differences in all three contexts, suggesting that tone language speakers' success outside of a linguistic context is an extension of their attunement to psychoacoustic pitch perception. An alternative is that tone language speakers' linguistic perception is the same as their non-linguistic perception of pitch, that is, linguistic experience does not reduce pitch perception; in fact it probably increases it. Furthermore, this behavioural data proposes that language experience may influence how auditory cues such as pitch are processed. As discussed in 3.6.1.2, herein lies the fundamental question of whether shared neural mechanisms at higher cortical levels are engaged for pitch perception of linguistic lexical tone and non-linguistic perception of pitch. Burnham et al., (1996) found equivalent discrimination of lexical tone and non-linguistic pitch by tone language (Cantonese) and pitch-accent (Swedish) speakers. Gandour et al. (1998) found similar behavioural results with tone language speakers, but differential cortical activation, with
lexical tone and non-linguistic pitch processed in the left and right hemispheres respectively.

3.6.2.3 The Role of Musical Training in Tone Perception

As language experience leads to the perception of tones in a linguistic mode (Burnham & Francis, 1997), and as tone does not equal pitch (Gandour & Harshman, 1978), it would appear that other non-linguistic training with pitch would not affect lexical tone perception. Burnham et al. (1992) investigated whether musical training has a bearing on tone perception. Musically trained adults with absolute pitch ability, musically trained adults without absolute pitch, and non-musically trained adults, all from non-tone language speaking backgrounds, were tested for their discrimination of Thai tone pairs, presented either as speech (carried on the syllable [ba]), filtered speech, or as violin sounds. Non-musicians discriminated tones better in music than in filtered speech, and better in music and filtered speech than speech. The musicians without absolute pitch also discriminated tones better in music and filtered speech than in speech but they performed significantly better in all three contexts compared with the non-musicians. Surprisingly musicians with absolute pitch had equally good discrimination of Thai tone pairs in all three contexts and this group performed significantly better overall than both the musicians without absolute pitch and the non-musicians. These findings suggest that for English speaking adults, musical training may facilitate speech perception and performance on general pitch perception tasks, and that the ameliorating effect of learning a non-tone language on pitch perception in speech is not evident in absolute pitch individuals.

3.6.2.4 Summary

The role of linguistic (and musical) experience in tone perception appears paramount. When discriminating between tone contrasts, tone language speakers use different perceptual cues and integrate multiple acoustic features in contrast to English (non-tone language) speakers. Tone language speakers perform significantly better in lexical tone discrimination than do non-tone language speakers, and demonstrate equivalent discrimination for tones presented in nonspeech contexts (Burnham et al., 1996). Interestingly, second language learners with some experience of learning a tone
LEXICAL TONE

language can discriminate particular Thai tone contrasts as well as native Thai speakers (Wayland & Guion, 2003), while non-tone language speakers with no experience learning Thai are better at discriminating musical tones than speech-like tones (Burnham et al., 1996). However non-tone language speaking adults do benefit from musical training and demonstrate better discrimination of tones in speech and nonspeech contexts compared to non-musicians, and musicians with absolute pitch perform even better again (Burnham & Brooker, 2002).

3.6.3 Is Tone Perceived Continuously or Categorically?

The investigation by Abramson (1961) was integral to the first decade of research into the categorical perception of tone. Abramson studied the identification and discrimination of phonemic tones in Thai by creating a set of five fundamental frequency contours with final points moving upwards from a level starting point and a diminution of F0 at the end. In an ABX task, native Thai listeners divided this continuum into the mid and high tones. Overall discrimination of the tones was uniformly high with little evidence for discrimination peaks at the category boundary, therefore indicating that perception was continuous, not categorical.

During the second decade, Abramson's (1961) longstanding conclusion that tone perception is continuous, was challenged by Chan, Chuang, and Wang (1975), who found categorical perception of a fundamental frequency continuum that approximated the rising versus level Mandarin tones. The results showed good correspondence of the crossover points in an identification task and the peak in the discrimination functions. However, contrary to what might be expected when the subjects were tested on the perception of nonspeech stimuli with identical F0 characteristics, a similar categorical boundary was found. In a later study Abramson (1977) responded by studying level tones, producing a continuum of flat variants that sought to yield the low, mid, and high tones of Thai. Here Abramson (1977) tested many more subjects than in the Abramson (1961) study. There was considerable overlap of labelling in the identification data. The results of the discrimination tests showed a high level of discrimination across the continuum with no apparent boundaries between categories. The results of the experiments reaffirm the work by Abramson (1961), therefore giving more strength to
the argument that a tone continuum is perceived continuously. Abramson (1977) cites differences between the level and contour continua used by Abramson (1961) and Chan et al. (1975) as a possible reason for the discrepant findings. Chan et al. (1975) synthesised the isolated vowel /i/ to create two tonally differentiated words, whereas Abramson (1977) used the syllable [khaa] to create three different words.

One possible reason for Abramson’s finding (1977) of relatively poor category boundaries in the identification of Thai level tones is that the stimuli were presented in isolation. Thus in a recent study of the perception of Cantonese tones, Francis, Ciocca, and Ng (submitted) examined whether carrier sentences (contextual information) assist in the categorical identification of level tones because the context allows the listener to normalise the speakers’ frequency range (Experiment 1). Listeners exhibited heightened sensitivity to F0 differences between pairs of tones that span category boundaries. In Experiment 2 and Experiment 3, Francis et al. (submitted) replicated the Abramson (1977) and Chan et al. (1975) studies but with Cantonese Level and Cantonese High Rising to High Level tone continua. In the experiment with level tones, the results were relatively consistent with Abramson’s findings for discrimination of Thai level tones. However using carrier sentences for the tone stimuli facilitated listeners’ sensitivity to category boundaries in the identification task when compared with (Abramson, 1977) results. The results of the experiment with contour tones (also using carrier sentences) provide tentative support for the categorical perception of contour tones and the findings of Chan et al. (1975), with the presence of a pronounced peak in sensitivity towards the middle of the continuum.

More recently, Burnham and Jones (2002) have investigated the perception of three artificial tone continua - mid-rise, mid-low and low-rise - in a two alternative forced-choice identification task, by Thai (tone) and Australian English (non-tone) speaking adults. The continua were presented in four different contexts, as speech, filtered speech, sine waves, and violin (music) sounds. There were two main findings. First, identification functions for the speech stimuli were more categorical for Thai than for Australian English participants. Therefore, it appears that long-term experience at making phonological distinctions on the basis of perceived pitch may induce categorical-like perception of native speech-tone (F0) continua. Secondly, Thai
listeners did not perceive the nonspeech continua in a more categorical-like manner than did Australian English listeners and nonspeech continua were generally perceived less categorically to speech continua and similarly to one another, with no more categorical-like perception for the filtered speech continua (assumed to be more speech-like) than the sine wave nonspeech. Together these findings show that whatever categorical speech perception there is for tone, it appears to be learned as a product of linguistic experience and specific to speech. While Burnham and Jones (2002) found convincing evidence for categorical identification, this response pattern should not be referred to as categorical perception because a discrimination paradigm was not used in conjunction with the identification task. A discrimination experiment by Schwanhausser, Jones, and Burnham (2003) in the vein of the Burnham and Jones study (2002) has found that tone and non-tone language speakers employ different strategies for discriminating a novel synthetic low-rise tone continuum. Vietnamese and Mandarin speakers appear to use a strategy in which they perceptually split the low-rise continuum - they split the continuum in half and treat it as two categories, upper and lower. On the other hand Australian English speakers use a more acoustic-based strategy in which the flat, not the contour tone, is taken as the perceptual anchor.

Contradictory evidence comes from Stagray and Downs (1993) who compared differential sensitivity to absolute frequency of standard (nonspeech) tones between Mandarin speakers and English speakers. In comparison to the English listeners, the Mandarin speakers exhibited a decreased sensitivity to pitch changes of similar frequency corresponding to within-category differences in tone. Thus practice with a tone language may have strengthened Mandarin speakers' categorical perception of nonspeech pitch changes. Therefore it is possible that under some conditions long-term experience at making phonological distinctions on the basis of perceived pitch may induce categorical-like perception of nonspeech F0 continua.

Is tone perceived categorically or continuously? Indeed, the evidence is at a stalemate. However there is hope of a resolution - a new research project being conducted at the

---

39 Burnham and Jones (2002) state that the Australian English participants reported not hearing the filtered speech as a form of speech, and this may account for their perception of the continua being different to how speech was perceived.
MARCS Auditory Laboratories, University of Western Sydney, Australia by B. Schwanhäuser and D. Burnham, hopes to reconcile the debate.

3.6.4 Are Tones Phones?

The issue in this section is whether tones are (or are not) segments. Evidence for and against the ‘tone as segment’ argument is considered, from the domains of orthography, phonetics, acquisition, and phonological versus tonological awareness.

3.6.4.1 Linguistic Considerations in the Tone as Segment Argument

In tone languages, the specification of each morpheme includes consonant and vowel features, as well as a distinctive pitch pattern manifested in the fundamental frequency of the voice. Tones are carried on the vowel of a syllable and there is some debate over whether tones should be treated holistically, with the syllable as the domain of the tone, or whether there should be a segmental treatment of tones with the vowel, single, double, or final sonorants (that the tone overlays), as its domain (Abramson, 1978a). The traditional view is that tones should not be segmented and that this is usually reflected in the orthography of tone languages. For example, in Thai the tone at which the syllable is spoken is integrated into the Thai script (for details of Thai orthography, refer to Burnham, Luksaneeyanawin, & Kantamphan, 2003), and children learn their basic vocabulary with the specific tone contour as part of the lexical item.

Segmentalists argue that there are linguistic constraints in the use of tones. Certain tones can only map onto certain segmental bases. For example, in Thai a syllable with a short vowel followed by a final stop can only carry the low or high tones, and a long vowel followed by a final stop can only carry the low tone.

Abramson (1978a) argues that a segmental treatment of tones may only be possible for languages that contain tone sandhi such as Cantonese and Mandarin but not Thai.

---

40 The Thai system for spelling dates back to the 13th Century. The writing is tiered so that initial letters are classed into high letters (top tier), mid letters and low letters (lowest tier). High letters represent voiceless aspirates (eg. [Ph, th, hm]), low letters represent the voiceless series ([b,d,m,n]). The mid levels represent the voiceless, non-aspirated stops or glottalised consonants (Haudricourt, 1972).
Under tone sandhi, vowels change their status across a morpheme boundary allowing the final static tone of the first morpheme and the initial static tone of the following morpheme to merge. Furthermore, if contour tones were just a sequence of level tones with a glide between them, then the phonetic evidence would be more consistent with a segmental analysis (Abramson, 1978a). However, there is no indication that contour tones are merely composites of static tones and therefore able to be segmented.

The most convincing evidence for the pro-segmental position comes from considering the laryngeal-muscle activity underlying the contours of the Thai tones. Erickson (1974) found distinctive muscle patterns for the five Central Thai tones and the separate and distinct muscle movements imply separate tones. The muscles implicated in the production of a rising tone were the thyrohyoid for its initial drop and the cricothyroid peak for rise. The falling tone showed a cricothyroid rise and a thyrohyoid peak for the sharp fall respectively. Distinct muscle groups could not be found for the static tones.

Some light may be shed on the tone as segment argument from behavioural data on the relative salience of tones versus phones, and tonological versus phonological awareness.

3.6.4.2. The Relative Salience of Tones versus Consonants and Vowels

Theoretical predictions regarding the relative salience of tones and segments and the ontogeny of tone and segmental perception can be gleaned from the production literature. As mentioned in section 3.5.1, infants begin to use both pitch and segments at a lexical level at 11 months (Tuaycharoen, 1977), and tone production is completed by 23 months ahead of full segmental production (Clumeck, 1980; Tuaycharoen, 1977). For children acquiring a tone language, it appears that pitch differences are more basic and more perceptually salient than segmental differences. However it is not clear whether these advantages occur simply because in tone languages there are less tone contrasts to learn in comparison to segments or whether tone contrasts are more salient psychoacoustically or linguistically than segmental contrasts.
When considering the experimentally-based rather than the case study literature, the conclusion that tone is more salient than segments becomes more tenuous. Firstly, Kuhl and Miller (1982) found that 1- to 4-month-old English language environment infants could detect a vowel change in the presence of an irrelevant pitch change, while they could not detect pitch changes in the presence of irrelevant vowel changes.

Secondly, Burnham and colleagues tested English-language children using a paradigm in which the identification of Thai tones was pitted against nonnative Thai consonants (Burnham & Francis, 1997; Burnham, Francis, & Webster, 1996), and Swedish pitch-accent was pitted against nonnative Swedish vowels (Burnham & Torstensson, 1995). They found that English-speaking children found nonnative consonant and nonnative vowel contrasts to be more perceptually salient than nonnative lexical tone or pitch-accent contrasts respectively, presumably because they have learnt that lexical level pitch related distinctions are not functionally relevant in their language environment. In contrast, English-speaking adults perceived nonnative tone and pitch-accent contrasts better than nonnative consonant and vowel contrasts respectively, and Burnham and colleagues argued that this was due to their ability to treat the experiment as a perceptual task such that the more acoustically salient F0 cues are used rather than the less salient segmental contrasts, and children’s inability to base their responses on psychoacoustic salience was due to their adoption of a language-specific mode of perception during the critical stages of language development. Both Kuhl and Miller (1982) and Burnham and colleagues (1996;1997) tested English-learning children so conclusions about the relative salience of consonant, vowels and tones may reflect language background and should not be generalised to how tone language speakers perceive tone.

3.6.4.3 Tonological Awareness versus Phonological Awareness

In one of the few studies of tonological awareness, Burnham, Davis et al. (2003) investigated the awareness of tones as independent speech units. Tonological awareness as they refer to it, is analogous to phonological awareness, the ability to use or manipulate information about phonemes of the spoken language (Wagner & Torgesen, 1987), for example the ability to add or delete phonemes to a word or
nonword. Testing tonological awareness assumes that tones can be segmented and treated as separate units autonomous to the lexical item to which they are attached. Burnham, Davis et al. (2003) examined the development of phonological and tonological awareness skills in Thai speaking children (5-, 7-, 9- and 11-year-olds) and adults, in relation to education level and reading ability. A control group of English subjects were also tested. The participants were asked to select the ‘odd one out’ of the three words, with the ‘odd one out’ being specified by a Phone, Tone, or a Tone + Phone difference. Understandably, for English language participants, tonological awareness tasks were more difficult than phonological awareness tasks and this difficulty was accentuated as a function of age. Suprisingly, the Thai children also experienced relatively greater difficulty with the tonological awareness tasks, and improved less over age on this task than the phonological awareness task. However Thai children did perform better on the Tone +Vowel task than on the Tone or Phone task, suggesting that the children had some implicit knowledge of tone. These results are in line with Abramson’s position (1978a) that the segmental representation of tones does not fit the grammar of Thai, and furthermore, that children learn tone in conjunction with the lexical item, rather than as separate items. With regard to this issue, tonological awareness was related to education level. The tonological awareness of primary educated adults was equivalent to the tonological awareness of 11-year old children, whereas tertiary educated adults had greater (near perfect) tonological awareness.

3.7 Infants’ Perception of Tone

There is a paucity of controlled experimental studies of lexical tone perception by infants. Yet in related areas, infants’ sensitivity to pure tones, pitch in music, and pitch in connected speech, have been studied quite extensively.

3.7.1 Threshold Sensitivity to Pure Tones

There has been a vast amount of research into infants’ perception of pitch. In one of the first published investigations of threshold sensitivity to pure tones by infants Olsho, Koch, Carter, Halpin, and Spetner (1988) tested 3-, 6- and 12-month-old infants and adults on their discrimination of 250, 500, 1000, 2000, 4000 and 8000 Hz pure tones.
Thresholds improved over age: For the youngest group, the 3-month-olds thresholds were 15-30 dB higher than those of the older (6- and 12-month-old) infants, and the older infants’ thresholds were 10-15 dB higher than the adult group.

In a more recent study, Werner and Baiko (2001) exposed infants aged 7 to 9 months, and adults with normal hearing, to four 500ms bursts of a computer-generated 1000 Hz tone and a 1000 Hz broadband noise. These were played at different levels of loudness in various signal-to-noise ratios or in quiet. It was found that on average infants are relatively better at detecting broadband noise than tones. The infant-adult difference for detection in quiet conditions was 14 dB for noise and 10 dB for quiet, and for the background noise condition, 7 dB for noise detection, and 5 dB for tone detection. An examination of the infants’ psychometric functions for detecting broadband noise shows that they have similar psychometric functions for broadband noise and tone detection, and do not attend more to broadband noise than to tone. Rather, the infants detect multiple simultaneous frequencies, that is, broadband, more easily than narrowband tones, probably because they have not yet developed the selective attention capacities of adults that are required for listening to narrowband sounds. Furthermore as broadband sounds are complexes of multiple frequencies, as is speech, and more specifically, lexical tone, then these sensitivities for detecting broadband noise and pure tones provides an important starting point – in order to perceive tone distinctions, detecting the presence of the tone sounds is necessary.

3.7.2 Pitch in Music

Infants are also able to discriminate pitch, and then find pitch contours particularly salient. For example, infants 5 to 11 months of age readily discriminate tone sequences or tunes differing in pitch contour (Trehub, Bull, & Thorpe, 1984; Trehub, Thorpe, & Morrongiello, 1985) and categorise tone sequences on the basis of pitch contour (Thorpe, 1986). Thorpe (1986) investigated whether infants could categorise variable two-tone sequences on the basis of rising and falling pitch. Infants of 7 to 10 months of age were trained to head turn when two-tone rising pitch sequences of different frequency intervals (6 and 2 semitones, 4 and 2 semitones, 2 and 1 semitones, plus a no-change interval) changed to a two-tone sequence with falling pitch. The infants
could successfully detect directional changes in pitch for all interval sizes, and performance was not influenced by the degree of frequency change, thus indicating that they responded to the two-tone sequences on the basis of rising or falling pitch only. Furthermore the infants could extract information about pitch even from intervals of one semitone, the smallest musically relevant frequency change in Western musical structure.

In a related study, Trainor and Trehub (1992) played repeated transpositions of a 10-note melody to Canadian 8-month-old infants and Canadian adults for their detection of changes to the melody. Adults found a 4-semitone within-key change to the melody significantly more difficult to detect than a 1-semitone outside-key change, while infants discriminated both changes to the melody equally well. Trainor and Trehub proposed that adults’ perception of melody is influenced by conventions associated with Western musical structure whereas infants have not yet incorporated these constraints. More recently Trehub, Schellenberg, and Kamenetsky (1999) have shown that Canadian infants can also readily detect subtle pitch changes in Western musical scales, but cannot detect fine pitch changes in a non-Western musical scale, suggesting that the influence of experience in musical tone perception may occur earlier than expected. However the inclusion of a control group of non-Western infants is required before strong conclusions can be drawn.

Saffran and Griepentrog (2001) have shown that 8-month-old infants better discriminate sequences of bell-like tones that differ in absolute pitch than relative pitch, whereas adults do the opposite, discriminating changes in relative pitch significantly better than changes in absolute pitch. This suggests that pitch perception may be absolute early in life, and with experience the listener learns to categorise sounds in terms of relative pitch, maybe because absolute pitch is too detailed to be functionally relevant (see discussion of classification, 2.3.1, and categorical perception, 2.4 and 3.6.3). Saffran and Griepentrog (2001) argue that absolute pitch also has important implications for learning tone languages because absolute pitch is required to understand the subtle differences between similar sounding words. Indeed, Deutsch and colleagues have data to suggest that tone language speakers demonstrate absolute pitch in their native language speech production (Deutsch, Henthorn, & Dolson, 1999).
In their first experiment, seven Vietnamese speakers read aloud a list of ten Vietnamese words (at a rate of one word every two seconds) in two test sessions across different days and it was found that the average F0 of each word was reliably similar, with averaged absolute pitch differences less than 1.1 semitone for all seven speakers and only .5 of a semitone for four of the seven speakers. In the second experiment 15 Mandarin speakers read aloud 12 Mandarin words twice, with a 20-second interval between readings, and two different days. Average pitch differences in enunciation both between and across sessions were not significantly different - less than .5 of a semitone for the majority of speakers, with five of 15 speakers averaging pitch differences less than .25 of a semitone. This suggests that the Mandarin speakers possess stable absolute pitch templates in word enunciation. In a third experiment with English speakers reading aloud a list of English words, absolute pitch consistencies were significantly greater within a test session than across test sessions. Deutsch, Henthorn, and Dolson (2004) argue that the differences between tone and non-tone language speakers’ pitch consistency across sessions is evidence that they are processing pitch in qualitatively different ways, but that English speaker’s possess absolute pitch under certain circumstances (i.e., for within-session repetition of items). Deutsch and colleagues surmise that the qualitative difference between Mandarin and English speakers is due to the Mandarin speakers’ learning of pitch-to-verbal label associations in infancy.

In a similar study Peretz, Burnham, Schwanhäußer, Tsukada, and Bollwerk (2003) investigated whether there is absolute pitch in the productions of Vietnamese and English speakers when order (fixed versus random) and stimulus onset timing of words (fixed 2 second intervals versus varying 3-5 second intervals) are manipulated, and subjects are tested on two different days. They found a small degree of variation in pitch, less than or equal to 1 semitone, between the first and second testing day, with slightly less variation in pitch across days for the Vietnamese compared to the English speakers in both the varying and the fixed stimulus onset conditions, and a greater difference in pitch across language groups evident in the varying condition. Thus the English speakers to an extent demonstrated absolute pitch – less than 1 semitone difference across sessions and more reliable absolute pitch enunciation in the fixed condition. Peretz et al. (2003) state that if this is absolute pitch, then both tone and non-
tone language speakers possess it. Yet while their results concur with Deutsch et al. (2004), Peretz et al. (2003) mount a different argument. Whereas Deutsch claims that pitch for speech and pitch for music share common mechanisms, and that speech affects musicality, or absolute pitch, Peretz and colleagues propose the opposite - that musicality affects tone language perception. Indeed, Burnham and Brooker (2002) have shown that musicians with absolute pitch are better at discriminating lexical tones than musicians without absolute pitch, who are in turn better at discriminating lexical tones than non-musicians.

Clearly, infants are able to discriminate pitch differences in musical stimuli (e.g., Trehub et al., 1984; Trehub et al., 1985), and are more sensitive to absolute rather than relative pitch transitions (Saffran & Griepentrog, 2001). However, whether infants’ initial perceptual sensitivity to pitch is absolute and only maintained if a tone language is acquired or musical training follows, is difficult to conclude. Indeed it cannot even be definitively concluded whether tone language speaking adults have absolute pitch perception, because Deutsch and colleagues (1999; 2004) and Peretz et al. (2003) investigated tone language production only. Of further interest might be to determine if tone language speakers who speak English as a second language, show greater absolute pitch in their production of English words than native English speakers.

3.7.3 Pitch in Speech

The pitch contours of speech hold special significance for infants. From at least one month of age, infants attend preferentially to the highly modulated infant-directed speech over adult-directed speech (Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod, 1989; see also chapter 2, section 2.5.4). The heightening of F0 and broader F0 range characteristic of IDS is even present in speech of tone language mothers, such as Mandarin (Grieser & Kuhl, 1988) and Thai (Kitamura, Thanavisuth, Burnham, & Luksaneeyanawin, 2002). These findings are particularly interesting given that tone languages use fluctuations in F0 to signal phonemic distinctions and the pitch fluctuations of IDS could easily disrupt tone information and compromise the phonemic message. However it is noteworthy that in their study, Kitamura et al. (2002) had Thai phonologists rate the integrity of lexical tone information in the IDS utterances and they
found that tones were relatively identifiable (74.5%), although less identifiable than
tones in ADS utterances. Note also in this regard that Thai tones retain their integrity
over intonational modulation (see 3.4.2.1).

An IDS study by Fernald and Kuhl (1987) is particularly relevant here. They used sine
wave tracks of the fundamental frequencies of the IDS and ADS speech recordings, and
found that 4-month-old infants’ selective preference for IDS is based on the pitch of
IDS rather than amplitude or durational characteristics. However, they did not test
infants’ relative preference for sine wave IDS versus natural IDS, so it is unclear
whether pitch contour alone was sufficient to evoke this preference. In contrast,
Kitamura and Burnham (1998) have shown that infants’ preferences for IDS are related
more to affect than pitch (see chapter 2, section 2.5.4).

The prosody of IDS is also argued to aid in speech discrimination. The prosodic
bootstraping hypothesis has been proposed to account for the acquisition of some of
the structural properties of language (Morgan & Demuth, 1996). One study by Karzon
(1985) showed that 1- to 4-month-old infants could discriminate a change in the second
syllable of a three-syllable utterance (‘malana’ vs. ‘marana’) only when the second
syllable had the pitch contour, duration, and intensity of IDS. In relation to pitch
though, infants are able to discriminate two words whose final vowels differ in
fundamental frequency contour (Karzon & Nicholas, 1989) and discriminate vowels /i/
and /I/ with IDS pitch contours (Trainor & Desjardins, 2002).

Nazzi, Floccia, and Bertoncini (1998) investigated whether newborn infants are
sensitive to the pitch contour of words and whether they can extract this information
from lists of words, rather than from the presentation of two stimuli. Using the high
amplitude sucking procedure (see section 2.1.1), it was found that French newborn
infants were able to discriminate two lists of Japanese phonetically-varied bisyllabic
words that differed in pitch contour (ascending versus descending). This finding
extends those of Karzon (1985) and Karzon and Nicholas (1989), by suggesting that the
infants’ sensitivity to pitch variations is an early and basic (language-general) ability.

---

41 Japanese is a pitch-accent language. Pitch contour is lexically specified in this language resulting from
the succession of high- and low-pitch morae, the rhythmic unit of spoken Japanese.
Young infants are particularly sensitive to pitch characteristics in speech. This may in part be determined by their preference for IDS over ADS and the heightened attention to suprasegmental aspects of native language speech. There is also some evidence that the F0 features typical of IDS may facilitate the discrimination of speech contrasts. Despite this knowledge, the manner in which pitch contour in the form of lexical tone is discriminated by infants is relatively unexplored. Lexical tone perception has been relatively forgotten by infant language researchers. Only a few studies have attempted to investigate this type of speech contrast with infants but the results so far do not provide a comprehensive account of lexical tone discrimination. The available evidence on lexical tone discrimination by infants is the focus of the next section.

3.7.4 Lexical Tone

One of the earliest studies of infant speech perception included an investigation of the discrimination of intonation. In one of four conditions Morse (1972) used a synthesised syllable [ba] with the fundamental frequency contour either rising from 120 Hz to 194 Hz or falling from 120 Hz to 70 Hz. These F0 changes occurred in the last 150ms of the 500ms stimuli, and were accompanied by parallel drops in amplitude. Six\(^{42}\) infants aged between 40 and 54 days were tested using the HAS procedure. Post-shift sucking functions showed that these very young infants discriminated the changes in F0. While the purpose of the study was to assess infants’ ability to discriminate intonation contours, for two reasons these intonation contours could equally be thought to be similar to lexical tone contours. Firstly, the change of F0 contour was presented in the final 150ms of the stimulus so that it was carried on the vowel of the [ba] syllable. If these stimuli were truly intonational then the pitch variation should have been a component of the structure of larger syntactic units (Anderson, 1978) and not of one-syllable stimuli. Secondly, the magnitude of F0 differences 74 Hz and 50 Hz for rising and falling respectively, are analogous to the average F0 change in the falling (73 Hz) and rising (48.5 Hz) Thai tone contours produced by Thai native speakers (Erickson,

\(^{42}\) As well as the falling versus rising intonation contrast \((n = 6)\), Morse (1972) tested infants for discrimination of a place of articulation contrast \((n = 6)\), control \((n = 4)\), and nonspeech place of articulation contrast \((n = 9)\). Only results pertaining to the discrimination of the intonation contours are reported here.


LEXICAL TONE

1974). A major difference between the Morse (1972) and Erickson (1974) stimuli was that the onset to mid F0 transitions of the synthetic intonation stimuli were flat in the Morse’s study but variable in Erickson’s. Nevertheless, the pitch of the Morse (1972) intonation stimuli are very similar to Thai lexical tone contours, so the Morse data provide good evidence that infants may be able to discriminate lexical tone-like stimuli.

There is only one known study of infants’ discrimination of lexical tone. This study by Harrison (2000) concerns the discrimination of tone in Yoruba, a Benue-Congoid language spoken in Western Nigeria and Benin. Yoruba has three level tones - high, mid, and low. Following a pilot study with adults, Harrison (2000)\(^3\) tested 6- to 8-month-old infants from Yoruba and English speaking families on the discrimination of a continuum of Yoruba lexical tones in order to determine whether infants’ perception of tone is (a) categorical (see section 3.6.3), (b) prototypical (see section 2.8.1.3), or (c) dependent completely upon the pitch relationships of an utterance. Natural tokens of the Yoruba words [ki-high] (greet), [ki-mid] (thick), and [ki-low] (praise) were synthesised so that the only distinction between the words was in pitch. Infants were tested using the Visually Reinforced Infant Speech Discrimination (VRISD)\(^4\) paradigm. Infants were trained in one of two conditions, ba-low vs. ki-high contrast, and pitch-only stimuli, (synthesised [ki]) at 175 Hz vs. 225 Hz. Thus there was a difference of 50Hz between the pitch-level of each tone in the pair.

Following training, infants were tested on similar pitch-only pairs of stimuli but with a 20 Hz pitch differential (140-160, 160-180, 170-190, 180-200, 190-210, 210-230, 220-240). Harrison (2000) found that the Yoruba language environment infants responded to the isolated syllables of different pitch but their English counterparts did not. Furthermore there appeared to be language-specific influences facilitating Yoruba infants’ discrimination of tone pairs, biasing them towards best discriminating between the 190 and 210 Hz tokens, a point in the frequency continuum that corresponds to a phonological tone boundary in Yoruba\(^5\).

\(^3\) Parts of Harrison (2000) were published in earlier papers. The pilot data with adults is reported in Harrison (1996) and the infant data is presented in Harrison (1998).

\(^4\) The VRISD is similar to the CHT procedure, see Eilers, Gavin, and Wilson (1979).

\(^5\) In the pilot study, Yoruba speaking adults showed a similar boundary. The adults identified stimuli above 210 Hz as ki-high (greet), and stimuli below 190 Hz as ki-mid (thick).
Despite these promising findings there are various methodological shortcomings in Harrison's research and some limitations on the conclusions that were drawn. Firstly, Harrison assumed that the Yoruba infants knew\(^{46}\) they were listening to speech due to their experience with hearing Yoruba. However, Yoruba infants typically hear native language tones in natural settings and the tones used in this study were synthetically produced. An alternative possibility is that the tones did not sound like speech to the Yoruba infants but that their tone language experience allowed them to use the cues available for discrimination more effectively than non-tone language environment infants.

Secondly, with regard to participant numbers only six infants from each language background participated \(N = 12\), with useable data for only 10 infants and from the point of view of infant testing (e.g., infants can become fussy, inattentive) and statistical power, this is an extremely small sample size. Thirdly, Harrison's training stage was not very successful. Ten infants were trained on the naturally produced real-word [ba-low] vs. [ki-high] Yoruba contrast and nine were able to discriminate the stimuli that differed across multiple cues. The infant that failed was from an English language environment. When Harrison trained three infants from each language group on the pitch-only distinction (50 Hz differential), only two infants, both Yoruba, could discriminate these tone contrasts. Although only two of six infants successfully passed training on the pitch-only distinction at 50 Hz differential, Harrison followed by testing all infants on a finer pitch differential of 20 Hz. Naturally the infants (especially those who had not passed, it might be thought), had difficulty discriminating these contrasts. Therefore of all the stimulus pairs only the phonologically relevant pair, 190-210 Hz, was discriminated at a statistically significant level, and then by only three of the 12 infants, all from Yoruba speaking families. This is a relatively good indication that this is a language-specific phenomenon but requires stronger statistical proof.

A subsidiary problem lies in Harrison's assertion that because infants discriminated the phonologically relevant pair, they perceived tone categorically. By definition, categorical perception requires both discrimination and identification functions for the

\(^{46}\) This was the expression used by Harrison (2000).
stimulus continuum. Harrison (2000) has the former but not the latter. In addition, better discrimination of this phonologically-relevant contrast may have been due to a methodological artefact. Harrison (2000) did not counterbalance the order of presentation for the different tone pairs. Rather, based on a pilot investigation, in which the 190-210 Hz tone contrast was best discriminated, infants were tested on this contrast early in the test phase (as the first, second, or third of 6 pairs tested) to decrease the likelihood that the infants were fatigued or distracted on this test trial. However, fatigue effects may of course have influenced the infants’ ability to discriminate the contrasts presented later in the test phase, thus giving the erroneous impression that the 190-210 Hz pair was better discriminated.

The experimental studies in this thesis investigate the discrimination of lexical tone by infants as a function of language background, linguistic context, and age. The experiments are designed to be as methodologically as sound as possible and to provide a clear, comprehensive account of infants’ ability to discriminate lexical tone. Following from the critical issues discussed here, infants from tone and non-tone language backgrounds are tested, both lexical tone and non-linguistic pitch stimuli are used, and age differences in performance within the important period of 6 to 9 months are investigated.
CHAPTER 4
Experimental Rationale, Framework, Hypotheses, and General Method

‘Though this be madness yet there is method in it’ (‘Hamlet’, W. Shakespeare)
In this chapter the studies that are the focus of this thesis are introduced. The chapter begins with a rationale for the experiments and their place in the infant speech perception literature, and this is followed by an outline of the studies and methods used. The original plan was to conduct the experiments in two locations - MARCS Auditory Laboratories, University of Western Sydney, Australia (testing of English-learning infants), and the Division of Speech and Hearing Sciences, University of Hong Kong, China (testing of Chinese-learning infants). Unfortunately, the outbreak of Severe Acute Respiratory Syndrome (SARS) in 2003 prevented the candidate from returning to Hong Kong to test the Chinese infants recruited for participation, and once the outbreak was contained, only three infants were of the appropriate age for participation. Consequently, it was decided to test the majority of Chinese infants in Sydney Australia, where there is a large Chinese population (see Appendix A1 for details of recruitment procedures and strategies in Australia and Hong Kong).

4.1 The Need for Research on Lexical Tone Discrimination by Infants

There are three main lacunae in speech perception research that will be filled by the investigation of lexical tone discrimination in infancy:

1. *Language development for non-tone but not tone languages*: Over half the world's population speak a tone language, however there is little knowledge of how infants from tone language speaking environments acquire the ability to discriminate lexical level utterances.

2. *Development of consonant and vowel but not tone perception*: Studies of speech perception development have focused on the discrimination of various consonant and vowel contrasts and have found that early speech discrimination abilities are language-general and become more language-specific between approximately 6 months and 12 months of age (see section 2.7). In contrast, the development of lexical tone perception is understudied, and it is unknown whether the developmental course of speech perception for lexical tone is akin
to that of consonant and vowel discrimination or if it follows a different course altogether.

3. *Adults and children but not infants*: There is a large amount of cross-language research into lexical tone discrimination by adults and children (see section 3.5.2.1) but no comprehensive account of what happens earlier in speech perception development for infants raised in either in tone or non-tone language speaking environments.

Unmistakably, there is a void in the infant speech perception literature regarding the development of lexical tone perception by infants. In this thesis this line of enquiry is pursued.

4.1.1 **Framework for Experiments: Perceptual Reorganisation for Speech**

The developmental speech perception literature is replete with studies demonstrating that infants are superior to adults in the discrimination of nonnative speech sound contrasts (for reviews see Burnham, 1986; Maye, 2002; Werker & Tees, 1999). In a long series of studies beginning with Werker, Gilbert, Humphrey, and Tees (1981), it has been established that during the second half of the first postpartum year and before infants begin to speak, the initial language-general pattern of speech sound discrimination becomes more language-specific and adult-like (refer to section 2.7 for details of other related research).

The experiments in this thesis are derived from a perceptual reorganisation orientation, and this has theoretical implications for the hypotheses (detailed later in section 4.1.5). Some aspects and expressions of the perceptual reorganisation notion are stated below:

1. Perceptual reorganisation occurs when infants begin to attend to word meaning (Best, 1995b; Lalonde & Werker, 1995).

2. Perceptual reorganisation occurs on the basis of mere exposure to the ambient phonology (Jusczyk et al., 1990; Maye et al., 2002).
3. Perceptual reorganisation occurs earlier and more clearly for nonnative *fragile* contrasts, that is, sounds that are less common in the world's languages, and less psychoacoustically salient than robust (more common, more salient) contrasts (Burnham, 1986).

4. Perceptual reorganisation may be related to general cognitive development (Lalonde & Werker, 1995).

5. Perceptual reorganisation may take place when the infant begins to *produce* speech sounds (Vihman, 1996).

It has been suggested that perceptual reorganisation favouring ambient language characteristics not only occurs for segments (consonants and vowels) but also for the suprasegmental features of the language, namely prosody (Hirsh-Pasek et al., 1987), and the stress patterns of words (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk & Kemler Nelson, 1996). However, as yet, there is no similar body of evidence regarding lexical tone. Cutler and Mehler (1993) suggest that infants have a *periodicity bias* to attend to the smallest level of rhythmic regularity in the ambient language. If this is so, infants should discover whether intonation, rhythm, stress, and also tone are used in the ambient language environment, and attend to their relevant regularities and ignore irrelevant variation. Thus infant learners of a tone language should attend to lexical tone, while infants from non-tone language environments should attend less or ignore tone. However, while intonation, rhythm, and stress assist in word segmentation it is possible that tone does. Thus unlike intonation, rhythm and stress, there appear to be no *regularities* in tone that allow the listener to split the speech stream into words. Nevertheless, it may be suggested that tone-learning infants' consistent exposure and experience with lexical tone signalling distinct lexical items in the ambient language environment, will result in their greater attention to lexical tone in speech than their non-tone language environment counterparts.
The experiments in this thesis are the first known studies specifically investigating infants' discrimination of lexical tone contrasts across age. The experiments provide a systematic programme of enquiry into the perceptual reorganisation of tone contrasts, which allows comparison with the perceptual reorganisation for segmental speech perception. More specifically the experiments in this thesis are designed to investigate whether discrimination of lexical tone contrasts is language-general at 6 months regardless of language background, and whether such discrimination is maintained (or facilitated) in tone language (Chinese\textsuperscript{47}) infants and attenuated in non-tone (Australian-English) infants.

4.1.2 Language Background, Age of Participants, and Stimulus Selection

Generally studies concerned with perceptual reorganisation have shown a decline in discrimination of nonnative speech contrasts over infant development. Many of these studies have included tests of how these contrasts are perceived by native language subjects, mostly adults (see Werker & Tees, 1999), though few have also included tests of native language infants (the inclusion of Hindi and Salish infants by Werker and Tees, 1984a is one exception). While the results of these studies strongly suggest perceptual reorganisation, ideally both Ontogenetic and Differential Experience methods should be incorporated, so native language listeners of the same ages should be tested. Now that greater correspondence and mobility between labs and countries is possible, combined method studies are more feasible. A combination of the Ontogenetic and Differential Experience method is be used here.

Infants are drawn from two different linguistic environments, English and Chinese (Differential Experience Method), for which lexical tone is nonnative and native respectively. The English-learning infants were recruited from families in Sydney, Australia in which Australian English was the sole language spoken, and the Chinese language-learning infants were recruited from Cantonese and/or Mandarin (both tone

\textsuperscript{47} Because Cantonese and Mandarin are Chinese languages/dialects and the infant participants are from one or both of these language backgrounds, hereon for simplicity this group of infants will be referred to as Chinese-learning infants, hereafter.
languages\textsuperscript{48} speaking families in Hong Kong, China, and Sydney, Australia. As the stimulus materials in the experiments were from the Thai language these Chinese infants have experience with a tone language but not the particular stimulus language. Via this manipulation it is expected that lexical tone discrimination would not be biased by familiarity with the language because these are not native language tones.

Both 6-month-olds (6.0 months to 6.5 months) and 9-month-old infants (9.0 to 9.5 months) participated in the experiments (Ontogenetic Method). The reorganisation of speech perception is thought to begin for vowels around 4 months of age, and consonants around 7 months of age (Jusczyk, 1998; Werker & Polka, 1993a;1993b; Werker & Tees, 1999), and so the younger age here, 6 months, was chosen to be between these two ages. The older age, 9 months, was selected because it is towards the upper end of the age continuum for the reorganisation of speech perception (see Werker and Tees, 1999 for a review).

Across the series of experiments two different sets of stimuli were used, speech and nonspeech. Infants at both ages and from both language environments were tested on both speech stimuli and equivalent nonspeech stimuli in order to examine the specificity of reorganisation for speech perception per se. The lexical tone speech stimuli were recorded by a native female Thai speaker from Chulalongkorn University, Thailand,\textsuperscript{49} and are exemplars of three voiced bilabial stop plus [a] vowel syllables [ba] (phonemic in Cantonese, Mandarin and English) with rising, falling, or low Thai tone contours (more details to follow in section 5.2.3). These stimuli were used by Burnham, Francis, and Webster (1996) to test discrimination of Thai lexical tone contrasts by English-speaking and Thai-speaking children and adults.

For the nonspeech sounds the F0 contours of the lexical tones were extracted and used to combine the F0 of the lexical tones with the keyboard-generated violin sounds.

\textsuperscript{48} None of the Chinese-learning infants had experience with Thai (the stimulus materials language), but many had exposure to English and other tone languages including other Chinese languages/dialects.
\textsuperscript{49} The assistance of Dr. Sudaporn Luksaneeyanawin, Director of the Centre for Research in Speech and Language Development at Chulalongkorn University, Bangkok, Thailand, who recorded the lexical tone speech stimuli, is greatly appreciated.
Details of the stimuli and procedure for resynthesis are provided in section 5.5.3 (see also Appendix A2).

4.1.3 Task and Method

Infants were tested on a simple lexical tone discrimination task of [bā] vs. [bâ] or [bã] vs. [bà] and in separate experiments the nonspeech analogue of these tone contrasts, rising vs. falling, and rising vs. low. Based on the findings of Burnham and Francis (1997) with English-speaking children, the lexical tone contrast [bā] vs. [bâ] was chosen as it was found to be relatively easy to discriminate in relation to the other possible tone pairings, while the [bã] vs. [bà] contrast was chosen as it was found to be relatively more difficult to discriminate. The similar starting F0 values of the Thai rising and low tones contributes to this contrast being more difficult to discriminate (see Abramson, 1962 and Figure 3.1, section 3.3.1). These two contrasts [bā] vs. [bâ] and [bã] vs. [bà] and their nonspeech counterparts, rising vs. falling, and rising vs. low, will henceforth be labelled as the Easy and Difficult contrast respectively.

The CHT procedure, typically used for speech research (e.g., Kuhl, 1980, 1983, 1985; Werker & Lalonde, 1988; Werker & Tees, 1984a) was the method chosen for testing tone discrimination. The CHT procedure is suited to testing infants between 5 ½ and 18 months of age (Polka, Jusczyk, & Rvachew, 1995), and hence is appropriate for the age range of infants in these experiments. Infants younger than 6 months have difficulty producing head-turns and the procedure is too lengthy for them (Polka et al., 1995), while infants over 10 months are easily bored by the reinforcer and are more mobile and therefore less content to sit quietly on the parent’s lap for long periods of time (Werker et al., 1998).
4.1.4  Aim and Purpose of Studies

Given young infants' ability to discriminate nonnative consonant and vowel contrasts and older infants' more language-specific speech perception\(^{50}\), an important question arises: What is the developmental course of tone perception; does tone perception change as a function of age and linguistic experience?

In addressing this question, the experiments here take a *cross-age, cross-language, cross-linguistic context* approach in order to investigate the development of lexical tone and nonspeech pitch perception in 6- and 9-month-old English and Chinese language-learning infants. The main aim of the experiments is to determine whether the discrimination of lexical tone is language specific (Chinese vs. English), speech specific (speech vs. nonspeech), and at what age it may become either or both language and speech specific. The outcomes of these experiments should allow conclusions to be drawn about whether perceptual capacities for lexical tone develop in the same way as do those for segmental speech perception, namely with a reorganisation from an initial *language-general* to more *language-specific* speech perception (Werker, 1994; Werker & Lalonde, 1988; Werker & Polka, 1993; Werker & Tees, 1984a; 1992).

4.1.5  General Hypotheses

The design of the set of experiments allows the formulation of a number of hypotheses specific to changes over age, changes over language background, and with the inclusion of speech and nonspeech stimuli, changes over stimulus type. In line with the perceptual reorganisation literature, it is hypothesised that the English-learning infants will show an age-related decline in discrimination performance for nonnative tone speech contrasts - English-learning infants should perform significantly better on the lexical tone discrimination task at 6 months than at 9 months. On the other hand it is hypothesised that the Chinese-learning infants, for whom tone contrasts are phonologically relevant, will not show an age-related decline in discrimination performance for lexical tone. In fact, it is possible that the discrimination of lexical tone will improve between 6 and 9 months of age as a function of tone language

---

\(^{50}\) Note an exception, English-learning infants Zulu click discrimination performance does not decline over age Best, McRoberts, and Sithole (1988).
experience, as Polka et al., (2001) have found across age for English listeners perceptual discrimination of the native /d-ð/ consonant contrast. Additionally, consistent with results for older children (Burnham et al. 1996), the Easy tone contrast, [bã] vs. [bâ], is predicted to be easier to discriminate than the Difficult, [bã] vs. [bâ] contrast.

Previous research with nonspeech tone contrasts shows that non-tone language speakers find pitch-based distinctions more easily discriminable outside of a speech context (Burnham & Francis, 1997; Burnham et al., 1992; Burnham, Tyler, & Horlyck, 2002). Therefore with regard to the nonspeech tone discrimination task, it is hypothesised that English-learning infants’ discrimination performance will not decline between 6 and 9 months of age. It is further predicted that there will be no significant difference in English-learning infants’ discrimination of the nonspeech rising vs. falling (Easy) and the nonspeech rising vs. low (Difficult) contrasts, because the removal of the phonetic cues signalling speech should serve to highlight pitch information in these nonspeech stimuli such that pitch-based contexts, at least those within pitch discrimination thresholds (see section 3.7.1), should be discriminated. It is also hypothesised that for English-learning infants nonspeech tone discrimination will be better than lexical tone discrimination performance at both 6 and 9 months. Furthermore, based on the findings of Burnham and colleagues (1996) with tone language speakers’ (Cantonese and Thai), pitch-accent speakers’ (Swedish), and non-tone speakers’ (English) discrimination of speech tone, filtered speech tone, and violin tone contrasts (see section 3.6.2.2), it is predicted that Chinese-learning infants’ discrimination of pitch variations in nonspeech will be facilitated due to their experience with tone in language, such that they will perform better on the nonspeech tone discrimination task than infants from English speaking families.
4.2 Method

4.2.1 Participants

The participants were 6-month-old and 9-month-old infants from Australian English-language and Chinese language backgrounds. Specific demographics including age and gender of infants are reported for each experiment in chapters 5, 6, and 8. English-learning infants were recruited from families in Sydney, Australia, on the basis of details recorded in the MARCS Baby Register database. In addition parents/guardians of infants were asked to fill out an Infant Details form (see Appendix CD4 on the CD supplement) to ascertain language exposure, birth age status (that the infant was born within 2 weeks of full term i.e., between 38-42 weeks gestation), and hearing status (no known hearing disabilities\(^5\)). Infants of Chinese language backgrounds were initially recruited from Hong Kong, whose official language is Cantonese, although Mandarin (the official language of Mainland China) and English are also widely spoken\(^6\). The parents/guardians of the Chinese-learning infants were asked to fill out a questionnaire on their infants’ language exposure (see Appendix CD4 on the CD supplement). The answers were important for determining the degree of experience that infants had with Cantonese or Mandarin and any other language(s). All participants were given $20 AUD to cover travel expenses, a selection of baby products from our sponsors (see Appendix A1 for details of product donation), and for the infant, a bib or t-shirt with the ‘I’ve been to Uni, you can go too!’ motif, and a Young Scientist Award (refer to Appendix A1 for full details of participant reimbursement).

4.2.2 Design

The experimental design and its justification are outlined below. Specific details of each experiment are given in the Method section of chapter 5 (section 5.2).

---

\(^5\) It was not until 2002 that it was compulsory for newborns to be given a hearing test in Australia. Testing of the English-language group commenced in October 2001 and it is possible that some infants suffered from undetected hearing disabilities.

\(^6\) As Hong Kong is no longer under British rule, English is being taught less in schools, although the teaching language of some Universities, including the University of Hong Kong, is English.
4.2.2.1 Experimental design

There are eight experiments. Of these, four are cross-sectional studies as follows: Experiment 1, English-learning infants’ lexical tone discrimination \((N = 48)\); Experiment 2, English-learning infants’ nonspeech tone discrimination \((N = 48)\); Experiment 3, Chinese-learning infants’ lexical tone discrimination \((N = 48)\); and Experiment 4, Chinese-learning infants’ nonspeech tone discrimination \((N = 40)\). There are also four longitudinal studies, English-learning infants’ discrimination of lexical tone (Experiment 5, \(N = 12\)) and nonspeech tone (Experiment 6, \(N = 11\)), and Chinese-learning infants’ discrimination of lexical tone (Experiment 7, \(N = 10\)) and nonspeech tone (Experiment 8, \(N = 11\)). Each study employed a 2 x 2 design. The first factor is Age with 2 levels, 6 months and 9 months. The second factor refers to the discrimination condition, that is, the Easy contrast or Difficult contrast condition. Difficulty was always a between-subjects factor. Age was a between-subjects factor in Experiments 1 to 4 and a within subjects factor in Experiments 5 to 8. Within the Easy and Difficult conditions, the tone that served as the background and target sound of a contrast was counterbalanced.

In addition to the basic Age and Difficulty factors within each experiment, the four cross-sectional and four longitudinal studies also form what could be called a 2 x 2 design, with language background (English and Chinese) and task (speech and nonspeech) as the across experiment independent variables. The dependent variables in all studies were the proportion of infants reaching criterion for discrimination and percent correct discrimination.

4.2.3 Stimuli

The lexical tone contrasts were [bā] vs. [bâ] (Easy contrast) and [bâ] vs. [bā] (Difficult contrast). The nonspeech stimuli were synthesised analogues of the Thai lexical tones. A pitch track of each exemplar was made using the program STRAIGHT (Kawahara, Katayose, de Cheveigne, & Patterson, 1999), and the program APT (Haszard Morris, Stainsby, Malloch, & Burnham, 2002) was used to remove the phonetic information from the speech signal and resynthesise the speech onto a
keyboard-generated violin sound. Further details of the stimuli are given in section 5.2.3 for lexical tone and section 5.5.3 for nonspeech tone.

4.2.4 Procedure

A one-experimenter Windows version of the traditional conditioned head-turn (CHT) procedure was used. The traditional procedure is described ahead of the Windows version.

4.2.4.1 Traditional Conditioned Head-turn Procedure

The traditional Conditioned Head-turn (CHT) procedure is conducted by two experimenters and is primarily used for assessing auditory perception and speech perception in infants (Kuhl, 1985; see also Polka et al., 1995 for a review). The infant is seated on the caregiver’s lap, opposite an experimental assistant who displays toys to the infant. A loudspeaker and visual reinforcer box are co-located to one side of the infant. Speech stimuli are presented through the loudspeaker. The infant is taught to turn their head towards the reinforcer (mechanical toys in a plexiglass box) when they hear a change in the speech stimulus. If the infant makes a correct head-turn the box is illuminated and the toy becomes animated. The experimental assistant also smiles and praises the infant for producing a correct head-turn. No reinforcement is given for incorrect head-turns.

The mother, infant, and an experimental assistant (E1) are seated in a sound-attenuated room, while a second experimental assistant (E2) controls the computer and observes the infant either through a one-way mirror, peephole, or on real-time video. E1 and the parent listen to masking music and masking sounds via headphones so that they cannot hear the stimuli being delivered to the infant and therefore cannot influence their child’s performance in any way. E2’s task is to observe the infant carefully and initiate a trial by pressing a button whenever the infant is watching E1 (but not totally engaged by the toys being manipulated by E1). Then, under computer program control, either a change trial, in which the speech stimulus changes, or a control trial, in which there is no change in the speech stimulus is presented. E2 monitors the infant’s behaviour and pushes a button if a head-turn occurs in the direction of the reinforcer box.
There are typically two stages in the procedure. Following an initial training stage in which the infant is familiarised with the reinforcer, a conditioning stage is given, in which every trial is a change trial. In this stage, E2 must gradually shape infants’ head-turn responses such that anticipatory head-turns toward the reinforcer occur when they hear a change in speech sound category. During the first trials, the reinforcer is activated immediately following the presentation of a target sound. Subsequently, a delay is gradually introduced to give the infant an opportunity to produce a head-turn prior to the reinforcement delivery. The final stage is a test stage that commences once the infant reaches a pre-set performance criterion in the conditioning phase. During the test stage change and control trials are presented in random order. E2, who is deaf to what trial type is being presented, monitors the infant’s behaviour and presses a button whenever the infant produces a head-turn to the reinforcer. If the button press occurs within a criterion window of 4-6 seconds (approximately 3 change stimulus presentations), the reinforcer is activated and E1 verbally praises the infant (‘Good boy/girl’), smiles, and claps. This is recorded as a ‘Hit’. If no head-turn occurs during a change trial, a ‘Miss’ is recorded. If no head-turn is made during a control trial, this is recorded as a ‘Correct Rejection’. If the infant turns his/her head during a control trial, this is a ‘False Positive’.

Infants seem to enjoy this procedure perhaps because they are in control of the reinforcer – the CHT procedure is fun, interactive, and interesting. Infants who are interested typically learn the association between the change in sound and the activation of the reinforcer very quickly, within about 15 trials.

One strength of the CHT procedure is that the stimulus and reinforcer are independent events (in contrast to them being intertwined as in the habituation procedure). A second strength of the CHT procedure is that each infant receives multiple test trials, making it relatively easy to evaluate whether an individual infant can detect a change in the stimulus.

Unfortunately, as with other infant procedures, this one also has its limitations. Variable attrition rates, from 5% to 50% are possible, and the method is not particularly
suitable for testing infants younger than about 6 months, nor for testing speech patterns longer than a syllable, such as multi-syllabic words and melodic patterns.

4.2.4.2 Windows Conditioned Head-turn Procedure

The Windows version of the Conditioned Head-turn procedure was originally designed and developed by Dr. Rachel Hayes and Prof. Alan Slater at the University of Exeter, and it is the procedure used for testing infants in the present studies (Exeter Hayes & Slater Windows Conditioned Headturn program, version 2, 1999).

The basic concept is the same as the traditional CHT procedure with the main difference being that there is only one experimenter involved in testing. As with the traditional version, the experimenter and parent listen to masking sounds through headphones so as not to influence infants’ responses or hear whether a change or control trial is being presented to the infant; and the single experimenter engages the infant with toys, uses hand buttons and foot pedals to initiate a trial when the infant’s gaze is centred, and records infant head-turns. The experimenter is unable to tell whether a particular trial is a change or control trial because the computer in the control room controls trial selection. If the experimenter presses the foot pedal to indicate that the infant has produced a head-turn and the visual reinforcer is activated, the experimenter then provides the infant with additional social reinforcement, for example ‘Clever Boy/Girl’, smiles, and clapping.

This version of CHT also has various stages that the infant progresses through, although these vary slightly to the traditional CHT. The first is the training stage, in which the reinforcer is automatically activated to a change in speech sound. A delay is gradually introduced to induce the infant to produce anticipatory head-turns for visual and social reinforcement when they hear a change in sound. All trials in the training stage are change trials.

The infant progresses to the Conditioning stage once they have produced three consecutive anticipatory head-turns. Both change and control trials are presented in the Conditioning stage. A criterion of 7 out of 8 correct responses to change and control
trials is the criterion for discrimination and progression to the next stage\textsuperscript{53}. Infants who fail to produce a head-turn in three consecutive change trials revert to retraining in which, upon activation of the next change trial, the reinforcer is automatically and simultaneously activated. This is to remind the infant of the desired response. The Test stage is similar to the Conditioning stage, but without the opportunity for retraining.

4.2.5 Laboratories and Equipment

The experiments were conducted at two laboratories, MARCS in Sydney, and the Infant Perception Laboratory in Hong Kong. In each there was a Control and a Test room. The control and test rooms are shown in Figure 4.1 for MARCS and Figure 4.2 for Hong Kong. General details are provided below with specific details provided in sections 5.2.4 and 6.2.4.

4.2.5.1 Testing Chambers

\textit{Control Room}

\textbf{BabyLab, MARCS Auditory Laboratories, Sydney:} A PC connected to an audio amplifier was located in the control room to control presentation of the stimuli. The PC was also used to record the results. The control room and test room were connected via a window. With the lighting dimmed in the control room, the window acted as a one-way mirror and the participants and experimenter in the test room were unable to see observers in the control room. The computer screen in the control room was angled away from the experimenter so as not to reveal the contents of the screen while testing was in process, and so the screen would not attract the infant’s attention if they happened to turn towards it during the test session. However, due to its illuminance the edge of the computer screen in the control room was just visible from the test room so the experimenter knew what stage of testing the infant was in, based on the colour of the screen in the different stages. See Figure 4.1 and photographs of the set-up in Appendix CD5, CD supplement).

\textsuperscript{53} Note that this was the criterion set by the candidate for the experiments in this thesis. This criterion is not fixed and can be adjusted to the preference of every experimenter.
Figure 4.1. Schematic diagram of the test and control rooms, MARCS BabyLab, University of Western Sydney
Rationale, Hypotheses & Method

Infant Perception Laboratory, University of Hong Kong: The set-up in Hong Kong was similar to that at MARCS in Sydney. The control room housed the PC and amplifier. The PC controlled presentation of the stimuli and recorded the results of each test session. As in the MARCS laboratory the computer screen was not visible from the test room except to reveal the colour of the screen in the different stages.

Test Room

BabyLab, MARCS Auditory Laboratories, Sydney: The test room contained two comfortable chairs (one for the parent and one for the experimenter), the audio loudspeaker (controlled by the PC and amplifier), visual reinforcer, audio tape player controlling the masking stimuli for parent and experimenter, toys to centre the infant’s gaze, a ‘stable table’\(^{54}\) for the toys, fairy wand, and button box. The parent with infant on their lap sat opposite the experimenter with their back to the control room window. Despite the lights in the control room being dimmed, the infant was positioned this way so that any movement in the control room could not be detected and distract the infant. A light-weight curtain was draped over the loudspeaker to prevent the infant from looking at the loudspeaker. Figure 4.1 is a schematic diagram of the test room and photographs of the test room are presented in Appendix CD 5 (CD supplement).

Infant Perception Laboratory, University of Hong Kong: The test room set-up in Hong Kong was almost identical to that at MARCS except that the control room window was to the left of the experimenter and to the right of the infant (rather than behind as in MARCS, see Figure 4.2). It was therefore imperative that there was no light or movement in the control room that could distract the infant. The loudspeaker and reinforcer were positioned to the left of the infant so that they were not near the control room window and also so that the infants tested in both Hong Kong and Sydney were required to produce a head-turn to the left. As in Sydney, a light-weight curtain covered the loudspeaker to prevent the infants from staring at the source of the sound during testing.

\(^{54}\) A ‘stable table’ is a flat tray with a cushioned underside. It serves as a table that sits on the lap of the user.
Figure 4.2. Schematic diagram of the test and control rooms, Infant Perception Laboratory, University of Hong Kong.

Equipment Wiring

The equipment was wired by an electrical engineer\textsuperscript{55}, and set up by the candidate and the MARCS lab manager\textsuperscript{56}. The wiring diagram is shown in Figure 4.3. Under experimenter control are two hand buttons and two foot pedals that each send signals to a button-box in the test room (see also Figures 4.1 and 4.2). The button box is connected to the games port of the PC in the control room such that the two hand buttons and two foot pedals function via a four button gamepad. The function of these four inputs is described as follows:

\begin{footnotesize}
\begin{itemize}
  \item [55] The assistance of John Fowler from the University of New South Wales with the circuit wiring and building of foot pedals, hand buttons, and electronic and junction boxes, is greatly appreciated.
  \item [56] The assistance of MARCS lab manager Colin Schoknecht with equipment wiring set-up in the Sydney lab, and the assistance of technicians Raymond Wu and Donald Chan with equipment wiring set-up in the Hong Kong lab, is greatly appreciated.
\end{itemize}
\end{footnotesize}
1. Pedal 1: The *start trial* pedal. The experimenter presses this pedal when the infant is centred and settled to initiate a trial.

2. Pedal 2: The *reinforcer* pedal. The experimenter presses this pedal when the infant produces a head-turn towards the visual reinforcer. If this head-turn coincides with the change in stimulus from the background to the target sound then the head-turn is correct and the reinforcer activated.

3. Button 1: The *start sounds* button. The experimenter presses this button to start the stimuli playing from the loudspeaker.

4. Button 2: The *reinforcement plus* button. The experimenter presses this button to extend the length of the reinforcement period.

A signal from the parallel port of the PC is sent to the button box and this activates the visual reinforcer. The circuit diagram is presented below in Figure 4.3. The presentation of the stimuli (saved as .wav files) is controlled via computer program and projected through the amplifier in the control room to a single loudspeaker in the test room. The audiotape playing the masking stimuli was controlled by a tape player that operated independently of the PC.

---

57 The pedal and button names are labels used by the candidate. The program developers may not use these labels.
Figure 4.3. Circuit diagram of the head-turn apparatus.

4.3 Experimenters and Experimental Training

There were two experimenters, one for Sydney testing, the candidate, and one in Hong Kong, a research assistant for this project. The candidate conducted the testing sessions for all the English-learning and Chinese-learning infants tested in Australia to maintain the reliability of judging a head-turn. The research assistant in Hong Kong tested three infants there. They were trained for testing as set out below.
4.3.1 Training the Experimenter and Testing in Sydney

At the MARCS laboratory it took a few months to perfect the running of the CHT Windows program\(^{58}\) and to practice testing infants using the CHT procedure. In the early stages of training, the candidate conducted sessions with adult volunteers - postgraduate students and staff of MARCS Auditory Laboratories. Approximately 15 mothers and their infants were then invited to participate in pilot sessions. The practice participants were infants from bilingual families or parents with older infants, (10 months or older) who were not suitable for the research project proper. The main purpose of the pilot sessions was to practise the task in action and to realise the potential problems that might arise during the testing session, for example, determining whether the infant produced a head-turn (a ‘head-turn’ is defined in section 5.2.5), the incidence of infant crying, and technical difficulties, as well as to develop confidence to overcome these problems. It was also important to practise meeting and greeting mothers and interacting with babies to ensure both mother and baby were happy and at ease for the testing session. Participants in these sessions received reimbursement for their participation as per section 4.2.1.

4.3.2 Training the Experimenter and Testing in Hong Kong

A research assistant at the University of Hong Kong was trained to test infants using the CHT and was briefed on the theoretical and empirical background of the project and the rationale for the series of experiments. The researcher provided one-to-one instruction for the research assistant on using the program interface and programming the stimuli and parameters for the experiments into the Windows CHT. The researcher also taught the assistant how to operate the hand buttons and foot pedals during testing and the assistant mastered their operation by testing the researcher in practise sessions. The experimental set-up and circuit wiring in Hong Kong were checked against photographs of the set-up in Sydney to ensure they were identical.

---

\(^{58}\) Thank you to Dr. Rachel Hayes and Professor Alan Slater from the University of Exeter, United Kingdom for guidance with trouble-shooting the Headturn v2 program, for advice on establishing the parameters of the experiments and for answering questions about the program. Their assistance is greatly appreciated.
The research assistant in Hong Kong also viewed the procedure in action via video recordings of test sessions with infants from Sydney. To perfect testing, the assistant tested postgraduate students from the Division of Speech and Hearing Sciences and progressed to testing Chinese infants in pilot sessions. As with the pilot sessions in Sydney, the purpose was to practice meeting and greeting mothers, interacting with babies, and increasing awareness of the potential difficulties that might arise during testing sessions and developing strategies to overcome these problems. The infants who participated in the pilot sessions were recruited from families who were friends of employees at the University. They were paid for their participation and the infant received a toy Koala.
CHAPTER 5

English-learning Infants' Discrimination of Tone
In this chapter two experiments are presented, both involving English language-learning infants. In Experiment 1 lexical tone stimuli were used, and in Experiment 2 nonspeech analogues of the lexical tones were used.

5.1 Experiment 1: English-learning Infants’ Discrimination of Lexical Tone Contrasts

In Experiment 1 English-language learning infants of 6 months and 9 months were tested using the CHT procedure for their ability to discriminate Easy or Difficult Thai lexical tone contrasts. It was expected that 6-month-olds should perform better on the task than 9-month-old infants. All testing sessions were held in the BabyLab at the MARCS Auditory Laboratories, Bankstown Campus, University of Western Sydney, Australia.

5.2 Method

5.2.1 Design

The study employed a 2 x 2 design with Age (6 month-olds, 9 month-olds) and Difficulty (Easy contrast, Difficult contrast) as between-subjects factors. The Easy contrast was [bâ] vs. [bâ] and the Difficult contrast was [bâ] vs. [bà], with allocation of background and target stimuli counterbalanced for each pair. Infants were randomly assigned to one of two discrimination conditions - Easy tone pair or Difficult tone pair - until a total of 12 infants per age group were tested in each condition. The proportion of infants reaching criterion and the percentage of correct responses (hits plus correct rejections divided by the number of trials) were the dependent variables.

5.2.2 Participants

A total of 48 infants, 24 6-month-olds (14 girls, 10 boys; \(M = 6.25\) months; range = 6.03-6.63 months, \(SD = .0167\)) and 24 9-month-olds (13 boys, 11 girls; \(M = 9.17\) months; range = 8.89-9.51 months, \(SD = .1710\)) participated in the study. All infants
ENGLISH INFANTS’ TONE DISCRIMINATION

were from monolingual Australian-English speaking families. The participants were recruited via newspaper advertisements circulated within the Sydney metropolitan area, shopping centre recruitment drives, or through pamphlet distribution to Early Childhood Centres and private hospitals in the Southern Sydney region (refer to Appendix A1 for more information). All infants were born within two weeks either side of full-term (38 - 42 weeks) with uncomplicated births, were judged to be healthy at the time of testing, and had no known hearing disabilities or a history of middle ear infections. An additional 24 infants were tested, but were not included in the final sample due to crying (2) or fussiness (3), experimenter error (12)\(^59\), failure to show responsiveness to the reinforcer (2), and failure to meet the language background requirements (5)\(^60\). Parental consent was obtained prior to testing in accordance with University of Western Sydney ethical guidelines for human participants (see Appendix CD6 to view the information hand-out and consent form). Parents were debriefed following the testing session and received remuneration of $20 AUD for travel expenses, a selection of baby products from the MARCS BabyLab sponsors, and a bib or t-shirt (for details see Appendix A1).

5.2.3 Stimuli

The experimental stimulus items were naturally produced bilabial stop consonant-vowel syllables [ba] (phonemically relevant in both English and Thai), spoken with rising, [bā], falling, [bâ] and low, [bà] Thai tones. The stimuli were originally created for use by Burnham, Francis, and Webster (1996) to test discrimination of Thai lexical tone contrasts by English-speaking and Thai-speaking children and adults. Permission to use these stimuli was obtained from the speaker and the first author. Five exemplars of each stimulus were used, and the exemplars of each tone were confirmed to be physically similar using the STRAIGHT program (Kawahara, Katayose, de Cheveigne,

\(^{59}\) In a pilot test with 12 infants, the criterion for discrimination was set at 4 out of 5 correct consecutive responses. Following inspection of the data and consultation with Professor Janet F. Werker, it was evident that the criterion of 4 out of 5 was too low. Data from the 12 infants were discarded and the criterion was changed to 7 out of 8 correct consecutive responses.

\(^{60}\) While parents were asked about languages spoken in the home prior to their infant being tested, there were a number of cases where the parents mentioned that they speak a second language to their infant occasionally or that a carer regularly minds their infant, and that this carer speaks a language other than English.
& Patterson, 1999)\textsuperscript{61}, on the basis of F0 height, F0 contour, and amplitude, and were segmented into individual sound files using Cool Edit 96 and saved to disk. The duration of the exemplars ranged from 621ms to 741ms for [bâ], \((M = 671.4, SD = 49.813)\), 627ms to 751ms for [bâ] \((M = 667.2, SD = 48.376)\) and 601ms to 695ms for [bâ] \((M = 648.4, SD = 39.564)\). These durations and the duration for consonant, periodic and prevoking segments of the tokens are specified in Appendix A2 (refer to Appendix CD-A2 on the CD supplement to hear the lexical tone stimuli). Neither stimulus duration nor any other characteristics of the speech sounds were modified so as to preserve natural phonetic variability and encourage linguistic rather than acoustic processing. Spectral distribution varied slightly for each tone. The falling tones showed the greatest spectral energy, followed by low tones, and rising tones showed the least spectral energy.

5.2.4 Materials and Apparatus

Testing was conducted in a sound attenuated test room (250cm x 225cm). In the adjacent control room (250cm x 298cm), an IBM, Intel Pentium MMX (Microsoft Windows 98 operating system) PC computer, with a Pine PCI digital wavetable sound card (Crystal 428, version 3.0) controlled the presentation of stimuli and the Exeter Hayes and Slater Conditioned Head-turn program (Headturn version 2). The stimuli were projected through a NAD C320 Stereo Integrated amplifier (40 watt output) to a single loudspeaker (Creative, Cambridge Sounds Inc SBS51) in the testing room. Above the loudspeaker was a visual reinforcer (a mechanical monkey that played the cymbals, (see Appendix CD5 on the CD supplement for a photograph of the reinforcer). enclosed in a wooden box (height 35cm, width 30cm, depth 25.5cm) with a perspex front. It was not possible to see inside the reinforcer box except when the box was illuminated during the reinforcement delivery period. The parent (usually the mother) held their infant on their lap and was seated in a comfortable chair so that the loudspeaker and the visual reinforcer were located at a 50-degree angle, 135cm to their left. The loudspeaker and reinforcer were to the right of the experimenter who was seated 182cm from and directly in front of the infant. Both the parent and the

\textsuperscript{61} Thank you to Hideki Kawahara for permission to use the STRAIGHT program for research purposes.
experimenter listened to an audiotape consisting of music and a track of jumbled experimental stimuli so that they would be deaf to the experimental conditions. On this masking tape, the experimental items were concatenated in Cool Edit 96 (using the Mix Paste function) with no silence between each token and mixed with the Beatles 'Revolver' album (with no silence between each track) into a mono sound file which was presented diotically through closed ear headphones (Koss UR-20). The experimenter and the parent heard these continuous streams of the music and experimental stimuli to mask what the infant was hearing and prevent them from influencing the infant's behaviour. Pinto, Fernald, McRoberts and Cole (1998) report that this method is the most effective means of masking sounds presented to the infant. The experimenter controlled two foot pedals (Pedal 1: start trial pedal and Pedal 2: reinforcer pedal) and two hand buttons (Button 1: start sounds button and Button 2: reinforcement plus button). Pedal 1, Button 1, and Button 2 fed into a button junction box, connected to the games port of the PC. Pedal 2 connected to the printer port via the button box (refer to chapter 4, Figure 4.3).

Activation of the start trial pedal signalled to the computer program that a trial should commence. Activation of the reinforcer pedal signalled to the computer that the infant has made a head-turn toward the reinforcer. Pressing the start sounds button started the background sounds playing. Pressing the reinforcement plus button increased the duration of reinforcement by 1000ms (i.e., from 4000ms to 5000ms) and was used if the experimenter decided the length of the reinforcement period was not salient enough to reward correct head-turns. This button was seldom used as the reinforcer engaged the majority of infants sufficiently. Pressing the reinforcer pedal (connected to the printer port via the button box) controlled activation of the reinforcer (see Chapter 4, Figure 4.3). Throughout experimental sessions, the experimenter maintained infant attention to the midline by manipulating toys. Toys were revealed from a 'hidey bag' one by one so that infants did not habituate to the one toy and lose interest. Toys were manipulated on a stable table positioned on the experimenter's lap. The experimenter used a fairy wand to direct attention towards the reinforcer at the appropriate time as part of the shaping process.
5.2.5 Procedure

Infants were tested individually (one session per infant) using a variation of the go/no-go conditioned head-turn procedure (CHT, for a detailed description of this procedure, see section 4.2.4.2). Testing sessions averaged 13 minutes duration (range 8-25 minutes). The infant was presented with one of the Thai tones in a pair, the background sound, repeated continuously at fixed intervals of 1000msec from a loudspeaker on the infant’s left. The infant was conditioned to turn their head away from the experimenter and towards the sound and the visual reinforcer when there was a change in speech sound from the background to the target sound, the other sound in the pair. A head-turn is defined as a movement of the head of at least 45 degrees from the midline to the left of the infant towards the reinforcer. The experimenter (who directly faces the infant) judges whether a head-turn occurs so it was vital that the experimenter’s gaze never left the infant’s face, particularly while manipulating the toys. The experimenter was required to centre the infant’s gaze properly, that is, maintain the infant’s attention towards themselves for at least 2 seconds before initiating a trial. If the infant’s centred gaze wandered just after a trial was initiated, the experimenter was still required to judge whether a head-turn was produced. For example, if a trial was initiated and then the infant was looking at the floor and then clearly looked up at the reinforcer following the commencement of a trial, this was classed as a head-turn. However, if the infant’s gaze was centred and a trial was commenced, and then the infant looked around the room, glancing towards the reinforcer during this period of movement, this was not judged as a head-turn. Correct head-turns to a change in stimulus were rewarded by the visual reinforcer and by social reinforcement from the experimenter, that is, “Clever .... (Child’s name)”, smiling, waving the fairy wand, and clapping. Incorrect head-turns (false positives) were not reinforced.

The procedure involves three stages through which the infant progresses in the following order: Training, Conditioning, and Experimental 1 stages. In each stage the experimenter initiated a trial by pressing one of the foot pedals when she judged the infant to be in a state of readiness, that is, looking at the experimenter rather than at the visual reinforcer or around the room, and was not fussing or did not appear to be babbling. As trial initiation was dependent on infant readiness, the number of repeated
presentations of the background stimulus before the start trial pedal was activated was variable in order to prevent a response pattern based on temporal cues.

5.2.5.1 Training Stage

In the training stage, all trials were change trials (i.e., the stimulus changed from the background stimulus to the target stimulus when a trial was signalled from the experimenter by pressing the start trial pedal). During the first two trials of this stage the reinforcer was automatically activated for 4000ms by the computer program simultaneous with the change in speech sound, and the infant was directed to turn their head towards the reinforcer within the activation period. The interval between stimulus change and activation of the reinforcer was then gradually incremented by 1000ms per trial (to a maximum of 4000ms), giving the infant increasing opportunities to make an anticipatory head-turn within 4000ms of the onset of a change trial. If the experimenter judged that a head-turn of 45 degrees or greater was produced by the infant, they pressed the reinforcer pedal. Head-turns and thus reinforcer pedal presses that failed to coincide with presentation of the target sound were not reinforced but were recorded by the computer as an index of the rate of between-trial (or random) head-turning. Once the reinforcement period ended and the monkeys were deactivated, the experimenter used toys to attract the infant’s attention to the midline again. When the infant made three correct consecutive anticipatory head-turns within 15 trials, the Training stage was complete and the infant progressed to the Conditioning stage.

5.2.5.2 Conditioning Stage

In the Conditioning stage, there was a 75% probability that a trial would be a change trial and a 25% probability that the trial would be a control trial in which the background stimulus continued to play and the target stimulus was not presented. Trials were presented in a quasi-random order with the constraint that no more than two change trials or two control trials occurred consecutively. Reinforcement was contingent upon an anticipatory head-turn within 4000ms of the onset of a change trial. If the infant responded during this period, the head-turn was signalled by the experimenter (who pressed reinforcer foot pedal) and was recorded as a ‘hit’ by the computer program. If no head-turn occurred during a change trial, a ‘miss’ was
recorded. If the infant failed to produce anticipatory head-turns on three consecutive change trials, they entered 'retraining', in which the change stimulus was paired with the activation of the reinforcer for three change trials or one hit, depending on which came first. The purpose of retraining was to remind the infant that producing a head-turn to a change trial would be positively reinforced. Performance in retraining trials was not used in calculating the infant's final performance score. On control trials, a 'correct-rejection' was recorded in the absence of a head-turn and a 'false alarm' (incorrect response) was recorded if the infant produced an anticipatory head-turn. The criterion for discrimination was set at 7 out of 8 correct consecutive responses (hits plus correct rejections divided by the number of trials, as used by Werker & Tees, 1984), and infants reaching this preset criterion were deemed able to discriminate the contrast and progressed to the Experimental 1 stage. Infants who failed to reach criterion for discrimination within 25 conditioning trials did not progress beyond the Conditioning stage and the session was terminated.

5.2.5.3 Experimental 1 Stage

There were three variations to the procedure from the Conditioning stage to the Experimental 1 stage. Firstly, the probability of a change trial occurring when the start trial pedal was pressed was 50%, as opposed to 75% probability in the Conditioning stage. Secondly there was no preset criterion, rather infants were presented with 15 trials and the number of correct responses recorded. Finally, infants did not revert to retraining if they failed to produce anticipatory head-turns to three consecutive change trials. Infants progressed to the Experimental 2 stage after completing 15 trials.

5.3 Results

The results for the Conditioning and Experimental 1 stages are presented in separate sections below. Chi-square analyses were performed on data from the Conditioning stage to determine the proportion of infants who reached the conditioning criterion and a t-test on the number of trials correct for those reaching vs. not reaching criterion. Percent correct scores in the Conditioning and Experimental 1 stages were analysed using a Univariate Analysis of Variance (ANOVA). Alpha was set at .05 for all tests.
Raw data and statistical output for Experiment 1 are presented in Appendix CD7 on the CD supplement.

5.3.1 Proportion Reaching Criterion

Infants were dichotomously categorised according to their ability to discriminate the Thai tone contrasts using the preset criterion for discrimination of 7/8 correct consecutive responses within 25 trials. Overall 22 of the 48 participants reached criterion. The mean percent correct for these infants was 71% (SD = 15.051, range 44-100%), compared to the remaining 26 infants who averaged just 47.2% correct (SD = 6.301, range = 36-64%). This performance difference was confirmed to be statistically significant by an independent samples t-test, t(46) = 7.332, p = .001, and establishes the validity of this categorisation strategy. A series of chi-square tests was conducted to determine the proportion of infants successfully reaching criterion. A 2 Difficulty (Easy, Difficult) x 2 Criterion (Pass, Fail) chi-square test conducted on the proportion of infants passing and failing in each condition revealed a significant difference between the proportion of infants reaching criterion in the two conditions $\chi^2 (1, N = 48) = 16.448, p = .001$, (see Table 5.1), confirming the hypothesis that the Difficult contrast was significantly more difficult for the infants to discriminate than the Easy contrast. Eighteen of 24 infants passed the Easy condition, compared to only 4 of 24 in the Difficult condition

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>18 (75%)</td>
<td>6 (25%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>4 (16.67%)</td>
<td>20 (83.33%)</td>
<td>24 (100%)</td>
</tr>
</tbody>
</table>

Table 5.1

Proportion of English-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Lexical Tone Contrasts.

62 Note that it was possible for infants to obtain a score over 50% but still not reach criterion of 7/8 correct consecutive responses.
A 2 Age (6-month-olds, 9-month-olds) x 2 Criterion (Pass, Fail) chi-square test was conducted to test if the proportion of 6- and 9-month-old infants reaching criterion differed significantly. As shown in Table 5.2, of the 6-month-old’s, 14 of 24 infants reached criterion, and of the 9-month-old’s, 8 of 24 infants, reached criterion. There was found to be no statistically reliable relationship between age and whether criterion was reached, $\chi^2(1, N = 48) = 3.021, p = .082$.

Table 5.2

<table>
<thead>
<tr>
<th>Age</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>14 (58.33%)</td>
<td>10 (41.67%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>8 (33.33%)</td>
<td>16 (66.67%)</td>
<td>24 (100%)</td>
</tr>
</tbody>
</table>

A further set of chi-square analyses was conducted using pairwise comparisons between the Easy and Difficult condition, to test if the proportion of infants reaching criterion in each condition differed significantly for each age group. A chi-square test revealed that for 6-month-olds, the comparison between the Easy condition and the Difficult condition was statistically significant, $\chi^2(1, N = 48) = 10.971, p = .001$, as shown in Table 5.3, with more 6-month-olds reaching criterion in the Easy condition (11 of 12) than in the Difficult condition (3 of 12). Likewise, as shown in Table 5.4, the proportion of 9-month-olds who passed criterion in the Easy condition, (7 of 12) was significantly greater than the proportion of 9 month-olds who passed the Difficult condition, (1 of 12) $\chi^2(1, N = 48) = 6.750, p = .009$. These data clearly indicate that [bǎ] vs. [bâ] was less discriminable than [bā] vs. [bâ]. The Difficult contrast is just too difficult for these English language infants to discriminate.
Table 5.3
Proportion of 6-month-old English-learning Infants reaching criterion for discrimination of the Easy and Difficult Lexical Tone Contrasts.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>11 (91.67%)</td>
<td>1 (8.33%)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>3 (25%)</td>
<td>9 (75%)</td>
<td>12 (100%)</td>
</tr>
</tbody>
</table>

Table 5.4
Proportion of 9-month-old English-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Lexical Tone Contrasts.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>7 (58.33%)</td>
<td>5 (41.67%)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>1 (8.33%)</td>
<td>11 (91.67%)</td>
<td>12 (100%)</td>
</tr>
</tbody>
</table>

The chi-square results reported above demonstrate that infants from English speaking families had difficulty learning a task in which the Difficult contrast was to be discriminated. Thus a 2 Age (6-month-olds, 9-month-olds) x 2 Criterion (Pass, Fail) chi-square test was performed on the proportion of infants that reached criterion in the Easy condition only (see Table 5.5 below). Eleven of the 12 6-month-old infants and 7 of the 12 9-month-old infants reached criterion. Pearson’s chi-square test revealed that this relationship was just outside significance $\chi^2(1, N = 48) = 3.556, p = .059$. 

164
Table 5.5

Proportion of 6- and 9-month-old English-learning Infants Reaching Criterion for Discrimination of the Easy Lexical Tone Contrast.

<table>
<thead>
<tr>
<th>Age</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>11 (91.67%)</td>
<td>1 (8.33%)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>7 (58.33%)</td>
<td>5 (41.67%)</td>
<td>12 (100%)</td>
</tr>
</tbody>
</table>

5.3.2 Discrimination Performance

5.3.2.1 Entire Sample

The percentage of correct responses in the Conditioning stage was calculated for each infant (both those reaching and not reaching criterion) by dividing the number of hits plus correct rejections by the total number of trials (change plus control trials). There was no evidence of differences due to which tone was the background stimulus in either the Easy condition $t(22) = 1.102, p = .282$ (bā vs. bā, $M = 69.75\%$ correct, $SD = 16.24$; bā vs. bā, $M = 62.08\%$ correct; $SD = 17.8$) or the Difficult condition $t(22) = .265, p = .794$ (bā vs. bā, $M = 51\%$ correct, $SD = 14.08$; bā vs. bā, $M = 49.75\%, SD = .835$), so the data were collapsed across this control factor. A 2 (Age) x 2 (Difficulty) between-subjects ANOVA was employed to compare the percent correct discrimination by 6-month-olds and 9-month-olds to the [bā] vs. [bā] (Easy) and [bā] vs. [bā] (Difficult) contrasts. The mean percent correct discrimination by each age group is shown in Figure 5.1. The statistical test revealed a main effect for Age ($F(1,44) = 6.792, p = .012$), with 6-month-old infants ($M_{6\text{months}} = 63.27\%, SD = 17.70$) showing reliably better discrimination than 9-month-olds infants ($M_{9\text{months}} = 52.96\%, SD = 13.18$). In addition, there was a significant main effect of Difficulty, with better discrimination of the Easy ($M_{\text{Easy}} = 65.86\%, SD = 17.01$) than the Difficult contrast ($M_{\text{Difficult}} = 50.37\%, SD = 11.29$) $F(1,44) = 15.361, p = .001$. In light of this finding and given that only a small number of infants achieved criterion for discrimination in the Difficult condition ($n = 4$), no further analyses on percent correct for the Difficult contrast were performed. There was no Age x Difficulty interaction, $F(1,44) = .320, p = .574$. 

165
Figure 5.1. Mean percent correct discrimination of the easy and difficult lexical tone contrasts by 6- and 9-month-old English-learning infants. Error bars represent standard error of the mean.

An additional 2 x 2 ANOVA with Age (6 month-olds, 9 month-olds) and Criterion (Pass, Fail) as between-subjects factors, and the number of random head-turns between trials (collapsed across Difficulty) as the dependent variable was performed in order to determine whether the 6-month-olds who reached criterion did not score higher simply because of a greater rate of random head-turning between trials, thereby increasing the chance of producing a correct head-turn during a trial. The results indicated no systematic effects for Age, $F(1,44) = .261, p = .612$, ($M_{\text{months}} = 8.29, SD = 6.33$; $M_{\text{months}} = 8.08, SD = 5.63$) Criterion, $F(1,44) = .3.021, p = .089$, ($M_{\text{pass}} = 6.64, SD = 4.82$; $M_{\text{fail}} = 9.500, SD = 6.53$) and no interaction $F(1,44) = .251, p = .619$. This is evidence that the correct responses in the Conditioning stage are not due to an artefact of infant behaviour during testing.
5.3.2.2  Subsample of Infants Reaching Criterion

Percent correct scores for only those infants who reached the criterion for discrimination in the Easy condition were analysed separately. An independent samples t-test revealed that the 6-month-olds ($M = 75.18\%$, $SD = 14.33$) did not significantly outperform the 9-month-olds ($M = 69\%$, $SD = 13.01$), $t(16) = .923$, $p = .370$

5.3.3  Experimental 1 Stage Data

5.3.3.1  Discrimination Performance

Infants who successfully reached criterion for discrimination progressed to the Experimental 1 stage. An independent samples t-test revealed that for those reaching criterion the mean percent correct by 6-month-old ($n = 11$, $M = 64.73\%$, $SD = 13.80$) and 9-month-old infants ($n = 7$, $M = 56.20\%$, $SD = .14.33$) in the Easy tone condition did not differ significantly, $t(16) = 1.259$, $p = .266$. As stated above the data of the four infants in the Difficult condition were not statistically analysed. This finding of no age related decline in Experimental 1 stage performance for infants who reached criterion in the Conditioning stage is not unexpected, given that the subsamples have shown they can discriminate the [bʌ] vs. [bɐ] contrast by reaching the performance criterion.

5.3.4  Performance Across Stages

A paired samples t-test on the percent correct scores in the Easy condition in the Conditioning ($M = 72.78\%$, $SD = 13.791$) and Experimental 1 Stages ($M = 61.41$, $SD = 14.24$) revealed a significant difference, $t(17) = 2.120$, $p = .049$, with performance in the Conditioning Stage, the earlier stage of testing, being significantly superior to performance in the Experimental Stage. Thus there was a decrement in performance across the stages of the session. Two possible reasons are proposed for this, (a) fatigue effects (or some such related factor), and (b) the change from 75% change trials in the Conditioning stage to 50% of change trials in the Experimental 1 stage.
5.3.5 Summary of Results, Experiment 1

It was hypothesised that the 6-month-old English-learning infants would discriminate the lexical tone contrasts better than the 9-month-old English-learning infants, and that discrimination would be superior for the Easy compared to the Difficult contrast. Both hypotheses were supported and the main findings are summarised below:

1. Analyses of percent correct for all infants highlighted superior discrimination by the 6-month-old English-learning infants over the 9-month-old English-learning infants. This effect was not due to a higher incidence of random head-turns by the younger infants. In the Easy condition the proportion of 6-month-old infants reaching criterion was greater than the proportion of 9-month-olds reaching criterion. While this supports the percent correct data analyses, statistical analysis showed that the difference in proportions narrowly failed to reach significance.

2. Chi-square tests and ANOVA on percent correct data confirm that infants of both ages were significantly less able to discriminate the Difficult contrast than the Easy contrast. Thus both 6- and 9-month-old infants discriminated [bā] vs. [bâ], but the vast majority of infants in both age groups treated the [bā] and [bâ] speech sounds as perceptually similar.

5.4 Experiment 2: English-learning Infants’ Discrimination of Nonspeech Tone Contrasts

In Experiment 1 infants were tested on the discrimination of two Thai lexical tone contrasts, [bā] vs. [bâ] and [bâ] vs. [bâ]. The results showed that somewhere between 6 and 9 months of age lexical tone discrimination performance deteriorates for non-tone language environment infants. The purpose of Experiment 2 was to determine the specificity of these results to speech sounds. To this end 6- and 9-month-old infants were tested for their discrimination of the same pitch variations in Experiment 1, but these variations were presented in nonspeech stimuli.
The original [bă], [bâ] and [bă] tokens were used to synthesise violin sounds. This process preserved the fundamental frequency and durational properties of the original speech tokens but removed the phonetic information from the speech signal. It is expected that this removal of phonetic information will allow infants to attend more to the pitch information. It is hypothesised that if the perceptual reorganisation indicated by the decline in discrimination performance over age in Experiment 1 is specific to speech, then discrimination of nonspeech tone contrasts should be similar for the 6- and 9-month-old infants. Additionally, greater attention to pitch information may result in the Easy (rising-falling) and Difficult (rising-low) contrasts being equally discriminable. Consequently, no difference in discrimination ability is predicted for these contrasts.

5.5 Method

5.5.1 Design

A 2 x 2 design was employed with Age (6 months, 9 months) and Difficulty (Easy contrast, Difficult contrast) as between-subjects factors. The contrasts were synthesised violin versions of [bă] vs. [bâ] and [bă] vs. [bă], rising-falling and rising-low respectively. There were 12 infants per age group assigned to each tone contrast condition. The dependent variables were the proportion of infants reaching criterion, and the percentage of correct responses (hits plus correct rejections divided by the number of trials).

5.5.2 Participants

Infants were recruited via the same means as in Experiment 1. The final sample consisted of 48 infants, 24 6-month-olds (14 boys, 10 girls) and 24 9-month-olds (8 boys, 16 girls). The 6-month-olds had a mean age of 6.25 months (SD = .1944, range 6.00-6.82) and the 9 month-olds had a mean age of 9.22 months (SD = .15642 range 8.98-9.48 months). All infants were born between two weeks of full term and were being raised in monolingual Australian-English speaking families. The participating
ENGLISH INFANTS’ TONE DISCRIMINATION

infants had no known health or hearing problems, or a history of middle ear infections. An additional eight infants were tested but excluded from analyses due to crying (2) or fussiness (4), parental interference (1), or because they fell asleep during the testing session (1). Parental consent was obtained prior to testing. Participating parents received $20AUD to cover travel expenses and sponsors’ baby products. Infants were awarded a Young Scientist certificate and a t-shirt printed with ‘I’ve been to uni’.

5.5.3 Stimuli

For this experiment, the F0 from each of the five exemplars of [bǎ], [bâ] and [bà] was tracked using STRAIGHT (Speech Transformation & Representation using Adaptive Interpolation of weiIGHTed spectrogram: Kawahara et al. (1999) and resynthesised onto violin sounds using the Auditory Perception Toolbox (APT: Haszard Morris et al., 2002). Both programs are controlled in Matlab version 6 (The Mathworks Inc, 1999) using a command line interface. Resynthesis involved removing the phonetic information from the speech signal and mapping the F0 of the stimuli onto a keyboard generated violin sound. The violin template was created on a Roland KF Electronic Piano (Roland Digital Intelligent Piano, KF-90). The violin sound was played (the note A below middle C, i.e., 220Hz) for 800ms, recorded directly onto disk from the keyboard, and saved in MIDI format. The sound file was converted from MIDI to .wav via .aif on an Apple Macintosh Power PC G3 (MacOS 8.1, 266MHz, 128MB). This format is suitable for STRAIGHT and APT.

The speech stimuli were analysed in STRAIGHT. The pitch-tracker was used to extract the F0 from each token of [bǎ], [bâ], and [bà] and each individual file was then saved in .mxs format in APT (refer to Appendix A2 for specific details on how to analyse a source for F0). APT was used to morph each of the speech files with the violin template, without changing the F0. To create the nonspeech test stimuli the .mxs file pertaining to each token was accessed and the F0 of the file was mapped onto the violin template. Information on command line instructions can be found in Appendix A2.
The durations of the new exemplars were slightly shorter than the speech files because only the periodic portion of the syllables contributed to the F0 tracking. Importantly however, the information for tone is carried on the vowel (the periodic segment) and the length of the periodic segment for each resynthesised violin soundfile was equal to its equivalent speech token. Another reason for the difference in length of the speech and nonspeech files is due to some speech soundfiles containing prevoicing or creaky voice (a phonetic feature of Thai), and the amplitude of this prevoicing was too low for accurate pitch tracking. Details of the duration for consonant, periodic, and prevoicing segments for all tokens of the speech (Experiment 1) and nonspeech files (Experiment 2) are presented in Appendix A2. Duration of violin tokens derived from [b̥ɑ] ranged from 612ms to 691ms ($M = 651.25, SD = 36.53$), the [bʌ] tokens from 608ms to 657ms ($M = 631.60, SD = 21.55$) and [b̥ʌ] from 600ms to 692ms ($M = 641.2, SD = 35.99$). To preserve natural variability between the tokens and to remain consistent with Experiment 1, stimulus duration was not matched among exemplars. The test contrasts were Easy, rising tone vs. falling tone, and Difficult, rising tone vs. low tone, based on the [b̥ɑ] vs. [bʌ] and [b̥ʌ] vs. [bʌ] contrasts from Experiment 1 respectively. The nonspeech stimuli (audio files .wav format) are presented in Appendix CD-A2 (on the CD supplement).

5.5.4 Materials and Apparatus

The set-up was similar to that of Experiment 1 (refer to section 5.2.4) but with three important differences. First, the loudspeaker was changed to an Acoustic Research 18BXI speaker because the sound quality of this speaker was superior to the loudspeaker used in Experiment 1. The speaker was located 110cm from and 30 degrees to the left of the infant and the reinforcer instead of being directly underneath the reinforcer as in Experiment 1. Second, an additional reinforcer box (height 35cm, width 30cm, and depth 25.5cm) containing a different mechanical toy (a monkey playing the drums) was added to the experimental set-up to further increase positive reinforcement for correct head-turns and to decrease the chance that infants would become bored from repeated exposure to a single toy. The new reinforcer box was

---

63 The loudspeaker was located to the left of the reinforcer from the perspective of the participant.
positioned directly below the existing box (also at 50 degrees and 135 cm to the left of the infant). For each change trial one of the reinforcers was randomly selected by the computer program. Third, the audio masking tape used in Experiment 1 was replaced with one using the new stimuli and music, the ‘Love Songs to the Max’ compilation. This was created as for the original masking tape (see section 5.2.4) for presentation to the parent and experimenter through headphones during the session.

5.5.5 Procedure
The procedure was identical to that reported in Experiment 1 (section 5.2.5). Mean duration of testing was 13.8 minutes (range 7 to 20 minutes).

5.6 Results
Chi–square analyses were performed on data from the Conditioning stage to determine the proportion of infants who reached the 7/8 preset performance criterion. Percent correct scores in the Conditioning and Experimental 1 stages were compared by Univariate between-subjects ANOVA. For all analyses, alpha was set at .05. Raw data and statistical output for Experiment 2 is presented in Appendix CD7 on the CD supplement.

5.6.1 Proportion Reaching Criterion
In the Conditioning stage, 27 infants reached the discrimination criterion and 21 failed to reach criterion. Infants who reached criterion made correct responses on a mean of 73.87% trials (SD = 11.710, Range = 53-100%) and this was significantly higher than the mean percentage correct of 56.19% (SD = 9.59, Range = 32-68%) for infants who failed, t(46) = 5.606, p = .001. A 2 Age (6-month-olds, 9-month-olds) x 2 Criterion (Pass, Fail) chi-square test revealed a relationship between age and whether criterion for discrimination of the nonspeech contrasts was reached $\chi^2 (1, N = 48) = 4.148, p = .042$. Specifically, as shown in Table 5.6, fewer 6-month-old infants passed criterion (10 of 24) than did 9-month-old infants (17 of 24). Thus there was an improvement in performance over age with regard to the proportion of infants reaching criterion.
Table 5.6
Proportion of 6- and 9-month-old English-learning Infants Reaching Criterion for Discrimination of the Nonspeech Tone Contrasts.

<table>
<thead>
<tr>
<th></th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>10 (41.67%)</td>
</tr>
<tr>
<td>9 months</td>
<td>17 (70.83%)</td>
</tr>
</tbody>
</table>

The outcome of a 2 Difficulty (Easy, Difficult) x 2 Criterion (Pass, Fail) chi-square test is shown in Table 5.7, and highlights that there was no significant difference in the proportion of infants reaching criterion for discrimination in the Easy (15 of 24) and Difficult (12 of 24) conditions, $\chi^2(1, N = 48) = 7.62, p = .383$.

Table 5.7
Proportion of English-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Nonspeech Tone Contrasts.

<table>
<thead>
<tr>
<th></th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td><strong>Difficulty</strong></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>15 (62.5%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>12 (50%)</td>
</tr>
</tbody>
</table>

A final chi-square analysing the relationship between infant Age (6-month-olds, 9-month-olds) and Difficulty (Easy, Difficult) of the nonspeech tone contrasts for infants that passed, showed no relationship, $\chi^2(1, N = 48) = 1.342, p = .247$. Table 5.8 shows that 10 of 24 6-month-old infants reached criterion for discrimination of the nonspeech contrasts, and of these 10 infants, 7 reached the criterion for discrimination of the Easy contrast, and only 3 discriminated the Difficult contrast to criterion. Of the 17 9-month-olds who reached criterion, 9 of 17 passed the Difficult condition.
Table 5.8

Proportion of 6- and 9-month-old English-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Nonspeech Tone Contrasts.

<table>
<thead>
<tr>
<th>Age</th>
<th>Difficulty</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Easy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>7 (70%)</td>
<td>3 (30%)</td>
<td></td>
</tr>
<tr>
<td>9 months</td>
<td>8 (47.06%)</td>
<td>9 (52.94%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10 (100%)</td>
<td>17 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

5.6.2 Discrimination Performance

5.6.2.1 Entire Sample

The percent correct discrimination in the Conditioning stage was calculated for each infant (both those reaching and not reaching criterion) by dividing the number of hits plus correct rejections by the total number of trials (change plus control trials). No evidence of a statistical difference was found regarding which nonspeech tone in a pair was presented as the background sound for either the Easy, *t*(22) = 9.49, *p* = .353 (rising vs. falling, *M* = 66.74% correct, *SD* = 8.547; falling vs. rising *M* = 71.84% correct, *SD* = 16.55), or Difficult condition *t*(22) = .922, *p* = .902 (rising vs. low, *M* = 63.35% correct, *SD* = 15.23; low vs. rising, *M* = 62.61% correct, *SD* = 13.79). Therefore data were collapsed across this variable. Mean percent correct scores are shown in Figure 5.2. A 2 Age (6-month-olds, 9-month-olds) x 2 Difficulty (Easy, Difficult) between-subjects ANOVA on the percent correct scores, (Conditioning stage) yielded no main effects. There were no overall differences in performance for 6-month-olds vs. 9-month-olds (*F*(1,46) = 3.918, *p* = .054; *M*<sub>6months</sub> = 62.5%, *SD* = 14.80, *M*<sub>9months</sub> = 69.8%, *SD* = 12.20), or for the Easy (rising tone vs. falling tone) contrast; (MEasy = 69.3%, SD = 13.140) and Difficult (rising tone vs. low tone) contrast (*F*(1, 46) = 2.933, *p* = .094; MDifficult = 63%, SD = 14.22). Thus, the hypothesis that there would be no difference in discrimination performance between the groups was supported. A significant Age x Difficulty interaction (*F*(1,46) = 4.944, *p* = .031) was revealed; 6
month-olds showed superior discrimination for the Easy over the Difficult contrast while for 9 month-olds there was very little difference.

![Graph showing percent correct discrimination by age and condition](image)

**Figure 5.2.** Mean percent correct discrimination of the easy and difficult nonspeech tone contrasts by 6- and 9-month-old English-learning infants. Error bars represent standard error of the mean.

### 5.6.2.2 Subsample of Infants Reaching Criterion

Cell sizes were too small for analysis by a single 2 x 2 ANOVA so the data for infants who reached criterion in the Conditioning stage were analysed using two t-tests on the mean percent correct discrimination for 6-month-olds vs. 9-month-olds, and the Easy vs. Difficult contrasts. There was no significant difference, $t(25) = .623, p = .974$, in the percent correct scores for 6 month-olds ($M_{6\text{months}} = 73.97, SD = 11.48$) vs. 9 month-olds ($M_{9\text{months}} = 73.81\%, SD = 12.19$). There was also no significant difference between discrimination of the rising vs. falling ($M_{\text{easy}} = 74.07\%, SD = 14.39$) and the rising vs. low contrasts ($M_{\text{difficult}} = 73.63\%, SD = 7.77$) $t(25) = 2.096, p = .925$. 
5.6.3 Experimental 1 Stage Data

5.6.3.1 Discrimination Performance

The same pattern of results was found for mean discrimination in the Experimental 1 stage as in the Conditioning stage. The analysis comparing the 6-month-olds and 9-month-olds revealed no significant differences, \( t(25) = .005, p = .996 \). Six-month-olds obtained a mean percent correct of 68.67\% (\( SD = 13.72 \)) and 9-month-olds performed similarly, with a mean of 68.64\% correct (\( SD = 16.95 \)). A \( t \)-test comparison of mean percent correct discrimination in the Easy vs. Difficult condition was also not significant, \( t(25) = 1.430, p = .165 \). Infants scored 64.89\% (\( SD = 15.83 \)) and 73.34\% correct (\( SD = 14.49 \)) for the Easy condition (rising vs. falling) and the Difficult condition (rising vs. low) respectively.

5.6.4 Performance Across Stages

The results reported in the above two sections show that infants who reached the criterion for discrimination performed similarly in the Conditioning and Experimental 1 stages. A paired \( t \)-test confirmed that mean percent correct discrimination for these infants were not significantly different in the Conditioning (\( M = 73.93\% \), \( SD = 11.75 \)) versus the Experimental 1 stages (\( M = 68.65\% \), \( SD = 15.56 \)), \( t(26) = 1.965, p = .060 \).

5.6.5 Summary of Results, Experiment 2

In this study of English-learning infants’ discrimination of nonspeech tone it was hypothesised that 6-month-old and 9-month-old infants would perform equally well, and that there would be no difference in the ability of infants to discriminate the Easy and Difficult contrasts. There were two main findings:

1. Six-month-old and 9-month-old English learning infants performed equally well on the discrimination of nonspeech tone contrasts as measured by percent correct. However with regard to the proportion of infants reaching criterion, more 9-month-olds than 6-month-olds reached criterion, indicating an age-related improvement in discrimination of the nonspeech contrasts.
As hypothesised, statistical tests on percent correct and the proportion of infants passing vs. failing discrimination show that there were no differences in English-learning infants' ability to discriminate the rising vs. falling and rising vs. low nonspeech tone contrasts.

### 5.7 Discriminating Lexical Tone versus Nonspeech tone: Comparing Experiment 1 and Experiment 2

Analyses in this section directly compare the proportion of 6 month-olds and 9-month-olds reaching criterion for discrimination of lexical tone (Experiment 1) and nonspeech tone (Experiment 2). As 6-month-olds have been found to discriminate nonnative contrasts (see Werker & Tees, 1992), and as infants can generally discriminate musical melodies (Trehab, Thorpe, & Morrongiello, 1985), it is hypothesised that 6-month-old infants will discriminate lexical tone and nonspeech tone equivalently - an equal number of 6-month-olds are expected to reach criterion for discrimination on the lexical tone and nonspeech tone discrimination tasks. Sensitivity to nonnative speech contrasts significantly decreases during the second half of the first year for consonants and vowels (see Werker & Tees, 1992), as infants become increasingly attuned to perceiving native language sound patterns. Consequently, it is expected that significantly fewer 9-month-olds reached criterion for discrimination in Experiment 1 (lexical tone discrimination) than in Experiment 2 (nonspeech tone discrimination).

#### 5.7.1 Results

Two $2 \times 2$ chi-square analyses were conducted on each age group, the 6-month-olds and the 9-month-olds. Task (Lexical, Nonspeech) and Criterion (Pass, Fail) were the factors. There was an equal proportion of 6-month-olds passing and failing each task in the two conditions, $\chi^2(1, N = 48) = 1.333, p = .248$. Table 5.9 shows the proportion of 6-month-old English-learning infants passing and failing criterion in the lexical and nonspeech experiments. In Experiment 1, 14 of 24 6-month-old infants reached criterion for discrimination of the lexical tone contrasts, and in Experiment 2, 10 of 24 6-month-old infants reached the criterion for discrimination of the nonspeech contrasts. Conversely, Table 5.10 shows that there was a significant difference in the proportion...
of 9-month-old infants who passed and failed the two tasks $\chi^2 (1, N = 48) = 6.762, p = .009$. Nine-month-olds found the lexical tone discrimination task relatively difficult in comparison to the nonspeech discrimination task, with 8 of 24 infants reaching criterion for discrimination in Experiment 1, and 17 of 24 9-month-olds reaching criterion in Experiment 2. Statistical output for the comparison of Experiment 1 and 2 is presented in Appendix CD7 on the CD supplement.

Table 5.9
**Proportion of 6-month-old English-learning Infants Reaching Criterion for Discrimination on the Lexical Tone (Experiment 1) and Nonspeech Tone (Experiment 2) Tasks.**

<table>
<thead>
<tr>
<th>Task</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>14 (58.33%)</td>
<td>10 (41.67%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>Nonspeech</td>
<td>10 (41.67%)</td>
<td>14 (58.33%)</td>
<td>24 (100%)</td>
</tr>
</tbody>
</table>

Table 5.10
**Proportion of 9-month-old English-learning Infants Reaching Criterion for Discrimination on the Lexical Tone (Experiment 1) and Nonspeech Tone (Experiment 2) Tasks.**

<table>
<thead>
<tr>
<th>Task</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>8 (33.33%)</td>
<td>16 (66.67%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>Nonspeech</td>
<td>17 (70.83%)</td>
<td>7 (29.67%)</td>
<td>24 (100%)</td>
</tr>
</tbody>
</table>
5.7.2 Summary

Comparing the proportion of English-learning infants passing or failing the lexical tone discrimination experiment and the nonspeech tone discrimination experiment provides two important results:

1. An equivalent proportion of 6-month-old English-learning infants reached the discrimination criterion for the lexical tone contrasts in Experiment 1 and the nonspeech tone contrasts in Experiment 2.

2. A greater proportion of 9-month-olds reached criterion for discrimination of the nonspeech tone (Experiment 2) than with lexical tone stimuli (Experiment 1).
CHAPTER 6
Chinese-learning Infants’ Discrimination of Tone
In this chapter two experiments are presented - Chinese language-learning infants’ discrimination of lexical tone (Experiment 3) and nonspeech tone contrasts (Experiment 4).

6.1 Experiment 3: Chinese-learning Infants’ Discrimination of Lexical Tone Contrasts

Chinese language-learning infants were tested using the Conditioned Head-turn procedure (CHT) on their ability to discriminate the Easy Thai lexical tone contrast [bā] vs. [bâ] or the Difficult Thai lexical tone contrast [bā] vs. [bâ]. Forty-eight infants participated in this study. Three of the 48 Chinese infants were tested at the Division of Speech and Hearing Sciences, University of Hong Kong and the remaining 45 were tested at the BabyLab, MARCS Auditory Laboratories, Bankstown Campus, University of Western Sydney, Australia.

Discrimination performance for the lexical tone contrasts was not expected to decline over age; instead it was hypothesised that the 9-month-old infants would discriminate the lexical tone contrasts better than the 6-month-olds, demonstrating facilitated discrimination of the linguistically relevant lexical tone contrasts. It was further hypothesised that the Easy and Difficult contrasts would be discriminated equally well.

6.2 Method

6.2.1 Design

The study employed a 2 x 2 design with Age (6 months, 9 months) and Difficulty (Easy contrast, Difficult contrast) as between-subjects factors.

6.2.2 Participants

A total of 48 infants, 24 (8 boys, 16 girls) 6-month-olds ($M = 6.08; \text{Range} = 5.61-6.85, \text{SD} = .277$) and 24 (11 boys, 13 girls) 9-month-olds ($M = 9.103; \text{Range} = 8.459, \text{SD} =$
Chinese Infants’ Tone Discrimination

.206) from Chinese speaking families participated in the study. To be a participant in this study at least one of the infant’s parents was required to be a Mandarin or Cantonese speaker, and regularly interact with the infant using one or both of these languages. Infants were not excluded if family members also spoke English. The participants were recruited via publicity on television, radio and newspaper, and through pamphlet distribution to Early Childhood Centres, Community Centres, and prominent Chinese businesses in suburbs of Sydney with a large Chinese population (refer to Appendix A1 for details of recruiting procedures). All infants were born within 38 to 42 weeks gestation, had no known hearing disability or medical problem at the time of testing, and no recorded history of middle ear infection.

An additional 8 infants were tested, but were rejected from the final sample set on the grounds of crying (5) and fussiness (3). Parental consent was obtained prior to testing in accordance with University of Western Sydney and University of Hong Kong ethical guidelines for human participants (see Appendix CD6 for information and consent forms). Parents were debriefed following the testing session and received remuneration of $20 AUD for travel expenses, a selection of baby products from the MARCS BabyLab sponsors and a bib or t-shirt.

6.2.2.1 Participant Numbers

The decision to move the Chinese portion of the project to Sydney, while necessary and unavoidable due to the SARS pandemic, was not ideal for two reasons. Firstly, the population of Chinese speakers in Sydney is much smaller than in Hong Kong and secondly, there was a greater possibility that these infants were exposed to languages other than Chinese, and that this could influence their perception of tone language sounds.

In order to maximise sample sizes in both the lexical tone (Experiment 3) and nonspeech tone (Experiment 4) discrimination tasks, the majority of Chinese infants participated in both experiments. The order of experiments was counterbalanced across participants, as was Difficulty. For example, if infants were presented with the Easy
lexical tone contrast in their first test experiment, they were presented with the Difficult nonspeech tone contrast in their second experiment.

Infants either completed the first experiment on Day 1 and then returned for the second experiment on Day 2 (within a week of the first test), or completed both tasks on the same day. The latter arrangement was more suitable for parents travelling a great distance to the University campus. For infants that were tested for both experiments on the same day, parents were asked to feed and rest their baby between experiments. Analyses of order effects are presented later in this section 6.8.

6.2.2.2 Assessment of Infants’ Exposure to a Chinese Language

A questionnaire was administered to the Chinese-speaking parent of each infant to assess their language use and the proportion of time that the infant would hear Cantonese or Mandarin in comparison to English. Parents were asked to report biographical details, and answer questions about their predominant language use, preferred language (if more than one), language proficiency, and language input to their infant. Responses to each question were coded as 0 = predominately Chinese influence to 1 = predominately English influence. These 0/1 scores were then totalled to yield a proportion of English use score. The families of 38 infants completed the questionnaire, a response rate of 79.17% (the questionnaire is presented in Appendix CD4). Participants’ responses are shown in Appendix CD8. Analyses of questionnaire data are reported later in section 6.9.

Of the 38 respondents 21 were native Cantonese speakers, 16 were native Mandarin speakers, and 1 was a native bilingual Cantonese/Mandarin speaker. All respondents were asked to indicate the percentage of language input to their infant that is Chinese, compared to the percentage of English language input. Mean language input for infants was 61.10% Chinese and 38.1% English.

6.2.3 Stimuli

The lexical tone stimuli were the same as those used for the English-learning infants in Experiment 1 (see sections 5.2.3).
6.2.4 Materials and Apparatus

The set-up for testing at MARCS was as per section 5.2.4. However for the three infants tested at the University of Hong Kong (HKU), the experimental set-up varied slightly. Testing at HKU was conducted in a sound attenuated test room. In the adjacent control room, a PC computer with (Microsoft Windows 98 operating system), with a Pine PCI digital wavetable sound card (Crystal 428, V3.0) controlled the presentation of stimuli and the Exeter Hayes and Slater Windows Conditioned Headturn program version 2 (1999) to a single custom made loudspeaker via an amplifier (40 watt output). Above the loudspeaker was a visual reinforcer (an elephant that played a drum), (see Appendix CD5 for a photograph of the reinforcer) enclosed in a wooden box (height 35cm, width 30cm, depth 25.5cm) with a perspex front. The loudspeaker and the visual reinforcer were located at a 50-degree angle 135cm to the left of the infant. The loudspeaker and reinforcer were to the right of the experimenter who was seated 182cm from and directly in front of the infant. The audio masking stimulus tape was presented diotically through Koss UR-20 closed ear headphones. The experimenter and the parent heard continuous streams of the music and experimental stimuli so as to be unaware of what the infant was hearing and unable to influence the infant’s behaviour systematically.

6.2.5 Procedure

The procedure for testing the Chinese language-learning infants was identical to the procedure used to test the English-learning infants reported in Experiment 1. Mean length of testing session was 14 minutes (range 8-20 minutes). All infants were tested by the same experimenter (the candidate) as in Experiment 1 with the exception of the three infants in Hong Kong who were tested by the research assistant at the University of Hong Kong, Hong Kong, China.
6.3 Results

To remain consistent with Experiment 1 and to allow comparisons across language groups the same statistical analyses were performed on the Chinese language learning infants’ data as were performed on the data from the English-language learning infants. Thus chi-square analyses and an independent samples t-test are reported for Conditioning stage data to determine the proportion of infants who reached the conditioning criterion and a t-test on the number of trials correct for those reaching v not reaching criterion. Percent correct scores in the Conditioning stage for all infants were analysed using Analysis of Variance (ANOVA). To analyse the Conditioning stage and Experimental 1 stage data for infants who reached criterion t-tests were employed. Alpha was set at .05 for all tests. Raw data and statistical output for Experiment 3 are presented in Appendix CD8 on the CD supplement.

6.3.1 Proportion Reaching Criterion

Infants were categorised using the criterion of 7 out of 8 correct responses on consecutive trials in the Conditioning stage with respect to whether they passed or failed to reach the criterion for discrimination. A total of 12 of the 48 infants tested reached the discrimination criterion. Infants who passed obtained a mean percentage correct of 74.80% (SD = 11.66, Range = 36), compared to the 36 infants who failed to reach criterion, with a mean of 52.65% (SD = 10.18, Range = 48). This difference in performance was significant, t(46) = 6.294, p = .000.

Two separate chi-square analyses were performed to determine whether the proportion of infants successfully reaching criterion for discrimination was significantly different across Age and Difficulty.

A 2 Age (6-month-olds, 9-month-olds) x 2 Criterion (Pass, Fail) chi-square test was conducted to determine if there was a relationship between age and reaching criterion. Table 6.1 shows the proportion of 6- and 9-month-olds infants reaching criterion. Five of 24 6-month-olds and 7 of 24 9-month-olds passed the task, but there was no difference between ages in success at reaching the performance criterion $\chi^2 (1, N = 48) = 4.44, p = .505.$
Table 6.1

Proportion of 6- and 9-month-old Chinese-learning Infants Reaching Criterion for Discrimination of the Lexical Tone Contrasts.

<table>
<thead>
<tr>
<th>Age</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>5 (20.83%)</td>
<td>19 (79.17%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>7 (29.17%)</td>
<td>17 (70.83%)</td>
<td>24 (100%)</td>
</tr>
</tbody>
</table>

A second chi-square, a 2 Difficulty (Easy, Difficult) x 2 Criterion (Pass, Fail) analysis, assessed the proportion of infants who reached criterion in the Easy and Difficult conditions respectively, and this is presented in Table 6.2 below. In both the Easy and Difficult conditions, 6 out of 24 infants correctly responded to 7/8 trials, and no relationship was found between Difficulty of contrast and whether criterion for discrimination was reached $\chi^2 (1, N = 48) = .000, p=1$.

Table 6.2

Proportion of Chinese-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Lexical Tone Contrasts.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>6 (25%)</td>
<td>18 (75%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>6 (25%)</td>
<td>18 (75%)</td>
<td>24 (100%)</td>
</tr>
</tbody>
</table>
Lastly a 2 (Age) x 2 (Difficulty) chi-square was conducted for only the infants who reached criterion. Three of six 6-month-olds and three of seven 9-month-olds reached criterion in the Easy condition while for the Difficult condition three of six 6-month-olds and four of seven 9-month-olds reached criterion (Table 6.3). Again the analysis revealed that there was no relationship between the age of the infants and whether they passed the performance criterion $\chi^2 (1, N = 13) = .066, p = .797$.

Table 6.3

Proportion of 6- and 9-month-old Chinese-learning Infants Reaching Criterion for Discrimination of the Lexical Tone Contrasts.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Easy</th>
<th>Difficulty</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>3 (50%)</td>
<td>3 (50%)</td>
<td>6 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>3 (42.86%)</td>
<td>4 (57.14%)</td>
<td>7 (100%)</td>
</tr>
</tbody>
</table>

The chi-square analyses clearly show no effect of age or difficulty of contrast on the proportion of infants successful at reaching the criterion of 7 out of 8 correct consecutive responses.

6.3.2 Discrimination Performance

6.3.2.1 Entire Sample

Percent correct responses to trials in the Conditioning stage were computed for every infant and became the dependent variable. There was no evidence of differences due to which tone was the background stimulus for either the Easy $t(22) = .843, p = .408$ (bâ vs. bâ, $M = 59.34\%$ correct, $SD = 13.11$; bâ vs. bâ, $M = 55.53\%$ correct, $SD = 6.553$) or Difficult $t(22) = .735, p = .470$ (bâ vs. bâ, $M = 56.02\%$ correct, $SD = 14.21$; bâ vs. bâ, $M = 61.24$, $SD = 20.09$) contrasts. A 2(Age) x 2(Difficulty) between-subjects ANOVA was performed on percent correct scores. This analysis compared the mean
percent correct responses by 6- and 9-month-old Chinese-learning infants on their discrimination of the Easy [bā] vs. [bâ] and Difficult [bă] vs. [bâ] lexical tone contrasts. The mean percent correct for the Easy and Difficult contrasts is plotted for each age group in Figure 6.1. Consistent with the prediction that there would be no decline across age in discrimination of the lexical tone contrasts, there was no significant difference between the performance of 6-month-olds ($M = 59.64\%, SD = 11.460$) and 9-month-olds ($M = 56.74\%, SD = 16.7$), $F(1,44) = .472, p = .496$. There was also no effect of Difficulty, with infants discriminating the Easy ($M = 57.75\%, SD = 10.84$) and Difficult ($M = 58.63\%, SD = 17.23$) contrasts equally well, $F(1,44) = .043, p = .836$.

![Graph](image)

*Figure 6.1.* Mean percent correct discrimination of the easy and difficult lexical tone contrasts by 6-and 9-month-old Chinese-learning infants. Error bars represent standard error of the mean.

### 6.3.2.2 Subsample of Infants reaching Criterion

Twelve infants successfully reached the criterion for discrimination of the lexical tone contrasts. Percent correct for the performance of these 6-and 9-month-olds in the Easy and Difficult conditions was analysed in two independent samples $t$-tests instead of ANOVA due to small cell sizes. The $t$-test comparing 6-month-olds vs. 9-month-olds
Chinese infants’ tone discrimination

on percent correct found that the mean percent correct of 69.58% (SD = 10.21) for 6-month-olds and 78.53% (SD = 11.87) correct for 9-month-olds was not a statistically reliable difference $t(10) = 1.360, p = .204$. The $t$-test comparing discrimination of the Easy vs. Difficult contrast revealed no significant difference between mean percent correct discrimination of the Easy ($M = 71\%, SD = 9.34$) and the Difficult contrast ($M = 78.6\%$ (SD = 13.31)). Taken together the $t$-tests provide no evidence of differences between discrimination of the lexical tone contrasts by 6-month-old or 9-month-old infants or of differences between discrimination of the Easy and Difficult contrasts.

6.3.3 Experimental 1 Stage Data

6.3.3.1 Discrimination Performance

Infants who passed the Conditioning stage were tested in the Experimental 1 stage. Independent samples $t$-tests were performed on Age (6-month-olds, 9-month-olds) and on Difficulty (Easy, Difficult). The first $t$-test analysis showed that there was no significant difference in the mean percent correct for 6- and 9-month-olds, 58.67% (SD = 14.45) and 64.52% (SD = 14.23), respectively, $t(10) = .698, p = .501$. The second $t$-test revealed that the mean percent correct discrimination for infants in the Easy condition, 70.83% (SD = 5.75), was significantly greater, $t(10) = 2.731, p = .021$, than mean percent correct discrimination for infants in the Difficult condition ($M = 53.34\%, SD = 14.61$).

6.3.4 Performance Across Stages

A paired samples $t$-test on percent correct discrimination in the Conditioning ($M = 75.78\%, SD = 11.71$) and Experimental 1 stages ($M = 61.92, SD = 13.40$) for infants who completed both, revealed a significant difference, $t (12) = 2.783 p = .017$, with performance in the earlier Conditioning stage being significantly better than performance in the later Experimental stage.
6.3.5 Summary of Results, Experiment 3

Overall, the pattern of results for Chinese language-learning infants’ discrimination of lexical tone supports the hypothesis that there is no decrement in performance (either in the proportion of infants reaching criterion or in percent correct responses) between the 6-month-old and 9-month-old age groups on the discrimination of [bǎ] vs. [bâ] or [bǎ] vs. [bâ]. The main findings are summarised below:

1. Analyses of percent correct discrimination and chi square tests of the proportion of 6 and 9-month-old Chinese-learning infants reaching criterion provide evidence that Chinese-learning infants’ perceptual discrimination of the lexical tone contrasts is maintained across age because tone is used phonemically in Chinese languages. The possibility of increased discrimination performance in the 9-month-old age group (due to their greater experience with lexical tone) compared to the 6-month-old age group was not upheld.

2. The hypothesis that infants in the Easy and Difficult conditions would perform equivalently was supported for the most part, with a difference only found in the Experimental 1 stage, the third stage of testing. The emergence of a difficulty effect in the Experimental 1 stage suggests that the duration of testing may have increased the cognitive demands associated with discriminating the more psychoacoustically difficult contrast.

6.4 Experiment 4: Chinese-learning Infants’ Discrimination of Nonspeech Tone Contrasts

In Experiment 4 Chinese language-learning infants were tested for their discrimination of synthetic nonspeech tones, with identical F0 trajectories to the lexical tones as used with English-learning infants in Experiment 2. There are two hypotheses. Firstly, it is expected that the Chinese-learning infants will perform equivalently on this task and the lexical tone task, with no difference in discrimination performance by 6-month-old and 9-month-old infants. Secondly, it is expected that Chinese language-learning infants’ experience with complex lexical tone sounds will facilitate their discrimination.
of tone in a presumably easier pitch discrimination task, where only F0 cues are present.

6.5 Method

6.5.1 Design
A 2 x 2 design was employed with Age (6 months, 9 months) and Difficulty (Easy contrast, Difficult contrast) as between-subjects factors. Infants were randomly assigned to the Easy and Difficult conditions until 10 infants per age were tested in each condition. See section 5.2.1 for all other design considerations.

6.5.2 Participants
Forty infants, 20 6-month-olds (8 boys, 12 girls) and 20 9-month-olds (8 boys, 12 girls) participated in the study. The mean age of the 6-month-old infants was 6.12 months (SD = .25, Range = 5.84-6.85) and the mean age of the 9 month-old infants was 9.04 months (SD = .22, Range = 8.53-9.77). All subjects were born within two weeks either side of full term (38 to 42 weeks) and were recruited from Cantonese and/or Mandarin speaking families. The participating infants had no known health or hearing problems at the time of testing and no history of middle ear infection. An additional six infants were tested but excluded from analyses due to crying (2), fussiness (3) or because they fell asleep during the testing session (1). Infant recruitment and remuneration for participation was as per Experiment 3, section 6.2.2).

6.5.3 Stimuli
The stimuli for this experiment were the same rising vs. falling (Easy) and rising vs. low (Difficult) contrasts used in the nonspeech tone discrimination task with English-learning infants (Experiment 2).
6.5.4 Materials and Apparatus

Materials and apparatus were identical to those for the nonspeech tone discrimination task with English-learning infants (Experiment 2) tested in Sydney (and none of the infants in Experiment 4 were tested in Hong Kong).

6.5.5 Procedure

The procedure was identical to that reported in Experiment 2. Mean duration of the testing sessions was 12 minutes (Range = 9-16 minutes).

6.6 Results

Chi-square analyses were performed on data from the Conditioning stage to determine the proportion of infants who reached the 7/8 preset criterion. Percent correct scores in the Conditioning and Experimental 1 stages were compared by ANOVA and by t-test when cell sizes were small. Alpha was set at .05 for all statistical tests. Raw data and statistical output for Experiment 4 is presented in Appendix CD8 on the CD supplement.

6.6.1 Proportion Reaching Criterion

In the Conditioning stage of the nonspeech tone discrimination task, 13 of 40 (32.5%) infants reached the criterion for discrimination. Of these 13 infants, four were 6-month-olds and nine were 9-month-olds (see Table 6.4). Infants that passed criterion scored a mean percent correct of 74.63%, \(SD = 13.47\), \(Range = 26.6\), and this was significantly better than the mean percent correct for infants that failed 51.76% \(SD = 8.8\), \(Range = 32\), \(t(46) = 6.967, p = .000\). A 2 Age (6 months, 9 months) \(\times\) 2 (Pass, Fail) chi-square indicated that there was no effect of age on whether infants reached the discrimination criterion \(\chi^2(1, N = 40) = 2.568, p = .109\). A second chi-square was used to analyse the proportion of infants who reached criterion in the Easy and Difficult conditions. Nine of 20 infants reached criterion in the Easy condition, and 11 of 20 infants in the Difficult condition, and this difference was not significant, \(\chi^2(1, N = 40) = 2.849, p = .091\) (Table 6.5).
Table 6.4
Proportion of 6- and 9-month-old Chinese-learning Infants Reaching Criterion for Discrimination of the Nonspeech Tone Contrasts.

<table>
<thead>
<tr>
<th>Age</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>4 (20%)</td>
<td>16 (80%)</td>
<td>20 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>9 (45%)</td>
<td>11 (55%)</td>
<td>20 (100%)</td>
</tr>
</tbody>
</table>

Table 6.5
Proportion of Chinese-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Nonspeech Tone Contrasts.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>9 (45%)</td>
<td>11 (55%)</td>
<td>20 (100%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>11 (55%)</td>
<td>9 (45%)</td>
<td>20 (100%)</td>
</tr>
</tbody>
</table>

A third chi-square examined if there was a relationship between Age (6-month-olds vs. 9-month-olds) and Difficulty (Easy vs. Difficult) of contrast for infants who reached the discrimination criterion. All four 6-month-olds who reached the criterion for discrimination, did so for the Easy contrast, with no 6-month-olds reaching criterion of the Difficult contrast. In the 9-month-old group, 5 of 9 infants passed in the Easy condition and 4 of 9 infants passed in the Difficult condition, and this difference was not significant, $\chi^2(1, N = 13) = .066, p = .797$. 

193
Table 6.6

Proportion of 6- and 9-month-old Chinese-learning Infants Reaching Criterion for Discrimination of the Easy and Difficult Nonspeech Tone Contrasts.

<table>
<thead>
<tr>
<th>Age</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Easy</td>
</tr>
<tr>
<td>6 months</td>
<td>4 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>5 (55.6%)</td>
</tr>
</tbody>
</table>

6.6.2 Discrimination Performance

6.6.2.1 Entire sample

Two independent samples t-tests were performed to investigate whether there were any differences in percent correct discrimination based on which tone of a pair was the background sound for either the Easy or Difficult contrast. No order effects were found for the Easy contrast $t(18) = 1.206, p = .243$ (rising vs. falling, $M = 67.53\%$ correct, $SD = 16.40$; falling vs. rising, $M = 59.47\%, SD = 13.34$) or the Difficult contrast $t(18) = .810, p = .429$ (rising vs. low, $M = 57.48\%, SD = 15.53$; low vs. rising, $M = 52.13\%, SD = 12.12$). Percent correct scores for all infants were entered into a 2 x 2 ANOVA with Age and Difficulty as factors. The mean percent correct for 6- and 9-month-old infants in the Easy and Difficult conditions are shown in Figure 6.2. Mean percent correct discrimination for 6-month-olds was 55.4\% ($SD = 12.42$) and 62.9\% for 9-month-olds ($SD = 16.75$), and this difference was not significant, ($F(1,38) = 2.598, p = .116$). There was also no significant difference between performance in the Easy ($M = 62.9\%, SD = 15.66$) and Difficult ($M = 55.3\%, SD = 13.71\%$) conditions ($F(1,38) = 2.736, p = .107$), and no interaction, $F(1,38) = .105, p = .748$.
Figure 6.2. Mean percent correct discrimination of the easy and difficult nonspeech tone contrasts by 6- and 9-month-old Chinese-learning infants. Error bars represent standard error of the mean.

6.6.2.2 Subsample of Infants Reaching Criterion

The percent correct data for 6-month-old and 9-month-old infants reaching criterion in the Conditioning stage was analysed by an independent samples t-test. Six-month-olds who reached criterion obtained a mean percent correct of 69.81% ($SD = 16.50$) and there was no significant difference between this and the mean percent correct discrimination by 9-month-olds ($M = 76.44\%$, $SD = 13.28$) $t(11) = .775$, $p = .455$. A second independent samples t-test revealed no significant difference between mean percent correct for infants who passed criterion in the Easy condition ($M = 73.6\%$, $SD = 13.62$) compared to infants in the Difficult condition ($M = 75.3\%$, $SD = 16.93$), $t(11) = .152$, $p = .882$. 

195
6.6.3 Experimental 1 Stage Data

6.6.3.1 Discrimination Performance

Percent correct data from infants who passed criterion in the Conditioning stage and then went on to complete the Experimental 1 stage were analysed by two independent samples t-tests, the first comparing discrimination by 6-month-olds vs. 9-month-olds and the second comparing discrimination of the Easy vs. Difficult contrasts. The 6-month-old age group obtained a mean percent correct of 60.74% (SD = 5.59) and the 9-month-olds obtained a mean percent correct of 71.11% (SD = 16.67) and this difference was not significant $t(10) = 1.029, p = .328$. The second $t$-test revealed no difference between the mean percent correct for the Easy ($M = 72.78\%$, $SD = 16.14$) and Difficult contrasts ($M = 60$, $SD = 9.43$), $t(10) = 1.433, p = .180$.

6.6.4 Performance Across Stages

A paired samples $t$-test showed that there was no significant difference between percent correct scores in the Conditioning ($M = 76.6\%$, $SD = 12.03$) vs. Experimental 1 ($M=68.52\%$, $SD = 15.16$) stages, although this difference was close to being significant, $t(11) = 2.024, p = .068$.

6.6.5 Summary of Results, Experiment 4

The main findings from the study of Chinese-learning infants’ discrimination of nonspeech tone are as follows:

1. Six-month-olds and 9-month-olds performed equally well on the discrimination of nonspeech tone contrasts in terms of percent correct.

2. There was no effect of age on the proportion of infants who reached criterion in either the Easy or Difficult conditions and there were no significant differences in the discrimination of the Easy contrast or the Difficult contrast, except in the Experimental 1 stage where the Chinese language-learning infants showed better discrimination of the Easy contrast.
6.7 Discriminating Lexical Tone versus Nonspeech Tone: Comparing Experiment 3 and Experiment 4

In this section the proportion of Chinese-learning infants reaching criterion for discrimination of lexical tone (Experiment 3) and nonspeech tone (Experiment 2) contrasts is compared. Two 2 Task (Lexical, Nonspeech) x 2 Criterion (Pass, Fail) chi-square analyses were performed on data from 6- and 9-month-old Chinese-learning infants. No differences in the proportion of infants reaching criterion in the lexical tone or nonspeech tone discrimination tasks are hypothesised for either the 6-month-olds or the 9-month-old infants. Statistical output for the comparison of Experiments 3 and 4 are presented in Appendix CD8 on the CD supplement.

6.7.1 Results

Both chi-squares comparing the proportion of infants reaching criterion in the lexical tone vs. nonspeech tone experiments revealed no relationship between task and whether infants passed or failed the task. For 6-month-old infants (Table 6.7), 5 of 24 infants and 4 of 20 infants reached criterion for discrimination performance in the lexical tone and nonspeech tone tasks respectively $\chi^2(1) = .005, p = .946$. For the 9-month-old infants (Table 6.8), 7 of 24 infants (29.2%) reached criterion for discrimination in the lexical tone experiment, and 9 of 20 (45%) reached the criterion for discrimination in the nonspeech tone experiment $\chi^2(1, N = 88) = 1.182, p = .277$.

Table 6.7

<table>
<thead>
<tr>
<th>Task</th>
<th>Difficulty</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Easy</td>
<td>Difficult</td>
<td></td>
</tr>
<tr>
<td>Lexical</td>
<td>5 (20.83%)</td>
<td>19(79.17%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>Nonspeech</td>
<td>4 (20%)</td>
<td>16 (80%)</td>
<td>20(100%)</td>
</tr>
</tbody>
</table>

197
Table 6.8
*Proportion of 9-month-old Chinese-learning Infants Reaching Criterion for Discrimination on the Lexical Tone (Experiment 3) and Nonspeech Tone (Experiment 4) Tasks.*

<table>
<thead>
<tr>
<th>Task</th>
<th>Criterion</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td></td>
<td>7 (29.2%)</td>
<td>17 (70.8%)</td>
<td>24 (100%)</td>
</tr>
<tr>
<td>Nonspeech</td>
<td></td>
<td>9 (45%)</td>
<td>11 (55%)</td>
<td>20 (100%)</td>
</tr>
</tbody>
</table>

### 6.7.2 Summary

Together these findings indicate similar performance of the Chinese language-learning infants in the lexical tone discrimination experiment (Experiment 3) and the nonspeech tone discrimination experiment (Experiment 4).

### 6.8 Order Effects

There is a possibility that results for Chinese language-learning infants’ discrimination of lexical tone and nonspeech tone are confounded by the fact that most of the Chinese infants completed both tasks. Of the 48 infants who participated in the lexical tone study (Experiment 3), 32 also completed the nonspeech tone task (Experiment 4). The remaining 16 infants were either too unsettled to complete another experiment later in the day, or the parent was unable to return for testing on Day 2. To check for order effects a 2 Age (6-month-olds, 9-month-olds) x 2 Task (Lexical tone, Nonspeech tone) x 2 Order (Order 1 lexical tone first, Order 2 nonspeech tone first) mixed plot repeated measures ANOVA was performed on percent correct data. Age and Order were the between-subjects factors and Task was the within subjects factor. There were no significant differences between percent correct as a function of Age ($M_{6\text{months}} = 57.6\%, SD = 12.04; M_{9\text{months}} = 59.5\%, SD_{9\text{months}} = 17.038; F(1,28) = .254, p = .618$) or Task ($M_{\text{Lexical}} = 57.75, SD = 14.1; M_{\text{Nonspeech}} = 62.24, SD = 17.05; F(1, 28) = .451, p = .507$).
In addition, no effects of Order between the task completed first and the task completed second were found $M_{\text{Order1}} = 57.9$, $SD = 13.401$; $M_{\text{Order2}} = 59.2$, $SD = 18.429$, $F(1, 28) = .111$, $p = .741$ indicating that infants performed equivalently across the two testing sessions. There were no significant interactions between any of the three factors (all $p > .05$). Statistical output for order effects is presented in Appendix CDB on the CD supplement.

Thus the results for Chinese language-learning infants' discrimination of lexical tone and nonspeech tone are robust and unaffected by testing order.

### 6.9 Questionnaire Data

It was necessary to determine if an infant's relative degree of exposure or experience with the Chinese and English languages is related to their ability to discriminate the tone contrasts. This was of a particular concern because Chinese infants were tested in Sydney, where there was a greater likelihood of exposure to English. However, it must be noted that English is widely spoken in Hong Kong where testing was originally planned, so this was always a potential problem for generalisation of the results. In this section, Pearson product moment correlations, factor analyses and standard regression tests will determine whether parents' overall language use, and infants experience with one or more languages, is related to their performance on the lexical tone discrimination task.

#### 6.9.1 Coding Responses

Questions on the questionnaire were grouped into five question categories - biographical details, predominant language use, preferred language (if more than one), language proficiency and language input to their infant. Answers to questions were coded as $0 =$ Chinese influence to $1 =$ English influence and summed for each of the five categories to obtain an overall proportion of English use score for each parent, a total language use score. Missing data on any question was substituted with the mean score of all other responses in the question category, so that correlational, factor analysis and regression analyses could be conducted appropriately. Mean proportion of overall English language use for the respondents was $38.10\%$, $(SD = 10.34)$, Range
6.9.2 Correlations

Pearson product moment correlations were conducted between the five questionnaire variables - *biographical details* (Bio), *predominant language use* (Pred), *preferred language* if more than one (Pref), *language proficiency* (Prof) *language input* (Input), *total language use* (Total) – and infants’ percent correct discrimination of lexical tone (Infants’ Score), using an alpha level of .05. Table 6.9 displays the correlations. *Language input to infant* was significantly correlated \( (r = .49) \) with *preferred language use* at the \( p < .05 \) level. All other questionnaire variables were significantly positively correlated with one another at the \( p < .01 \) level, the highest positive correlation was between *total language use* and *language proficiency* \( (r = .892) \). This indicates that a high score on one questionnaire variable is related to a high score on other questionnaire variables. However, no questionnaire variables (administered to parents) were significantly correlated with infants’ percent correct discrimination of lexical tone contrasts. Correlations between questionnaire variables and infants’ percent correct discrimination ranged from \( r = -.005 \) to \( r = .197 \). Statistical output for the correlation analyses is presented in Appendix CD8 on the CD supplement.
Table 6.9

Pearson Product Moment Correlations Between Parents’ Language Use and Infants’ Mean Percent Correct Lexical Tone Discrimination.

<table>
<thead>
<tr>
<th></th>
<th>Infants’ Score</th>
<th>Bio</th>
<th>Pred</th>
<th>Pref</th>
<th>Prof</th>
<th>Input</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infants’ Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bio</td>
<td>-.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pred</td>
<td>.003</td>
<td>.574**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pref</td>
<td>.168</td>
<td>.422**</td>
<td>.478**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof</td>
<td>.197</td>
<td>.723**</td>
<td>.624**</td>
<td>.590**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input</td>
<td>.041</td>
<td>.436**</td>
<td>.585**</td>
<td>.409*</td>
<td>.584**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.097</td>
<td>.809**</td>
<td>.804**</td>
<td>.672**</td>
<td>.892**</td>
<td>.786**</td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$
* $p < .01$

6.9.3 Factor Analysis

The five questionnaire variables and the total language use score were factor analysed. Hills (2003) states that there must be at least a minimum of five participants per variable in a factor analysis. In this instance there are 38 participants, greater than the 30 required by this rule.

A principal components analysis with varimax rotation was performed on the questionnaire variables (statistical output for this analysis is presented in Appendix CD8, on the CD supplement). There were no missing data. A single component with an eigenvalue greater than unity was extracted, accounting for 69.67% of the variance. The factor was labelled Language Use. The varimax rotation was not required because only one component was extracted. Component loadings and extraction communalities ($h^2$) are shown in Table 6.10. The total language use variable had the highest extraction communality indicating that the Language Use factor is well represented by this variable. The lowest communality was the preferred language use variable.
**Table 6.10**  
*Principle Component Analysis Loadings for Chinese Language Use Questionnaire Variables.*  

<table>
<thead>
<tr>
<th>Component</th>
<th>Factor 1</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biographical</td>
<td>.802</td>
<td>.643</td>
</tr>
<tr>
<td>Predominant</td>
<td>.818</td>
<td>.669</td>
</tr>
<tr>
<td>Preferred</td>
<td>.704</td>
<td>.495</td>
</tr>
<tr>
<td>Proficiency</td>
<td>.893</td>
<td>.797</td>
</tr>
<tr>
<td>Language</td>
<td>.762</td>
<td>.580</td>
</tr>
<tr>
<td>Total</td>
<td>.998</td>
<td>.996</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>69.677</td>
<td>69.677</td>
</tr>
<tr>
<td>Label</td>
<td>Language Use</td>
<td></td>
</tr>
</tbody>
</table>

### 6.9.4 Regression

The *Language Use* factor scores for each participant ($N = 38$) (requested as part of the factor analysis) were added to the dataset and entered as a predictor variable into a standard regression analysis, predicting the percent correct lexical tone discrimination of the 38 infants. Tabachnick and Fidell (1989) state that an absolute minimum sample size is at least five times as many participants as the number of predictor variables. According to this rule, an absolute minimum sample size would be 5. Thus the sample size of 38 was sufficient for a regression analysis.

An unstandardised regression coefficient ($B$) of 0.01295 and the standardised regression coefficient ($B$) of .099 were obtained. A $t$-test on these regression coefficients was not significant ($p>.05$), indicating that the predictor variable *Language Use* does not make a significant contribution to predicting infants' percent correct lexical tone discrimination. The $R$ and $R^2$ and adjusted $R^2$ values are .099, .010 and -.018 respectively. The low $R$ squared ($R^2$) value shows that only 9.9% of the proportion of variation in infants' percent correct discrimination of lexical tone is explained by the
regression model, indicating that the model does not fit the data well. When $R^2$ was adjusted to more closely reflect the goodness of fit of the model in the population, the model only explains 1% of variation in infants’ percent correct discrimination performance. This is borne out by statistical test - *Language Use* of parents did not predict their infants’ discrimination of lexical tone, $F(1,36) = .359, p = .553$. Statistical output for the regression analysis is presented in Appendix CD8 on the CD supplement.

### 6.9.5 Summary

Pearson’s correlations revealed no association between the proportion of English relative to Chinese used by parents, and their infants’ discrimination of lexical tone. Moreover, while the questionnaire variables loaded onto a single factor, *Language Use*, this model failed to be a good predictor of infants’ discrimination performance. Thus the Chinese language-learning infants’ ability to discriminate lexical tone contrasts was independent of their parents’ relative use of English and Chinese, so differences between results for the English language infants and Chinese language infants can be attributed to the difference in their language environments.
CHAPTER 7
Interlude: The Story so Far...
The cross-sectional studies, Experiments 1, 2, 3 and 4, suggest that lexical tone perception is shaped specifically by exposure to and experience with a lexical tone language. The results with the English-learning infants in Experiment 1 establish that initial perceptual abilities appear to shift from an unimpeded, language-general form towards a more constrained, language-specific pattern. Without lexical tone in their speech input, there appears to be attenuation of the ability of 9-month-old (compared with 6-month-old) English-learning infants to discriminate lexical tone contrasts.

A different pattern emerged for the Chinese-language learning infants whose native speech input is tonal - their perceptual abilities appear to be maintained through their regular exposure to and experience with a lexical tone language. Considering that lexical tone plays such an important role in tone languages, it is surprising that the 9-month-old Chinese language infants greater experience with lexical tone did not play a facilitative role in their phonetic perception, as has been found for some native consonant contrasts (Polka et al., 2001).

While complete in their support of a perceptual reorganisation for normal tone perception, these cross-sectional studies do not provide evidence regarding individual development. Is there a decline in discrimination of lexical tone contrasts over age for infants with no tone language experience? Longitudinal studies, unlike cross-sectional studies, measure and compare changes in behaviour over time and reveal patterns of individual change. Thus a longitudinal study is more likely to indicate whether an individual infants' initial ability to discriminate lexical tone contrasts really is a function of age and experience with the ambient language.

The next chapter reports longitudinal adaptations of the lexical tone and nonspeech tone experiments in which English language-learning and Chinese language-learning infants were tested for their discrimination of the test contrasts at 6 months of age, and again at 9 months of age. The aim was to track infants' perception of tone across a 3-month developmental window. It is expected that the results of the longitudinal studies of lexical tone and nonspeech tone perception by English- and Chinese-learning infants
will confirm what was found in the cross-sectional studies, and allow more
generalisable conclusions to be drawn. There is a risk in conducting these experiments: It is possible that performance on the task at 9 months of age may be confounded by
earlier participation at 6 months, with the possible consequence that the cross-sectional
data pertaining to the English-learning infants will be overturned, suggesting no perceptual reorganisation of tone.
CHAPTER 8
Longitudinal Data: English- and Chinese-learning Infants’
Discrimination of Tone
In this chapter, 4 longitudinal experiments are reported. The result of English-learning infants’ discrimination of lexical tone (Experiment 5) and nonspeech tone (Experiment 6) are reported first, followed by the Chinese-learning infants’ discrimination of lexical tone (Experiment 7) and nonspeech tone (Experiment 8).

8.1 Longitudinal Study of English-learning Infants’ Discrimination of Lexical Tone: Experiment 5

In Experiment 5, a longitudinal study of the discrimination of lexical tone contrasts, 12 English language infants were tested on the Easy contrast [bǎ] vs. [bâ] when they were 6 months of age and were retested on discrimination of the same contrast 3 months later, at 9 months of age. There was counterbalancing between subjects of which sound in each pair was the background and which was the target stimulus. Infants were not tested on the Difficult [bǎ] vs. [bâ] contrast because the cross-sectional analyses on proportion of infants reaching criterion and mean percent correct indicated that this lexical tone contrast was extremely difficult for English language-learning infants to discriminate.

To maintain consistency with the cross-sectional study, the same statistical tests were performed on the longitudinal data. Thus, chi-square analyses were performed on data from the Conditioning stage to determine the proportion of infants who reached criterion for discrimination as measured by the 7/8 preset performance criterion, and percent correct scores in the Conditioning and Experimental 1 stages were compared by paired samples t-test. For all analyses, alpha was set at .05.

8.2 Method

8.2.1 Participants

Twelve infants (7 girls and 5 boys) were tested at 6 months of age ($M = 6.41$ months, $SD = 0.11$, Range = 6.13-6.79) and again at 9 months of age ($M = 9.34$ months; $SD = 0.11$, Range = 9.18-9.61). An additional two infants were tested, but were rejected
from the final sample on the grounds of fussiness. The infants were being raised in monolingual Australian-English speaking families residing in the Sydney metropolitan area. All infants were born full-term or within two weeks of full-term with uncomplicated births, were judged to be healthy at the time of testing, and had no known hearing disabilities or history of ear infections.

Parental consent was obtained prior to testing in accordance with University of Western Sydney ethical guidelines for human participants (see Appendix CD6 on the CD supplement for the information hand-out and consent form). Parents were debriefed following the testing session and received remuneration of $20 AUD for travel expenses, a selection of baby products from the project sponsors and a bib or t-shirt.

8.2.2 Stimuli

The Easy speech contrast exemplars [bā] vs. [bâ] used in the cross-sectional studies of lexical tone discrimination (Experiment 1 and Experiment 3) were used here.

8.2.3 Materials and Apparatus

Testing was conducted at the MARCS BabyLab, University of Western Sydney. The experimental materials and apparatus were the same as those used in the cross-sectional studies of lexical tone discrimination (see Experiment 1).

8.2.4 Procedure

As for the previous experiments, the CHT (Exeter Hayes and Slater Windows Conditioned Headturn program, version 2, 1999) was used for testing infants in this longitudinal study.

8.3 Results

Raw data and statistical output for Experiment 5 is presented in Appendix CD9 on the CD supplement.
8.3.1 Proportion Reaching Criterion

To be a participant in this longitudinal study, the infants needed to reach criterion for discrimination when they were tested at 6 months, in order to discover if and how the ability to discriminate the tone contrasts might change over age. Therefore in the Conditioning stage, all 12 infants (100%) comprising the final sample reached the criterion for discrimination (7 correct responses within 8 trials) when they were tested at 6 months of age as reported in Table 8.1. When retested at 9 months, only 7 of the 12 infants reached the criterion for discrimination. A 2 Age (6 months, 9 months) x 2 Criterion (Pass, Fail) chi-square test (Fischer’s Exact) determined that this difference was significant, $\chi^2 (1, N = 12) = 6.316, p = .037$, indicating a decline over age for English language infants’ ability to discriminate the Easy tone contrast to criterion.

Table 8.1

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>12 (100%)</td>
<td>0 (0%)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>7 (58.33%)</td>
<td>5 (41.67%)</td>
<td>12 (100%)</td>
</tr>
</tbody>
</table>

8.3.2 Discrimination Performance

8.3.2.1 Entire Sample

The percentage of correct responses in the Conditioning stage was calculated for each infant (both those reaching and not reaching criterion) by dividing the number of hits plus correct rejections by the total number of trials (change plus control trials)\(^{64}\). Mean percent correct scores are shown in Figure 8.1. A paired samples $t$-test with Age (6 months, 9 months) as the independent variable was performed on the percentage correct

\(^{64}\) For Experiments 5, 6, 7, and 8 no analyses of order effects for which tone was the background sound were performed, due to small cell sizes and/or unequal cell sizes.
scores of all infants in the Conditioning stage. There were no overall differences in percent correct scores for infants tested at 6 months ($M = 75.52\%$, $SD = 12.92$) and then tested again at 9 months ($M = 65.91\%$, $SD = 13.53$), $t(11) = 1.508$, $p = .160$. Thus, the hypothesis that performance on the Easy contrast at 6 months would be significantly better than at 9 months of age was not supported.

![Figure 8.1](image_url)  
*Figure 8.1.* Mean percent correct discrimination of the easy lexical tone contrasts by English-learning infants tested at 6 and 9 months of age. Error bars represent standard error of the mean.

### 8.3.2.2 Subsample of Infants Reaching Criterion

As stated in 8.2.1, all 12 infants reached the criterion for discrimination when they were 6-month-olds, but only seven of the twelve infants reached criterion again when they were 9 months of age. The data for the seven infants who passed the Conditioning stage on both test occasions, were analysed by paired samples $t$-test. There was no significant difference between percent correct scores for infants reaching criterion at 6 months ($M_{6\text{months}} = 74.014\%$, $SD = 6.82$) and 9 months, ($M_{9\text{months}} = 77.586\%$, $SD = 14.34$), $t(6) = .709$, $p = .505$.
8.3.3 Experimental 1 Stage Data

8.3.3.1 Discrimination Performance

The seven infants who reached the criterion for discrimination of the Easy contrast in the Conditioning stage at 6 months and 9 months of age progressed to participation in the Experimental 1 stage. When they were 6 months old they obtained a mean of 73.33% correct ($SD = 13.71$) and as 9-month-olds they obtained a mean of 77.129% correct ($SD = 15.46$). A paired samples $t$-test comparison revealed no significant differences at 6 months and 9 months of age between the mean percent correct scores in the Experimental 1 stage, $t(6) = .456$, $p = .665$.

8.3.4 Performance Across Stages

Percent correct scores in the Conditioning and Experimental 1 stages for all 12 infants when they were 6 months and 9 months old were compared. A paired samples $t$-test for the 6-month-olds confirmed that the mean percent correct for these infants was not significantly different in the Conditioning ($M = 75.52\%$, $SD = 12.92$) versus the Experimental 1 stages ($M = 68.33\%$, $SD = 15.34$) $t(11) = 1.965$, $p = .310$, despite the pattern indicating that they performed better in the earlier Conditioning stage.

Only 7 of the 12 infants completed both the Conditioning and Experimental 1 stages as 9-month-olds. Again, a paired samples $t$-test revealed no significant difference between the performance of the 9-month-olds in the Conditioning vs. Experimental 1 stage, $t(6) = 1.965$, $p = .616$, with the infants obtaining a mean percent correct of 73.56% ($SD = 9.47$) in the Conditioning stage, and 77.13% ($SD = 16.71$) in the Experimental 1 stage.

8.3.5 Summary of Results

There were two main findings in the longitudinal study of lexical tone discrimination by English language environment infants, as follows:
1. Age affected the proportion of infants who reached criterion for discrimination of the [bʌ] vs. [bâ] (Easy) contrast; significantly more infants reached the criterion for discrimination when they were 6 months than when they were 9 months of age.

2. Interestingly, the difference in performance across age as shown by the proportion of infants reaching criterion was not evident in a comparison of mean percent correct discrimination at 6 months vs. 9 months, but the effect was in the expected direction.

8.4 Longitudinal Study of English-learning Infants’ Discrimination of Nonspeech Tone: Experiment 6

While there was no difference in the ability of English-learning infants to discriminate the Easy and Difficult contrasts in the nonspeech cross-sectional study (Experiment 2), to be consistent with the longitudinal study of lexical tone discrimination (Experiment 5) the English-learning infants were only tested for their ability to discriminate the Easy (rising vs. falling) nonspeech tone contrast. Eleven English language-learning infants participated in this study and were tested for their discrimination of the rising vs. falling nonspeech contrast when they were 6 months old and again when they were 9 months old. The same statistical tests were performed on the data in the cross-sectional nonspeech study (Experiment 2), and the longitudinal study of lexical tone discrimination (Experiment 5). Chi-square analyses were performed on data to determine the proportion of infants reaching criterion, and paired sample t-tests were performed on percent correct scores in the Conditioning and Experimental 1 stages. Alpha was .05 for all statistical tests.

8.5 Method

8.5.1 Participants

Eleven infants (6 girls and 5 boys) were tested at 6 months of age ($M = 6.37$ months, $SD = .18$, Range = 6.03-6.82) and again at 9 months of age ($M = 9.22$, $SD = .13$, Range
LONGITUDINAL DATA

= 9.02-9.48). All infants were from monolingual Australian-English speaking families residing in the Sydney metropolitan area. All infants were born full-term or within two weeks either side of full-term with uncomplicated births, were judged to be healthy at the time of testing, and had no known hearing disabilities or history of ear infections. An additional three infants were tested, but were not included in the final sample on the grounds of crying (1) or fussiness (1) and failure to return for testing at 9 months of age (1). Parental consent was obtained prior to testing in accordance with University of Western Sydney ethical guidelines for human participants (see Appendix CD6 for the information hand-out and consent form). Parents were debriefed following the testing session and received remuneration of $20 AUD for travel expenses, a selection of baby products from the project sponsors, and a bib or t-shirt.

8.5.2 Stimuli

The experimental contrast rising vs. falling was one of two nonspeech tone contrasts used in the cross-sectional studies of nonspeech tone discrimination (as in Experiment 2). The rising vs. falling contrast is a synthetic analogue of the Thai [bā] vs. [bâ] lexical tone contrast and its F0 movement is identical to the original [bā] vs. [bâ] contrast. Refer to section 5.5.3 for details of how these nonspeech tones were created.

8.5.3 Materials and Apparatus

Testing was conducted at the MARCS BabyLab, University of Western Sydney. The experimental materials and apparatus were the same as those used in the cross-sectional studies of nonspeech tone discrimination (Experiment 2 and Experiment 4).

8.5.4 Procedure

The CHT procedure controlled by the Exeter Hayes and Slater Windows Conditioned Headturn program, version 2 (1999) was used for testing infants in this longitudinal study.
8.6 Results

Raw data for Experiment 6 and statistical output for the proportion of infants reaching criterion and discrimination performance is presented in Appendix CD9 on the CD supplement.

8.6.1 Proportion Reaching Criterion

As Table 8.2 shows, all eleven infants reached the criterion for discrimination when they were 6 months of age (and were selected for participation in this longitudinal study on this basis). Three months later, at 9 months of age, the infants were retested and 7 out of the 11, reached criterion. A 2 Age (6 months, 9 months) x 2 Criterion (Pass, Fail) chi-square test revealed no significant difference between the proportion of infants reaching criterion for discrimination at 6 months and 9 months, $\chi^2(1, N=11) = .788, p = .449$.

Table 8.2

<table>
<thead>
<tr>
<th>Age</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>11 (100%)</td>
<td>0 (0%)</td>
<td>11 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>7 (63.64%)</td>
<td>4 (36.36%)</td>
<td>11 (100%)</td>
</tr>
</tbody>
</table>

8.6.2 Discrimination Performance

8.6.2.1 Entire Sample

The percentage of correct responses in the Conditioning stage was calculated for each infant (both those reaching and not reaching criterion) by dividing the number of hits plus correct rejections by the total number of trials (change plus control trials). A paired sample $t$-test was performed on percent correct discrimination for the infants at 6
months vs. 9 months of age. Mean percent correct performance was 76.87% (SD = 11.30) on the rising vs. falling contrast when infants were tested at 6 months of age, and was 72.28% (SD = 13.61) when the same infants were tested on the rising vs. falling contrast again at 9 months of age. This difference was not significant, t(10) = .709, p = .505. Mean percent correct scores are presented in Figure 8.2. The results are in accordance with the hypothesis that there would be no decline in the ability to discriminate the nonspeech tone contrasts between 6 and 9 months of age.

![Graph showing percent correct over age](image)

*Figure 8.2.* Mean percent correct discrimination of the easy nonspeech tone contrast by English-learning infants tested at 6 and 9 months. Error bars represent standard error of the mean.

### 8.6.2.2 Subsample of Infants Reaching Criterion

Of the eleven infants reaching the criterion for discrimination of the rising vs. low contrast, only seven reached criterion when retested at 9 months of age. Percent correct scores from these seven infants at 6 and 9 months were entered in a paired samples *t*-test. The comparison of percent correct scores across age revealed no significant differences; the infants achieved a mean of 74.01% correct (SD = 6.80) for their first test session at 6 months, and a mean percent correct of 77.59% (SD = 14.34) for their second test session at 9 months, *t* (6) = .709, *p* = .505.
8.6.3 Experimental 1 Stage Data

8.6.3.1 Discrimination Performance

A comparison of percent correct discrimination of the rising vs. falling contrast was performed for the infants who reached criterion at both 6 and 9 months of age and completed the Experimental 1 stage. A paired samples t-test failed to show any significant differences between percent correct discrimination for infants tested at 6 months ($M = 65.71, SD = 11.19$) and tested again at 9 months of age ($M = 73.33, SD = 14.40$), $t(6) = 1.290, p = .244$.

8.6.4 Performance Across Stages

Percent correct discrimination for the Conditioning vs. Experimental 1 stages was compared by a $t$-test (paired samples) for the infants at 6 months and another $t$-test at 9 months. To be included in these analyses, infants were required to reach the criterion for discrimination when tested at both ages, and complete the Conditioning and Experimental 1 stages ($N=7$). The paired samples $t$-test revealed a significant difference in percent correct discrimination between the Conditioning stage ($M = 76.87\%, SD = 11.30$) and Experimental 1 stage ($M = 67.87\%, SD = 12.94$) for infants tested at 6 months, $t(11) = 3.643, p = .005$, but no difference between percent correct discrimination when the infants were tested again at 9 months ($M_{\text{Conditioning}} = 77.59\%, SD = 14.34$; $M_{\text{Experimental}} = 73.33\%, SD = 14.40$; $t(6) = .711, p = .504$.

8.6.5 Summary of Results

There were two main findings in the longitudinal study of lexical tone discrimination by English language infants, as follows:

1. There was no effect of age on the proportion of infants reaching criterion for discrimination of the rising vs. low nonspeech tone contrast.

2. There were also no significant differences in percent correct discrimination when infants were tested at 6 and 9 months of age.
8.7 Longitudinal Study of Chinese-learning Infants’ Discrimination of Lexical Tone: Experiment 7

In the longitudinal study of the discrimination of lexical tone contrasts, 10 Chinese language environment infants were tested on either the Easy contrast [bâ] vs. [bâ] or the Difficult contrast when they were 6 months of age and were retested on discrimination of the same contrast, 3 months later, at 9 months of age. Five of the Chinese-learning infants were randomly assigned to the Easy condition and five were assigned to the Difficult condition because the comparable cross-sectional study revealed no difference in discrimination of the Easy vs. Difficult contrast. There was counterbalancing between infants of which sound in each pair was the background and which was the target stimulus.

In line with the analyses for the earlier reported experiments, chi-square analyses were performed on data from the Conditioning stage to determine the proportion of infants who reached criterion for discrimination as measured by the 7/8 preset performance criterion. In addition, percent correct scores in the Conditioning and Experimental 1 stages were compared by paired samples t-test. The alpha level was set at .05 for all statistical tests.

8.8 Method

8.8.1 Design

Due to the difficulty with recruiting the sample of Chinese-learning infants, infants in the final sample did not necessarily reach criterion at 6 months before they were tested again at 9 months. This is different to the procedure of the longitudinal study of English-learning infants’ discrimination of lexical tone contrasts (Experiment 5) where it was required that all infants pass criterion at 6 months to be considered in the final data set.
8.8.2 Participants

The participants were ten infants (8 girls and 2 boys) from Cantonese and/or Mandarin speaking families who were tested at 6 months of age ($M = 6.02$, $SD = .35$; Range = 5.48-6.79) and again at 9 months of age ($M = 9.06$, $SD = .21$; Range = 8.6-9.48). All infants were born within 38-42 weeks gestation with reportedly uncomplicated births, were judged to be healthy at the time of testing, and had no known hearing disabilities or history of middle ear infections.

No additional infants were tested. Parental consent was obtained prior to testing in accordance with university ethical guidelines for human participants (see Appendix CD6 for the information hand-out and consent form). Parents were asked to complete a questionnaire on their language use and proficiency at English (see Appendix CD4). Parents were debriefed following the testing session and received remuneration of $20 AUD for travel expenses, a selection of baby products from the project sponsors and a bib or t-shirt.

8.8.3 Stimuli

The experimental contrasts were the Easy and Difficult contrasts that were used in the earlier experiments (1 and 3) on lexical tone discrimination.

8.8.4 Materials and Apparatus

Testing of all infants was carried out at the MARCS BabyLab, University of Western Sydney. The experimental materials and apparatus were the same as those used in Experiment 3.

8.8.5 Procedure

As in the previous reported experiments, the one experimenter CHT procedure (Exeter Hayes and Slater Windows Conditioned Headturn procedure version 2, 1999) was used for testing infants in this longitudinal study.
8.9 Results

Raw data for Experiment 7 and statistical output for the proportion of infants’ reaching criterion and discrimination performance is presented in Appendix CD9 on the CD supplement.

8.9.1 Proportion Reaching Criterion

Ten Chinese language-learning infants were tested at 6 months and 9 months of age. Two of the infants passed the Conditioning stage at 6 months and 8 failed as shown in Table 8.3. When tested again at 9 months, 3 of 10 infants passed the Conditioning stage. A 2 (Age) x 2 (Criterion) chi-square test illustrated that there was no effect of age on whether infants passed or failed the discrimination tasks $\chi^2(1, N=10) = 2.67, p = .606$.

Table 8.3

<table>
<thead>
<tr>
<th>Age</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>2 (20%)</td>
<td>8 (80%)</td>
<td>10 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>3 (30%)</td>
<td>7 (70%)</td>
<td>12 (100%)</td>
</tr>
</tbody>
</table>

8.9.2 Discrimination Performance

8.9.2.1 Entire Sample

The data from 10 infants was entered into a 2 x 2 repeated measures factorial ANOVA with Age (6 months and 9 months) as the within-subjects variable and Difficulty (Easy or Difficult) as the between-subjects variable. When tested at 6 months of age the infants achieved a mean percent correct of 64.65% ($SD = 17.35$) and when retested at 9
months of age it was 61.55% ($SD = 18.99$). This performance difference across age was not statistically significant, $F(1, 8) = .118$, $p = .740$. There was no difference in discrimination of the Easy ($M = 69.3\%, SD = 15.13$) vs. Difficult contrast ($M = 56.9\%, SD = 12.30$) $F(1, 8) = .1123$, $p = .112$ and no interaction $F(1, 10) = .019$, $p = .893$. Percent correct scores at 6 and 9 months of age are shown in Figure 8.3.

![Figure 8.3](image)

*Figure 8.3.* Mean percent correct discrimination of the easy and difficult lexical tone contrasts by Chinese-learning infants tested at 6 and 9 months of age. Error bars represent standard error of the mean.

As only two infants passed criterion when they were 6 months and a different three infants passed criterion when they were 9 months, no statistical tests could be performed on the data for infants reaching criterion in the Conditioning stage, performance in the Experimental 1 stage or to compare performance in the Conditioning vs. Experimental 1 stages.
8.9.3 Summary of Results

The outcome of the chi-square and repeated measures ANOVA tests, confirm the results of the cross-sectional study of Chinese-learning infants’ discrimination of lexical tone:

1. There was no difference in the performance of infants at 6 and 9 months of age

2. There was no evidence of a difference in discrimination of the Easy and Difficult contrasts.

Thus, as hypothesised the Chinese-learning infants show no evidence for a decline in the ability to discriminate tone language contrasts between 6 and 9 months of age. However the small sample of infants reaching criterion warrants cautious interpretation.

8.10 Longitudinal Study of Chinese-learning Infants’ Discrimination of Nonspeech Tone: Experiment 8

In Experiment 8, ten Chinese language-learning infants were tested when they were 6 months of age and again at 9 months of age. Five infants were tested on the Easy rising vs. falling contrast and five were tested on the Difficult rising vs. low contrast. Assignment to each condition was random and there was between-subject counterbalancing of which tone in a pair was the background or target tone. Chi-square tests were employed to analyse the proportion of infants passing or failing to reach criterion for discrimination, and percent correct discrimination in the Conditioning and Experimental 1 stages were subjected to paired samples t-tests. Alpha was .05 for all tests.

8.11 Method

8.11.1 Design

The design was the same as for Experiment 7 (see 8.3.2).
8.11.2 Participants

Ten infants (9 girls, 1 boy) from Cantonese and/or Mandarin speaking families participated in this longitudinal study of nonspeech tone discrimination and were tested at 6 months ($M = 6.06$ months, $SD = .358$, Range = 5.5-6.70 months) and 9 months of age ($M = 9.03$ months, $SD = .216$, Range = 8.60-9.48 months). No additional infants were tested. Infants were recruited from Sydney Australia and tested at the MARCS BabyLab, University of Western Sydney. Parents reported that their infant suffered no health problems at the time of testing and did not have a history of middle ear infection. All infants were born between 38 and 42 weeks gestation, and no complicated births were reported. Parental consent was obtained prior to testing in accordance with University of Western Sydney ethical guidelines for human participants (see CD6 for information hand-out and consent forms), and parents completed a questionnaire on their language background (see Appendix CD4 for the questionnaire). Parents were debriefed following the testing session, received remuneration of $20 AUD for travel expenses and a selection of baby products from the project sponsors, and a bib or t-shirt.

8.11.3 Stimuli

The experimental contrasts were the rising vs. falling and the rising vs. low nonspeech tone contrasts as used for the other nonspeech tone experiments (Experiments 2, 4 and 6).

8.11.4 Materials and Apparatus

Testing was conducted at the MARCS BabyLab, University of Western Sydney. The experimental materials and apparatus were the same as those reported in Experiments 2 and 4.

8.11.5 Procedure

The Exeter Hayes and Slater Windows Conditioned Headturn procedure version 2, (1999) outlined in section 4.2.4.2 was used for testing infants in this longitudinal study
8.12 Results

Raw data for Experiment 8 and statistical output for the proportion of infants’ reaching criterion and discrimination performance is presented in Appendix CD9 on the CD supplement.

8.12.1 Proportion Reaching Criterion

As shown in Table 8.4, three of the 10 infants reached criterion when they were 6 months and 6 of the 10 when they were 9 months old. A chi-square test revealed no significant difference in the proportion of infants reaching criterion for discrimination \( \chi^2(1, N=10) = 1.818, p = .178 \).

Table 8.4

Proportion of Chinese-learning Infants Reaching Criterion for Discrimination of the Nonspeech Tone Contrasts at 6 and 9 Months of Age.

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>3 (30%)</td>
<td>7 (70%)</td>
<td>10 (100%)</td>
</tr>
<tr>
<td>9 months</td>
<td>6 (60%)</td>
<td>4 (40%)</td>
<td>10 (100%)</td>
</tr>
</tbody>
</table>

8.12.2 Discrimination Performance

8.12.2.1 Entire sample

A 2 (Age) x 2 (Difficulty) repeated measures ANOVA was performed on the Conditioning stage data, with age as the within-subjects variable and difficulty as the between-subjects variable. Figure 8.4 shows that there were no effects of Age (\( M_{6\text{months}} = 55.56\%, SD = 14.74; M_{9\text{months}} = 67.26\%, SD = 17.51 \), \( F(1,8) = 2.473, p = .154 \)) or Difficulty (\( M_{\text{Easy}} = 62.9\%, SD = 16.33; M_{\text{Difficult}} = 59.9\%, SD = 17.21 \) \( F(1,8) = .143, SD = .715 \)), and no Age x Difficulty interaction were found for the Chinese-learning infants’ discrimination of nonspeech tone, (\( F(1,8) = .074, p = .792 \)).
Figure 8.4. Mean percent correct discrimination of the easy and difficult nonspeech tone contrasts by Chinese-learning infants tested at 6 and 9 months. Error bars represent standard error of the mean.

8.12.2.2 Subsample of Infants Reaching Criterion

As only three infants reached criterion at 6 months of age and only two of the three reached criterion when retested at 9 months, the sample size was too small to perform statistical analyses on data for infants reaching criterion in the Conditioning and Experimental 1 stages, and for comparisons to be made for discrimination in the Conditioning vs. Experimental 1 stages.

8.12.3 Summary of Results

The longitudinal investigation of Chinese-learning infants' discrimination of nonspeech tone revealed:
1. No differences between the discrimination of the contrasts by infants tested at 6 months and retested when they were 9 months old.

2. The results indicate that Chinese-learning infants discriminated the Easy and Difficult contrasts equally well and that this did not change over age.

These findings support the cross-sectional data in which no Age or Difficulty effects were found for the Chinese-learning infants. The chi-square findings are in line with the corresponding chi-square results in the cross-sectional study (see Experiment 4) with no effect of Age on the proportion of Chinese-learning infants reaching criterion for discrimination.
CHAPTER 9
The Complete Picture
9.1 An All–Inclusive Analysis of Variance

Experimental chapters 5, 6, and 8 have examined the developmental pattern of tone discrimination for the English-learning and Chinese-learning infants separately. The purpose of this chapter, is to analyse the data pertaining to the different cross-sectional studies together in a single Factorial ANOVA, a 2 Age (6-month-olds, 9-month-olds) x 2 Difficulty (Easy, Difficult) x 2 Language (English, Chinese) x 2 Task (Lexical tone, Nonspeech tone). This will enable direct cross-age, cross-language background, cross-tone contrast, and cross-task comparisons to be made. Statistical output for the ANOVA and simple effects are presented in Appendix CD10 on the CD supplement.

9.1.1 A Problem with the Analysis and a Proposed Solution

The first three variables in the 2 (Age) x 2 (Difficulty) x 2 (Language) x 2 (Task) factorial ANOVA are between-subjects factors. However the fourth factor, Task, is problematic. Separate groups of English-learning infants participated in the lexical tone and nonspeech tone discrimination experiments, whereas for the Chinese-learning infants, 32 infants completed both experiments. Thus for English-learning infants this is a between-subjects variable, but for Chinese-learning infants it is a within-subjects variable, and so it is possible that there is a reduced error term across the lexical and nonspeech tasks for the Chinese-learning infants. It is argued that the failure to find order effects, practice effects or performance decrements on the second task (see 6.10) renders a violation of the assumptions of between-subjects ANOVA in this case, tolerable. Specifically, if there are no differences in Chinese-learning infants' discrimination as measured by percent correct, then it is as if these infants act as independent subjects in the two experiments. Furthermore, a Pearson's correlation between percent correct lexical tone discrimination and percent correct nonspeech tone discrimination for the 32 infants who completed both experiments confirms that there is no significant relationship between Chinese-learning infants' performance on the two tasks, $r = -.096, p = .603$. Given these facts it is argued that, while technically illegitimate and acknowledged to be so, it seems reasonable to include the Task factor in this ANOVA as a between-subjects variable, in order to provide a final analysis of
the collected data across age, language, difficulty and task. In light of the problems with this analysis the results should be viewed as indicative of cross-age, cross-language, cross-task tone discrimination, and not considered absolute.

9.1.2 Results of the 2 x 2 x 2 x 2 factorial ANOVA

The 2 Age x 2 Difficulty x 2 Language x 2 Task Univariate ANOVA revealed that there were no main effects of Age, ($M_{6\text{months}} = 60.40\%, SD = 14.46; M_{9\text{months}} = 60.47\%, SD = 15.91$), $F(1,168) = .034, p = .854$, or Language ($M_{\text{English}} = 62.12, SD = 15.6 M_{\text{Chinese}} = 58.69, SD = 14.53$), $F(1, 168) = 2.878, p = .092$. However, a statistically significant difference was found for Difficulty, with percent correct being greater for the Easy ($M = 63.9\%, SD = 14.72$) than the Difficult contrast ($M = 56.88, SD = 14.83$), $F(1,168) = 12.042, p = .001$. The ANOVA also revealed a main effect for Task, with better performance on the nonspeech tone discrimination task ($M = 62.6\%, SD = 14.778$) than the lexical tone discrimination task ($M = 58.15\%, SD = 14.53$). There was also a significant Age x Task interaction ($F(1,168) = 11.523, p = .001$). Six-month-olds performed better on the lexical tone discrimination task ($M = 61.5\%, SD = 14.86$) than 9-month-olds ($M = 54.85\%, SD = 15.006$), and performed similarly on the lexical ($M = 61.5\%, SD = 14.86$) and nonspeech tone discrimination tasks ($M = 59.25\%, SD = 14.08$). Nine-month-olds showed better nonspeech tone ($M = 66.61\%, SD = 14.70$) than lexical tone discrimination ($M = 54.85\%, SD = 15.$). This Age x Task interaction is presented in Figure 9.1. A three way interaction between Difficulty x Language x Task interaction was significant, $F(1, 168) = 4.621, p = .033$. Table 9.1 presents the mean percent correct for English-learning infants’ discrimination of Easy and Difficult lexical and nonspeech tone contrasts. Mean percent correct discrimination for Chinese-learning infants’ discrimination of Easy and Difficult lexical and nonspeech contrasts is presented in Table 9.2. In the next section, analyses of simple effects are performed to demarcate the Difficulty x Language x Task interaction.
Figure 9.1. The age by task interaction collapsed across difficulty and language background. Error bars represent standard error of the mean.

Table 9.1

<table>
<thead>
<tr>
<th>English-learning Infants</th>
<th>Easy</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>65.9%</td>
<td>50.4%</td>
</tr>
<tr>
<td>Nonspeech</td>
<td>69.3%</td>
<td>63.0%</td>
</tr>
</tbody>
</table>
Table 9.2
*Chinese-learning Infants’ Mean Percent Correct Discrimination of the Easy and Difficult Lexical and Nonspeech Tone Contrasts.*

<table>
<thead>
<tr>
<th></th>
<th>Chinese-learning Infants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Easy</td>
</tr>
<tr>
<td>Lexical</td>
<td>57.7%</td>
</tr>
<tr>
<td>Nonspeech</td>
<td>62.9%</td>
</tr>
</tbody>
</table>

### 9.1.3 Simple Effects

Two post-hoc univariate ANOVAs were performed on lexical tone and nonspeech tone data, to analyse simple effects and ascertain the nature of the three way Difficulty x Language x Task interaction found in section 9.1.2. The first ANOVA was a 2 (Age) x 2 (Difficulty) x 2 (Language) analysis for the lexical tone data. Here a main effect of Age was revealed, with discrimination by 6-month-olds ($M = 61.5\%, SD = 14.86$) being superior to discrimination by 9-month-olds ($M = 54.8\%, SD = 15$; $F(1,88) = 5.210, p = .025$). There was also a main effect of Difficulty, with percent correct higher on the Easy contrast ($M = 61.8\%, SD = 14.70$) than the Difficult contrast ($M = 54.50\%, SD = 15$; $F(1,88) = 6.381, p = .013$). There was a Difficulty x Language interaction $F(1, 88) = 8.006, p = .006$. As shown in Figure 9.2 for English-learning infants, the Easy contrast ($M = 65.86\%, SD = 17.01$) was easier to discriminate than the Difficult contrast ($M = 50.4\%, SD = 11.29$) while discrimination of the Easy ($M = 57.64\%, SD = 10.84$) and Difficult ($M = 58.63\%, SD = 17.23$) contrasts was similar for Chinese-learning infants.
Figure 9.2. The interaction between difficulty and language for lexical tone discrimination.

The second analysis of simple effects was a 2 (Age) x 2 (Difficulty) x 2 (Language) factorial ANOVA for nonspeech tone discrimination. It revealed three main effects. First, percent correct discrimination of nonspeech tones by 9-month-old infants ($M = 66.6\%, \ SD = 14.70$) was significantly better than percent correct by 6-month-old infants ($M = 59.25\%, \ SD = 14.07$), $F(1, 80) = 6.383, p = .013)$. Second, discrimination of the Easy contrast ($M = 66.38\%, \ SD = 14.53$) was better than discrimination of the Difficult contrast ($M = 59.47\%, \ SD = 14.36$), $F(1, 80) = 5.719, p = .019$. Third, there was a Language effect $F(1.88) = 5.858, p = .018$ - infants of English language backgrounds could better discriminate nonspeech tone contrasts ($M = 66.14\%, \ SD = 13.91$) than infants of Chinese language backgrounds ($M = 59.08, \ SD = 15.03$).

9.1.4 Summary

The results of ANOVA (9.1.2) and simple effects (9.1.3) are summarised below:

1. Discrimination of lexical tone by 6-month-olds was greater than discrimination by 9-month-olds (collapsed across the Language and Difficulty factors).
2. Discrimination of the Easy (lexical and nonspeech) contrasts was better than discrimination of the Difficult (lexical and nonspeech) contrasts.

3. Discrimination was better for the nonspeech compared to the lexical tone contrasts.

4. 6-month-olds discriminated the lexical and nonspeech tones equally well, whereas 9-month-olds discrimination of nonspeech tone was greater than their lexical tone discrimination.

5. English-learning infants showed superior discrimination of the Easy lexical tone contrast, while Chinese-learning infants discriminated the Easy and Difficult lexical tone contrasts equally well.

6. English language-learning infants were better at discriminating nonspeech tone contrasts than Chinese language-learning infants.

### 9.2 Overview of Main Findings

In this section the main findings of Experiments 1 to 8 are presented. Figure 9.3 highlights the main findings from Experiments 1 through 4 (cross-sectional studies), that is, the mean percent correct discrimination of the lexical and nonspeech tone contrasts by English- and Chinese-learning 6- and 9-month-olds. The results are presented alongside one another for ease of comparison. The main findings pertaining to each experiment are summarised in the subsections below for each experiment.
Figure 9.3. Summary of infants' mean percent correct discrimination in Experiments 1 to 4.
9.2.1 Discrimination of Lexical Tone by English-learning Infants

9.2.1.1 Experiment 1: Cross-sectional Study

The results of the cross-sectional study on English language environment infants’ discrimination of lexical tone indicate:

1. Six-month-old English-learning infants discriminated the lexical tone contrasts better than the 9-month-old infants. The decline in the ability of the 9-month-olds to discriminate the nonnative lexical tone contrasts supports a perceptual reorganisation of lexical tone perception.

2. A greater number of 6-month-old compared to 9-month-old English-learning infants reached the criterion for discrimination of the Easy contrast (although just outside being significant), further supporting a decrease in lexical tone discrimination across age.


9.2.1.2 Experiment 5: Longitudinal Study

The results of the 6- to 9-month longitudinal study with English-learning infants support the cross-sectional data:

1. There was a significant reduction in the proportion of English-learning infants reaching criterion on the Easy lexical tone contrast when they were 6 months compared with 9 months of age.

2. While percent correct was higher for English-learning infants when they were tested at 6 months than at 9 months, this decline in performance over age was not significant.
9.2.2 Discrimination of Nonspeech tone by English-learning Infants

9.2.2.1 Experiment 2: Cross-sectional Study

The results of the cross-sectional study of English-learning infants’ discrimination of nonspeech tone contrasts indicate:

1. There was no statistically significant difference between the 6- and 9-month-old English-learning infants’ discrimination of the nonspeech contrasts both in (a) percent correct scores and (b) the number of infants reaching the criterion for discrimination.

2. Contrary to what was found for lexical tone discrimination, there was no significant difference in English-learning infants’ discrimination of the Easy and Difficult contrasts.

9.2.2.2 Experiment 6: Longitudinal Study

The results of the longitudinal study showed:

1. There was no significant difference between the proportion of English-learning infants reaching criterion at 6 months and again at 9 months.

2. Performance as measured by percent correct discrimination was consistent over age for the English-learning infants, with no significant differences.

9.2.3 Discrimination of Lexical tone by Chinese-learning Infants

9.2.3.1 Experiment 3: Cross-sectional Study

The results of the cross-sectional study of Chinese-learning infants’ discrimination of lexical tone indicate:
1. There was no statistically significant difference between 6-month-old and 9-month-old Chinese-learning infants’ ability to discriminate the tone contrasts, and hence no evidence for a decrease over age to discriminate the contrasts.

2. There was no significant difference in the number of 6- and 9-month-old Chinese-learning infants reaching the criterion for discrimination for the lexical tone contrasts.

3. The Easy and Difficult contrasts were discriminated equally well by the 6- and 9-month-old Chinese-learning infants.

### 9.2.3.2 Experiment 7: Longitudinal Study

The results of the longitudinal study showed:

1. There was no effect of Chinese-learning infants’ age on whether the criterion for discrimination of the lexical tone contrasts was reached. A similar proportion of infants reached criterion at 6 and 9 months.

2. Chinese-learning infants performed similarly at 6 and 9 months of age in terms of percent correct discrimination.

### 9.2.4 Discrimination of Nonspeech Tone by Chinese-learning Infants

#### 9.2.4.1 Experiment 4: Cross-sectional Study

The Chinese-learning infants’ performance on the nonspeech tone discrimination task indicates:

1. No difference between 6- and 9-month-old Chinese language-learning infants’ discrimination of the nonspeech tone contrasts as shown by both percent correct results and the proportion of infants’ reaching criterion for discrimination.
2. Equivalent discrimination of the Easy and Difficult contrasts by Chinese-learning infants.

9.2.4.2  **Experiment 8: Longitudinal Study**

The results of the longitudinal study showed:

1. There were no significant performance differences on Chinese-learning infants’ discrimination of nonspeech tone contrasts over age.

2. There was no effect of age on whether Chinese-learning infants reached the discrimination criterion.
CHAPTER 10
Discussion

‘What we call the beginning is often the end. And to make an end is to make a beginning. The end is where we start from’

(‘Four Quartets’, T.S. Eliot)
Research on infant speech perception began in earnest with the publication of the seminal paper by Eimas, Siqueland, Jusczyk, and Vigorito (1971) in Science. Since this initial account of infants’ categorical discrimination of synthetic consonant-vowel syllables, there has been an explosion of research on how the surrounding linguistic environment shapes infants’ speech perception. The particular focus has been whether infants are able to discriminate different types of phonetic contrasts, both native and nonnative, (Werker & Tees, 1984), and how experience with the native language leads to the development of phonemic categories (Grieser & Kuhl, 1989; Kuhl & Iverson, 1995; Kuhl et al., 1992).

A hitherto neglected area of infants’ speech perception is the discrimination of lexical tone contrasts and the developmental course of tone perception. This thesis has provided an investigation of the influence of linguistic experience (English language and Chinese language environments) on infants’ discrimination of lexical tone and nonspeech tone contrasts.

The research findings are summarised in section 10.1 and frameworks for interpretation are discussed in section 10.2. This is followed by some preliminary musings on how existing accounts of infants’ speech perception might incorporate tone. The final section considers the questions raised by the experiments in this thesis and directions for future research into infants’ tone perception.

10.1 Summary of the Results
In this research program an ontogenetic decline in the ability to discriminate lexical tone contrasts was found for infants who had not been exposed to lexical tone in the surrounding language. More specifically, between 6 and 9 months of age English-learning infants’ ability to discriminate phonologically irrelevant lexical tone contrasts was attenuated. This discovery supports and extends previous research that has established a similar decrement in the second half of the first year for nonnative consonants (Aslin et al., 1981; Best et. al., 1988; Werker et al., 1981; Werker & Tees,
1984a) and nonnative vowel discriminations (Polka & Werker, 1994; Trehub, 1976). The parallel between segmental and tone perception findings suggests that tones could be considered phones, since they are processed similarly to other speech sounds but not to nonspeech.

The development of nonspeech tone perception between 6 and 9 months of age was somewhat different; there was no decline over age in the ability of English-learning infants' to discriminate the contrasts, suggesting that when English-learning infants are released from a linguistic context they can more easily discriminate pitch-based contrasts. Direct comparison of the proportion of 6- and 9-month-old English-learning infants who reached the discrimination criterion in the lexical tone versus nonspeech tone experiments shows that an equal proportion of 6-month-olds reached criterion for discrimination of lexical tone and nonspeech tone, whereas significantly fewer 9-month-olds reached criterion for discrimination of the lexical tone contrasts than the nonspeech tone contrasts. Overall, English- and Chinese-learning infants treat nonspeech tones similarly but diverge on their perception of tones in speech: Tone language infants continue to discriminate tone contrasts in speech at 6 and 9 months equally well, while for non-tone language infants their tone language perception deteriorates over this period. It is here that the relative influence of an infants' ambient language environment in shaping speech perception is most clear.

For Chinese-learning infants who are being raised in a tone language environment, the ability to discriminate tone contrasts was maintained between 6 and 9 months of age for both lexical and nonspeech tones. Furthermore the Chinese-learning infants' discrimination of the Easy and Difficult contrasts was equivalent, even for the lexical tones. These findings are in accordance with an emergence of language-specific speech perception - tone is important to infants learning tone languages and tone discrimination is maintained in both 6- and 9-month-old Chinese-learning infants. English-learning infants also demonstrated less robust discrimination of the more psychoacoustically similar [bâ] vs. [bà] (Difficult) contrast than the more psychoacoustically distinct [bâ] vs. [bà] (Easy) contrast for the lexical but not for the nonspeech tone contrasts.
The nature of the difference in the perception of tone information in Chinese versus English language infants is enriched by consideration of their perception of the Easy and Difficult contrasts. The Difficulty x Language x Task interaction in the final ANOVA (see section 9.1.2, and Figure 9.1), and the analysis of simple effects (see section 9.1.3, and Figure 9.2), aids in the interpretation of these. The analysis confirms statistically that irrespective of age, English-learning infants’ ability to discriminate the lexical tone contrasts deteriorated with increasing difficulty, whereas discrimination of the lexical tone contrasts by the Chinese-learning infants remained constant, regardless of contrast difficulty. Thus, while it might have been expected that the psychoacoustically difficult [bā] vs. [bạ] contrast would be discriminated more poorly than [bạ] vs. [bạ], the Chinese-learning infants’ tone language experience appears to enable them to perform equally well on the Easy and the Difficult contrasts.

Regarding nonspeech tone discrimination, the English-learning infants performed better than the Chinese-learning infants overall. In addition, collapsed across language groups there was better discrimination of the Easy than the Difficult contrast\(^\text{65}\). It is interesting to consider the similarities in the pattern of performance of both language groups on the Easy and Difficult contrasts. On the Difficult nonspeech tone contrast, performance by both groups was comparable, indicating that infants from both language groups treated the difference between the Easy and Difficult nonspeech tones similarly. However, this pattern of performance on the Easy and Difficult lexical tone contrasts is somewhat different for the English- and Chinese-learning infants. Generally, the English-learning infants seem to treat the psychoacoustically easy and psychoacoustically difficult contrasts in speech in a similar fashion to nonspeech - the easy contrasts were easier and the difficult contrasts were more difficult. However, the Chinese-learning infants treated the psychoacoustically easy and the psychoacoustically difficult contrasts equivalently in both the nonspeech tone contrasts and the lexical tone contrasts. This is possibly because in a tone language environment all tones are important and so, as a product of their language experience, Chinese-learning infants are ‘immunised’ against psychoacoustically difficult tone contrasts.

\(^{65}\) Note however that with the language groups analysed separately for nonspeech tone discrimination (Experiments 2 and 4) there is no difference in discrimination of the Easy and Difficult nonspeech tone contrasts.
As the first cross-language investigation of lexical tone perception development by infants, it remains to be seen whether the results obtained here will be confirmed in future studies. It appears that there is a developmental decline in tone discrimination by infants from a non-tone language environment which mirrors the developmental pattern of nonnative consonant (Werker & Tees, 1984a) and vowel (Polka & Werker, 1994) discrimination, but no decline in tone language environment infants’ tone discrimination over the same period. Furthermore, for both tone and non-tone language infants the decline does not occur for psychoacoustically similar nonspeech stimuli. Pending confirmation by further studies, this is the state of our current knowledge. However there are still many things that we do not know about early tone perception, and some of these issues are addressed in the next section.

10.2 Frameworks for Interpretation

To become a mature language perceiver, infants must attune their initial language-general ability to perceive the phonetic variation in the world’s languages to the specific characteristics of the particular language in their environment. In interpreting the present findings, this section draws on some of infants’ developmental achievements - speech perception, tone perception, music perception, word learning, and differential attention to pitch in linguistic and non-linguistic contexts - to place the lexical and nonspeech tone perception findings into perspective. Furthermore in section 10.2.3 some suggestions are made regarding the notion of tone space and how it might be specified.

10.2.1 Segments and Tones

The results of the experiments in this thesis show that there is a parallel between nonnative tone and nonnative segmental discrimination. Discrimination of the Thai tone contrasts investigated here, and the existing evidence for the perception of nonnative segments (see Werker & Tees, 1999) is consistent with the notion that initial sensitivity to language-general speech contrasts is subsequently attenuated and becomes

66 Note that vowel perception development is slightly different, see Polka et al., (2001).
more language-specific. An important theoretical issue for developmental cross-language research is the nature of this language-specific effect in tone perception. One interpretation of the findings is that by 9 months of age, English-learning infants’ experience with English functions to selectively attune their perceptual attention to the collection of speech sounds that convey meaningful distinctions in their native language. As lexical tones are not used contrastively in English, discriminative sensitivity to tone is attenuated for English-learning infants.

A second important issue in tone perception relates to when the ambient language begins to modify perceptual sensitivity to nonnative tones. The studies here clearly highlight that some change in perception of tone occurs between 6 and 9 months for English-learning infants. While this implied developmental locus coincides with the reorganisation of English-learning infants’ perceptual sensitivities to nonnative consonant contrasts (e.g., the Hindi and Nthlakampx languages; Werker et al., 1981; Werker & Lalonde, 1988; Werker & Tees, 1984a), it is possible that the onset of perceptual reorganisation for tones occurs before 6 months of age, particularly since tone is carried on vowels, and experiential influences on vowel perception are evident at an earlier age than are consonant-related changes. For example, Polka and Werker (1994) revealed an effect of language environment on Canadian English-learning infants’ discrimination of German vowel distinctions between 4 and 6 months of age, but complete perceptual reorganisation was not evident until around 10 months. For the moment we may state that for lexical tones, age-related changes in perception are certainly evident between 6 and 9 months, but that these may begin earlier and may continue until later.

An emergence of the phonemic perception of tone raises the issue of the relationship between perception and production for tone language-learning infants. Production studies with infants acquiring Thai (Tuaycharoen, 1977) Cantonese (Tse, 1977), and Mandarin (Li & Thompson, 1977) indicate that tones are produced earlier than segments. What implications might this have for the relationship between tone and segmental perception? It is speculated that if there are differences in infants’ ability to discriminate various tone contrasts, then those that are more difficult to discriminate will be produced at a later age, the difficult to discriminate fricative contrasts (see
Eilers & Minifie, 1975; Holmberg et al., 1977). A comprehensive investigation of tone language perception and its relationship with tone language production is yet to be conducted\textsuperscript{67}.

Until now only Harrison (2000) (see section 3.7.4) has investigated lexical tone discrimination by infants. Harrison’s work and the experiments reported here had distinctly different purposes. Harrison investigated whether infants’ discrimination of synthetic Yoruba tone contrasts was categorical in nature, whereas in this thesis lexical tone discrimination was investigated in a cross-age, cross-language, cross-task design. Nevertheless Harrison’s results are consistent with those found in this thesis in that the tone language-learning Yoruba infants discriminated the tone contrasts better than did the English-learning infants. The results of this thesis extend these results and initiate description of a gradually emerging picture of tone perception development in and beyond the first year of life.

10.2.2 Pitch in Speech

There is considerable evidence that F0-related prosodic modifications in infant-directed speech (IDS) such as elevated pitch and extended pitch contours, contribute to infants’ preferences for listening to IDS over adult-directed speech and for speech with exaggerated over normal pitch information (Cooper, Abraham, Berman, & Staska, 1997; Cooper & Aslin, 1990; Fernald & Kuhl, 1987; Pegg et al., 1992; Werker, Pegg, & McLeod, 1994). There is also evidence that the pitch contours of IDS aid infants’ discrimination of speech sounds (Karzon, 1985; Trainor & Desjardins, 2002). In this light, it is interesting to speculate whether the stimuli were perceived as lexical tone, that is, in the same manner as other segmental variations, or whether they were perceived as variations more akin to those in IDS. This in turn gives rise to the more general question of how the tone stimuli were actually perceived by the infants. In this regard these further issues are of interest: (a) whether discrimination was psychoacoustically based such that better discrimination of the Easy lexical tone contrast is related to its greater psychoacoustic salience than the Difficult contrast (b)\textsuperscript{67}.

\textsuperscript{67} As a point of interest, no such perception-production study exists for segments either, so this too should be addressed.
whether Chinese-learning infants’ ability to discriminate the Thai tones is different to how they may discriminate Chinese tones, and (c) the ‘naturalness’ of these laboratory tasks and their generalisation to the discrimination of tones in real life contexts.

10.2.2.1 Was Tone Perceived as Infant-directed Speech?

Of the two language groups it seems more likely that the English-learning infants would treat the stimuli akin to IDS because their native language has not prepared them for the contrastive use of tones. If this is the case then the pitch contours could assist in the perception of the phonetic segments as Trainor and Desjardins (2002) have found for vowel contrasts. If lexical tone discrimination by the English-learning infants is influenced by the IDS-like nature of the stimuli then it is of interest to note that six months is the age when Australian English and also Thai mothers’ mean pitch in IDS is at its maximum (Kitamura & Burnham, 2003; Kitamura et al., 2002). Six months is also the age at which IDS contours have been shown to enhance discrimination of single vowel tokens with IDS pitch contours (Trainor & Desjardins, 2002), and the age at which the English-learning infants in this thesis best discriminated the tone contrasts. Furthermore, newborn infants can discriminate pitch-accented Japanese bisyllabic high-low words from low-high words (Nazzi et al., 1998), and 1- to -4-month-old infants can discriminate three-syllable utterances only if the second syllable of each is IDS-like (Karzon, 1985).

Australian English mothers’ IDS to 9-month-olds has a markedly lower mean F0 and a much higher pitch range than IDS to 6-month-olds (Kitamura & Burnham, 2003). Thus if IDS, and specifically pitch, really does enhance discrimination of speech contrasts then it is possible that the ability of the 9-month-olds to discriminate lexical tone is attenuated for this reason, and not because the tones are nonnative. However this appears unlikely because the pitch range of IDS is higher at 9 months, and infants presumably respond to this variation. To counter the problem that the lexical tone contrasts may have been perceived as instances of IDS-like syllables, it would be worthwhile inducing a different perceptual set in infants by investigating English-learning infants’ discrimination of lexical tone contrasts when they are pre-exposed to recordings of: (a) English IDS, (b) Thai IDS, and (c) Thai ADS, before participation in
the lexical tone discrimination experiments. It is unclear what predictions to make regarding performance under these three conditions. On the one hand, lexical tone discrimination may be superior following pre-exposure to familiar English IDS than following unfamiliar Thai possibly because infants could be basing their discrimination on intonation rather than lexical tone differences. On the other hand, lexical tone discrimination may be better following Thai because of the pre-exposure to lexical tone in the language. This would be in line with the findings of Maye et al., (2002) who found that after only a few minutes of exposure to the distributional properties of a language, infants’ sensitivity to the phonetic categories of this language were changed. If it was found that pre-exposure to Thai does facilitate tone discrimination, then in keeping with what has been found for vowels (Trainor & Desjardins, 2002), it is possible that Thai IDS would facilitate this more than Thai ADS.

For Chinese language-learning infants the distinctive contours of tones and the distinctive contours of IDS are combined in the speech they hear from their caregivers. It is of interest to speculate whether the IDS contours in tone languages facilitate tone language infants’ acquisition of tone distinctions. In this regard some distinct differences between the mean F0 and pitch range in tone and non-tone language IDS have been found, and these are possibly related to the use of exaggerated F0 to signal tone distinctions in tone languages (Grieser & Kuhl, 1988; Kitamura et al., 2002 ). Grieser and Kuhl (1988) suggested that when Mandarin-speaking mothers express vocal affect, they sacrifice tone identity in order to preserve the higher pitch and exaggerated pitch contours of IDS. However, Kitamura et al. (2002) found that compared to Australian-English mothers, Thai mothers restrict their pitch range to convey correct tone information, and only sacrifice tone identity at the end of sentences. Indeed, language-specific modifications to F0 in tone language-speaking mother’s IDS become evident between 9 and 12 months of age (Kitamura et al., 2002) in the form of less corruption of tones in syllable-final positions; and this happens to coincide with the narrowing of infants’ perceptual sensitivities to the characteristics of ambient phonology. Therefore it is quite possible that a relationship might exist between the nature of IDS to tone language infants of different ages, and infants’ ability to discriminate lexical tone contrasts in IDS. Clearly, further research on this issue is required.
10.2.2.2 Easy versus Difficult Contrasts

In their discrimination of the lexical tone contrasts, English-learning infants' responses may have been based primarily on the low-level acoustic differences in F0 rather than the phonological (ir)relevance of the tone contrasts. English-learning infants' superior discrimination of the more psychoacoustically distinct [bǎ] vs. [bâ] lexical tone contrast, over the psychoacoustically more similar [bǎ] vs. [bà] lexical tone contrast appears to support this explanation. However, if discrimination is psychoacoustically based then there should be no decline of English-learning infants' discrimination performance for the lexical tone contrasts from 6 to 9 months, for such a psychoacoustic mechanism would presumably result in maintained or even improved discrimination over age. Moreover, if psychoacoustically driven, the same pattern of response would be expected for English- and Chinese-learning infants. Thus it appears that a better account for the discrimination of lexical tone at 6 and 9 months of age is that perception is phonemically driven resulting in attenuation of tone discrimination by the non-tone English-learning infants, and consistent performance by tone Chinese-learning infants over age.

10.2.2.3 Chinese Infants; Thai Tones

Chinese infants hear tones in their native language but not Thai tones so discrimination of the Thai tone contrasts may have been more difficult than if the lexical tone contrasts were from the repertoire of Chinese tones. There is evidence that tone language-speaking adults are better at discriminating their own tones than nonnative tones (Lee et al., 1996), so it is quite possible that tone language-learning infants would show a similar pattern of discrimination. In addition, one developmental account of tone perception (Clumeck, 1980) documents that in order to acquire tones, tone language speakers must come to know which tones are used phonemically in their language, and which are not\(^{68}\). Presumably then, the tone language-learner must be able to

\(^{68}\) Contrary to this account, Burnham et al. (2003) have shown that neither Thai speaking children nor tertiary educated adults demonstrate explicit awareness of the tones in their language, and so it is assumed that this is also the case for infants.
discriminate successfully between native tones in order to produce them, and this language-specific tone perception may well begin in infancy. An experiment that investigates tone language-learning infants’ discrimination of native and nonnative tone contrasts might expect to find language-general discrimination by young infants, that is, no difference between discrimination performance for native and nonnative lexical tone contrasts, with an emergence of language-specific discrimination expected during the second half of the first year post-partum involving attenuation of nonnative tone discrimination performance.

10.2.2.4 Ecological Validity of the Task

In addition to the presentation of nonnative tones, the context of presentation may also have impeded the Chinese-infants’ tone discrimination. Specifically, presenting tones in isolation without familiarising the listener with the speakers’ pitch range is an unnatural speaker-listener relationship - there is evidence that even adult tone language speakers have difficulty correctly identifying lexical tones in isolation (Abramson, 1986). Consequently, in order for infants to accurately gauge the speaker’s pitch range, future studies of tone perception should provide context\(^6^9\), that is, infants should be presented with short recordings of the speaker’s voice prior to the commencement of the discrimination task (Burnham et al., 1996).

This opens the issue of how relevant a two-stimulus tone discrimination task is to natural tone language processing. At one level, the perceptual reorganisation for English-learning infants is clear; tone is not functional in English so perceptual sensitivity for lexical tone distinctions is attenuated. However at another level, the performance of the Chinese-learning infants on this simple discrimination task possibly cannot be generalised to account for infants’ perception of tones in connected speech, in which there is a complex influence of individual speakers’ pitch range, tone sandhi, sentence intonation (Abramson & Svastikula, 1983; Luksaneeyanawin, 1984), tone downstep (Anderson, 1978). coarticulation (Abramson, 1975), and most importantly for

\(^{69}\) In fact the same criticism can be made for phonetic discrimination experiments generally, especially with regard to normalisation of and familiarisation with vowel space.
the infant language-learner, infant-directed speech (Grieser & Kuhl, 1988; Kitamura et al., 2002; Liu et al., 2003).

In a similar vein, real-word communication involves face-to-face interaction between speaker and listener, and even infants can match phonetic information in the face and voice (Patterson & Werker, 1999, 2002; Werker, Corcoran, Fennell, & Stager, 2002). Moreover, recent research by Burnham, Ciocca, Lauw, and Stokes (2000) indicates that there is visual information for lexical tone which is used by adult perceivers when identifying tones and discriminating tone contrasts, and Erickson (1974) has suggested different cricothyroid muscle movements for each tone. An associated issue is the possible presence of tone-specific facial expressions. Chong, Werker, Russell, and Carroll (2003) have found that there are three dominant emotional expressions in IDS and these are remarkably similar to the facial shape involved in production of the corner vowels /a/, /i/, /u/ in most human languages. Since tone is carried on vowels, there may also be tone-specific facial expressions. Thus it is quite possible that mothers’ speech will contain tone-specific facial expressions and that infants, particularly tone language-learning infants, could be sensitive to the presence of these features.

10.2.3 Specifying Tone Space

The device of vowel space, in which particular vowels are plotted in terms of their F1 and F2 values, is a valuable tool in describing and differentiating vowels. The issues raised in sections 10.2.2.1 through 10.2.2.4 point to three important questions about what might be called tone space:

1. How might tone space be specified?

2. How might the possible acoustic-articulatory relationships be reflected in tone space?

---

70 The same argument for visual information can apply to the real-world relevance of two-stimulus segment discrimination tasks.
3. How might the features/characteristics of this tone space relate to the development of tone perception?

The term *tone space* is used here to refer to a schematic representation of tones in which tones may be plotted and distinguished via acoustic-phonetic and/or articulatory-physiological means. In a vowel space, vowels are plotted in a two-dimensional plane with F1 on the abscissa and F2 on the ordinate, and these phonetic-acoustic features relate to the front-back position of the tongue, and tongue height respectively. In this section the issue of which features of tone should be used to plot tones in space is considered, along with the implications of the positioning of tones in space in such a manner.

As stated in section 1.3.2.2, tone is more than simply F0. It is comprised of many cues, including F1 and F2 features, voice quality, F0 slope, and duration (tones are realised over time), over and above F0, and is typically described in terms of starting, middle and endpoint F0 values. Moreover, these acoustic cues are used to varying extents by tone language users in discriminating between tones and identifying tone classes (e.g., Gandour & Harshman, 1978; see also section 3.4.1). Given these various cues, it would not be informative to plot tones on the basis of only F1 and F2 values (as vowels are). In addition to the omission of significant cues, F1/F2 plots are based on the notion that vowels are steady states, and it is quite clear that temporal information is vitally important in tone perception (Liang, 1963). One possibility might be to include F0 starting point, a perceptual cue important in tone discrimination (Abramson, 1962; see also section 4.1.3), along with some measure of F0 slope or F0 variation over time.

Moreover, given that tones are carried on vowels, and many acoustic features of tone are vowel-related (such as height, duration, voice quality, and F1 and F2 values), should tone space *incorporate* these vowel features, or should it *ignore* them and only plot informative F0-based features of tone? At the very *least*, tone space should be three- rather than two-dimensional (as vowel space is) to include some measure of F0 movement over time.

The notion of an acoustic-articulatory relationship for tones plotted in space is interesting, and one that tone researchers should consider. Given that the cricothyroid
muscle contracts for rising F0 and extends for falling F0 (Erickson, 1974), it is possible that some specification of acoustic-articulatory connections is possible in tone space. While full consideration of this issue is beyond the aims of this thesis, now that the first steps have been taken to investigate infants’ tone perception, it is certainly of interest to consider how tone perception and production may be linked and how the perceptual and/or articulatory distinctiveness of tones may be represented and coded.

Another relevant issue regarding tone space concerns infants learning two tone languages with different tone systems. Linguistic input for these infants would be replete with tone distinctions. If it were the case that the resulting tone space was densely populated (and if this density could be coded and measured), this could have interesting implications for which tone contrasts are discriminated more or less easily. Moreover, it is also possible that some tone distinctions are perceived continuously and some are perceived categorically (either within or across languages), perhaps depending upon relative pitch heights and degree of pitch movement, and that relative exposure to and experience with a tone contrast would influence its discriminability.

A series of studies now converge on the notion that vowel hyperarticulation in IDS is a didactic tool that emphasises the identity of vowels in the particular language environment (Burnham et al., 2002; Grieser & Kuhl, 1989; Liu et al., 2003). Similarly, the degree of hyperarticulation for tones in tone language IDS may serve the didactic purposes of directing infants’ attention to salient features of tones, and emphasising the specific nature of the tone space in a particular language. In this regard, a method for the location of particular tones within a tone space would be a valuable heuristic tool.

A final point of interest generated by consideration of the tone space issue is the notion of initial sensitivities or biases (asymmetries) for particular tone types. In the realm of vowel discrimination with 6- to 8- and 10- to 12-month-old infants, Polka and Werker (1994) found that vowels that are more typical of vowels in an infants’ native language act as a perceptual attractor for other vowels, a language-specific effect. Later work by Polka and Bohn (1996) revealed that there appear to be universally favoured vowels — that irrespective of the language experience of the listener, relatively more peripheral vowels act as perceptual attractors for less peripheral vowels such that peripheral
vowels are more easily discriminated from more central vowels rather than the other way around. With respect to tone, Francis and Ciocca (2003) have shown that Cantonese-speaking adults have language-specific tone perception asymmetries that result in better discrimination of low-high F0 sequences than high-low F0 sequences. In this regard and due to the counterbalancing of stimulus order in the present experiments, it is possible to investigate whether such (or similar) asymmetries occur in infants’ tone perception. Percent correct discrimination for the Easy rising-falling tone contrast in two orders, rising-falling and falling-rising, and the Difficult contrasts in two orders, rising-low and low-rising, are shown below in Figures 10.1 and 10.2 respectively. While there were no statistically significant differences between orders (see sections 5.3.2.1 and 6.3.2.1), that is, no asymmetries, the pattern of performance for the two language groups is interesting. In Figure 10.1, it can be seen that both the English and Chinese-learning infants discriminated rising-falling better than falling-rising, suggesting a possible ‘universal’ language-general bias which would place the falling as the more ‘peripheral’ of the two tones. However this apparently universal effect is not found for the Difficult contrast. Figure 10.2 shows slightly better discrimination of low-rising compared to rising-low for Chinese- but not English-learning infants, a possible language-specific effect. Further research on the development of tone asymmetries is required. Given the evidence for language-specific asymmetries in tone perception by adults, it would be of interest to determine whether there are asymmetries in infancy or childhood, whether they are similar to vowel asymmetries, how the caregiver speaking to the infant may shape tone asymmetries, and whether any asymmetries are maintained across development from infancy through to adulthood.
Figure 10.1. Asymmetries in English- and Chinese-learning infants' discrimination of the easy lexical tone contrast.

Figure 10.2. Asymmetries in English- and Chinese-learning infants' discrimination of the difficult lexical tone contrast.
DISCUSSION

Future research is required to address the issues raised here regarding the nature of tone space, acoustic-articulatory relationships for tone, tone asymmetries, and the influence of all three of these on tone perception in general, and infants’ tone perception in particular.

10.2.4 Pitch in Music

In this section two important questions are addressed: (a) is pitch perception absolute or relative? and (b) is tone perception dependent on absolute pitch?

As outlined in section 3.7.2 infants can distinguish different pitch relations in musical stimuli, can categorise stimuli on the basis of pitch contours (Thorpe, 1986), and can discriminate music played in different keys or with different pitches (Trainor & Trehub, 1992). The results of the nonspeech tone discrimination experiments here (Experiments 2, 4, 6 and 8) provide further evidence that infants are sensitive to pitch changes in non-linguistic contexts. The studies by Thorpe (1986), Trainor and Trehub (1992) and the nonspeech studies here demonstrate that infants can process the relative pitch of notes and/or tones. However, infants have shown that they are also sensitive to absolute pitch differences between tones. Saffran and Griepentrog, (2001) demonstrated that infants’ discrimination of changes in the absolute pitch of bell-like sequences of tones is better than their discrimination of changes in relative pitch. However adults were better at discriminating changes in relative than absolute pitch.

An interesting parallel can be drawn between the developmental patterns in the study by Saffran and Griepentrog (2001) and the pattern found here for lexical tone. Saffran and Griepentrog (2001) found a shift from general pitch processing in infancy to more functional pitch processing in adulthood, suggesting a developmental reorganisation some time in between. Saffran and Griepentrog argue that unless one is a tone language speaker, it is not advantageous to understand subtle pitch differences between otherwise similar sounding words. This position implies a critical period hypothesis that pitch perception is absolute at birth and continues to be so for tone but not non-tone language environment infants. There are two implications of this argument for speech perception in general and lexical tone perception in particular. Generally, the selective
perceptual attunement of musical pitch processing from the general to the specific is comparable with attunement during the first year of life to native language-specific speech sounds. More specifically, based on Saffran and Griepentrog’s argument, it follows that perceptual reorganisation would only be evident in English-learning infants’ discrimination, because absolute pitch is not functional for non-tone language learners (as implied by the results of the present experiments). However, this does not mean that speakers of a tone language rely exclusively upon absolute pitch. It is argued here that tone language speakers’ tone perception is not dependent on absolute pitch. In fact their perception is more likely relative, given the difficulty they have in identifying (or labelling) a tone without reference to other tones or the speakers’ pitch range (Abramson, 1977; Francis et al., submitted), but their production is more likely to be based on absolute pitch (see Deutsch et al., 1999; 2004; Peretz et al., 2003). Nevertheless, at one level it must be acknowledged that this argument is false, for tone language experience is not a necessary condition for absolute pitch production - speakers from non-tone language speaking backgrounds show absolute pitch in enunciating words under certain conditions.

In summary, infants may have general sensitivities to absolute pitch distinctions and these sensitivities may become functional (i.e., relative) for infants, depending on the linguistic environment in which they are reared. However, the perceptual evidence suggests that in identifying tones, tone language speakers rely most upon the relative rather than absolute pitch distinctions; and in terms of production there are instances where both tone and non-tone language speakers show absolute pitch. Thus it cannot be definitively concluded that tone languages and absolute pitch go hand in hand.

10.2.5 Tone Acquisition and its Relationship to Word Learning

In this section, the implications of the current results for word learning in tone languages are discussed. It has been found that infants fail to discriminate previously discriminable phonetic distinctions when a perceptual task also involves word learning: Fourteen-month-old infants ignore various small differences between syllables, for example the place of articulation contrast in [bi] and [di] when learning new object-word pairings in the Switch paradigm even though they can perceive this distinction in
DISCUSSION

a standard minimal pair phonetic discrimination task (Pater et al., in press; Stager & Werker, 1997; Werker & Stager, 2000), and can discriminate in a word learning task when all phonetic segments differ, as in ‘lif’ and ‘neem’ (Stager & Werker, 1997). Stager and Werker (1997) attribute this to an added cognitive load in the word learning task that interferes with infants’ attention to phonetic detail. By 17 months, and in concert with increases in vocabulary size, infants overcome this limitation and associate phonetic differences with changes in word meaning.

It may be speculated that similar minimal pair effects with tone would not result in failure on a word learning task for tone language infants in the same way as non-tone distinctions. There are usually three to six lexical tone contrasts to learn in a tone language, and evidence from production studies shows that tones are acquired well before full segmental production, possibly because there are less tones to learn. Tse’s case study (1977) showed word learning for a 10-month-old infant on the basis of tone distinctions but not on the basis of segmental distinctions. This single case study evidence (especially if backed up by more extensive studies) may be taken to imply that for infants learning a tone language, tones are more perceptually salient and encoded earlier than consonants and vowels. It also suggests that especially in the early stages of word learning around 14 months, tone language-learning infants may learn word-object pairings more reliably if there is a change in tone as opposed to a change in phone. To investigate this issue more systematically, Chinese- and English-learning infants could be tested on word-learning and tone perception using both the Switch paradigm and a perceptual discrimination paradigm with the following conditions: (a) Phone + Tone change condition: Habituate infants to [bʌ] paired with object 1, and [pʰʌ] paired with object 2, then test their ability to recognize a switch in the sound-object pairing, (b) Phone change: Habituate infants to [bʌ] tone paired with object 1, and [pʰʌ] paired with object 2, then test their ability to recognize a switch in the pairing, and (c) Tone change: Habituate infants to [bʌ] paired with object 1, and [bʌ] paired with object 2, then test their ability to recognize a switch in the pairing. In the word learning switch task Chinese infants would be expected to discriminate a switch based on a tone change more readily than a change in phone, and English-learning infants would be expected to discriminate all
discriminations poorly. On the other hand in a perceptual discrimination task, Chinese infants would be expected to discriminate in all three conditions, and English-learning infants only the Phone + Tone and Phone change conditions.

10.2.6 Speech is Special: Tone versus Pitch

The results of Experiments 1 and 2 indicate that English language infants discriminate tone better in a nonspeech than a speech context and that as infants’ experience in an English language environment increases, their perception of non-phonemic tone is attenuated in speech but not in nonspeech. These results are consistent with the ‘speech is special’ argument (see section 1.5.3) and it could be argued that these results are more compelling evidence for the special status of speech than the parallel findings for segmental distinctions, because of the involvement of F0.

Although the F0 characteristics of the nonspeech stimuli and speech were identical, English-learning infants showed an age-related decline for discrimination of the speech but not the nonspeech stimuli. On this basis it may be concluded that English-learning infants perceived the lexical tone stimuli as speech (or at least linguistically or experientially relevant) but did not perceive the nonspeech tones (or pitch stimuli) as speech. This could imply separate processors for speech and nonspeech or alternatively it could mean that in nonspeech pitch-related cues have greater perceptual weighting for infants. It seems then that phonetic contextual features (consonants and vowels) are necessary for tone to be perceived in a lexical manner. These results are consistent with those of Burnham et al. (1996) who found that English speaking children and adults’ perception of pitch variations in Thai tones significantly improves when pitch contrasts are presented in filtered speech or as violin sounds. Thai and Cantonese speakers on the other hand, discriminated the tone contrasts in speech, filtered speech, and violin sounds equally well.

While the behavioural evidence for a dissociation of lexical tone and nonspeech tone discrimination processing by English language-learning infants is quite convincing, investigating both tone and non-tone language-learning infants’ neural activity via ERP in response to tone in different contexts could provide converging evidence for
differences in infants’ processing of lexical and nonspeech tone. Several studies with adults have shown preferential activation of left hemisphere speech centres for tone language speakers’ when processing tone (Gandour et al., 1998; Klein, Zattore, Milner, & Zhao, 2001), and right hemisphere activation consistent with pitch perception for non-tone language speakers (Klein et al., 2001). A differential outcome of brain activity in location (and possibly intensity) for tone and non-tone language learning infants would support the view that linguistically relevant properties of tone determine the manner in which infants perceptually encode these.

Of additional interest is whether very young infants from tone language environments would show neural activation (via fMRI, Mismatched Negativity ERPs) to tone and non-tone stimuli that is akin to tone language speaking adults. If the pattern of activity between infants and adults were different, then investigating the age at which infant and adult neural activity aligns, would shed further light on the ontogeny of lexical tone perception. If young infants process the tone speech sounds in a language-general fashion, then it would follow that young non-tone language learning infants would exhibit patterns of neural activation to lexical tone stimuli that are similar to those of tone language-learning infants. Then as the non-tone language-learning infants mature, possibly some time between 6 and 9 months, neural activation in the regions of the brain associated with speech processing would likely decline and concur with regions activated by non-linguistic pitch perception.

10.3 Fitting Tone into Accounts of Speech Perception Development

Over half the world’s population speaks a tone language, yet few studies have investigated infants’ tone perception development, and no current theory or model incorporates tone into accounts of infant speech perception development. Current theories of infants’ speech perception are segment focused, however tone may fit into existing models/theories of speech perception for three reasons: (a) tones and segments are similar in that they both function to distinguish word meaning (see also tone as segment argument, section 3.6.4.1), (b) the current state of knowledge about the pattern of development for tone perception during the infants’ first year, parallels the pattern of development for segmental perception, and (c) the results of the studies reported here
show that the developmental course of tone perception is more similar to segmental perception than to nonspeech perception. Therefore the aim of this section is to provide some preliminary suggestions on how tone may be included into existing models/theories of infant speech perception.

The discovery of an age-related change in perceptual tuning to the characteristics of lexical tone between 6 and 9 months implies that the timing of the perceptual reorganisation is remarkably similar for tones and segments. That is, between 6 and 9 months, infants access fine details of the phonetic and tonetic units that make up words in their language and form phonemic and tonomic classes of sounds. This perceptual tuning prepares infants for the next linguistic development, the application of these phonemic (and presumably tonomic) classes to later word learning and finally, word production.

10.3.1 The Epigenetic Model

Werker, Lloyd, Pegg, and Polka (1996) posit that the changing relationship between infants’ perceptual sensitivities and their attention to regularities in the linguistic input is a function of age, ambient language environment, the developing brain, the infant’s changing perceptual and linguistic abilities, and changes in domain-general as opposed to linguistic/phonetic cognitive abilities. One implication of the Werker et al. model (1996) is that the greater exposure infants have to linguistic input the more quickly should they ‘tune in’ to the various properties of native language phonology. Nevertheless, despite the variability in the input that could potentially produce variation in the age of onset of language-specific speech perception, there is remarkable regularity in the age of language-specific speech perception onset. Let us now look at the results of the present experiments to see how experience with English, Chinese, or both, predicted infants’ tone discrimination.

In the experiments conducted here with English-learning infants, over 90% of 6-month-olds versus 58% of 9-month-olds discriminated the lexical tone contrasts to criterion in the cross-sectional and longitudinal studies combined. It can be argued that this is good evidence for a consistent age of onset for language-specific speech perception. On the
other hand, 42%\textsuperscript{71} of 9-month-old Chinese-language infants failed, and only 52%\textsuperscript{72} passed the task. It could be argued that this hardly represents regularity in performance. However, as older infants have not been tested it is possible that the change from language-general to language-specific speech perception is not yet complete at 9 months. Converging evidence that infants' lexical tone discrimination capacities are shaped not purely by the quantity of language experience comes from the perspective of the native listener. The questionnaire data indicates that Chinese-learning infants have varying degrees of exposure to tonal linguistic input, (range 30-80% Chinese input) and `quantity' of Chinese language experience failed either to predict or correlate with performance on the lexical tone discrimination task, suggesting that perceptual biases may be shaped by factors other than quantity of input.

It has been proposed that statistical learning via sensitivity to distributional regularities in the language input could be the means by which phonetic categories are tuned during the first year of life. Maye et al. (2002) exposed infants to at least four voiced (native to English) and four voiceless unaspirated (nonnative to English) stop exemplars from a voicing continuum in a language input exposure learning stage. Half the infants were given greater exposure to both voiced and voiceless exemplars corresponding to a two voicing category language, and the other half were exposed to mostly voiced exemplars. Infants previously exposed to both voiced and voiceless stimuli were able to discriminate stimuli from across the continuum as two phonetic categories, whereas infants exposed to only voiced exemplars treated the whole continuum as a single phonetic category. Thus, familiarisation with phonetic distinctions and the relative distribution of these distinctions, results in the creation of perceptual categories in infancy. This evidence that exposure to distributional regularities can so easily change perceptual categories seems at odds with the position that quantity of input cannot fully account for infants' perceptual sensitivities.

As Werker (2003) asserts, the creation of phonetic categories through the manipulation of distributional regularities in input has important implications for bilingual-language

\textsuperscript{71} This value is the approximate percentage of 9-month-old Chinese infants that \textit{failed on average}, across the lexical tone cross-sectional and longitudinal experiments.

\textsuperscript{72} This value is the approximate percentage of 9-month-old Chinese infants that \textit{passed on average}, across the lexical tone cross-sectional and longitudinal experiments.
environment infants. Equal exposure to phonetic contrasts in two languages is predicted to maintain perceptual categories in both, and exposure to one language more than the other should result in greater sensitivity to phonetic distinctions in the most commonly heard language. In the case of infants learning a tone language concurrently with a non-tone language (like many Chinese infants tested in this thesis), an unequal balance of tone and non-tone language input should have the effect that at any given age, there will be greater perceptual sensitivity to the contrasts of one language to the detriment of the other (in so far as they conflict). Thus it would be interesting to investigate what might happen to the perception of English consonant and vowel contrasts given that tone contrasts are being learned in Chinese. Would there be any spill-over effect, or would the fact that tone contrasts were being learned simply have no effect on learning contrasts in a second non-tone language?

10.3.2 Native Language Magnet Theory

The Native Language Magnet (NLM) theory of vowel perception development (detailed in section 2.8.1.3) may be adapted to explain tone perception (a) given that tones are carried on vowels, and (b) the evidence that tones tend to be perceived continuously like vowels (Abramson, 1961, 1977). Kuhl et al. (1992) provide evidence for the existence of native perceptual vowel magnets by 6 months of age. In the NLM theory it is claimed that from early in life (around 4 months of age) there is an organisation of vowel space, with some regions more stably represented and more discriminable than others. With repeated experience with different instances of vowels, infants learn about the fine-grained features of native language vowel structure and begin to organise native language vowel instances around prototypical categories. The most frequently heard vowels become ‘perceptual magnets’, such that sounds close to them in phonetic/acoustic space are drawn towards them, making it difficult to discriminate between the magnet (prototype) and nearby sounds. Testing the existence of a perceptual magnet effect for tones might unveil a similar effect, although this would depend upon a satisfactory representation of tones in an acoustically/articulatory relevant tone space (see section 10.2.3). Given the organisation of tones in tone space,

73 Note here that there are also pre-existing preferences, (i.e., asymmetries in vowel perception). See sections 2.7.2 and 10.2.3
the most frequently heard tones could serve as prototypes or attractors, and similar sounding instances of tones to these prototypes would be more difficult to discriminate than tones further away from the prototype. Instances of tones that fall equally between two perceptual prototypes (which is possible if it is accepted that tones are perceived continuously), would not be drawn towards either prototype, but would be equally discriminable from the two prototypes.

This section has considered tone perception in light of NLM theory, and there seems to be potential for this theory to account for tone perception.

10.3.3 *Gestural Phonology*

There may also be a place for tones in the gestural phonology account of speech perception. In gestural phonology theory it is claimed that perceivers must recognise the articulatory-gestural properties of ambient speech in order to perceive speech (see 2.8.1.2). The argument that gestural information provides a basis for infant speech perception is supported by evidence that infants can match visual and auditory speech patterns in bimodal perception studies (e.g., Kuhl & Meltzoff, 1982; Patterson & Werker, 1999) and that infants integrate auditory and visual information in the McGurk effect⁷⁴ (Burnham & Dodd, 1996; Desjardins & Werker, 1996; Rosenblum, Schmuckler, & Johnson, 1997). The notion that the gestural information inherent in auditory and visual signals might be available from tone speech comes from Burnham et al. (2000) who found that tone language speakers use visual perception for tone in a degraded listening environment. Thus if it were found that infants can match and integrate visual and auditory speech in bimodal perception of tone as well as for segments, then this would be evidence that infants perceive gestural properties for tone language learning.

Another model that deals with gestural information is the Perceptual Assimilation Model (PAM), and a possible place for tone in this model is proposed below.

---

⁷⁴ The examples given here and later in 10.3.4 for the role of ‘gesture’ in speech perception have mostly visual connotations, but ‘gesture’ is in fact an *amodal* term, that can equally mean the visual, auditory, haptic, kinaesthetic etc cues for speech perception.
10.3.4 Perceptual Assimilation Model

The specification of a phonetic-articulatory relationship for tone (see section 10.2.3) could provide a conceptual space for tone in Best’s Perceptual Assimilation Model (PAM) (Best, 1994; Best, 1995b). Best (1994) proposed that native-language influence on the perception of nonnative phonetic perception takes place when older infants begin to recognise relationships among native phonemes and specific articulatory gestures, (e.g., the gesture for bilabial /b/). Let us assume that a phonetic-articulatory relationship exists for tone, that is, that there are laryngeal and facial gestures that distinguish a rising from a falling tone, and consider how tone- and non-tone language-learning infants in the experiments in this thesis might have perceived tone distinctions in relation to their native phonological categories. It would appear that sometime between 6 and 9 months Chinese language-learning infants are beginning to develop phonemic categories for the tones in their native language (i.e., Cantonese and/or Mandarin). The ability of the Chinese-learning infants to discriminate the two members of a nonnative Thai tone contrast would be related to the pattern of assimilation of nonnative tone to native tone categories. The English-learning infants would not have or be developing native toneme categories because tone is not used contrastively in English. Therefore it might be suggested that English-learning infants have just one single ‘tone category’, and that all tones in nonnative tone contrasts would be assimilated to this single category. Thus according to Best’s PAM their ability to discriminate the two tones would be based on the degree of acoustic similarity/difference between the two tones.

Alternatively, it might be more fruitful to consider both the tone and vowel characteristics of a tone (and non-tone) sound in concert. In this sense, English language listeners would have a single /a/ + tone category, while Thai would have five /a/ + tone categories. However because tone is also specified by F1 and F2 the quality of the [a] in each /a/ + tone category may be slightly different. Then the degree of discrimination of the two tones in a pair would depend on the degree of phonetic similarity/discrepancy either of the Thai tones carried on the vowel [a] shares with the native English [a] category.
Which pattern of perceptual assimilation to native toneme/phoneme properties can best account for the English- and the Chinese-learning infants’ discrimination of the Thai tone contrasts? It is speculated that the Chinese-learning infants’ assimilation is a Two Category (TC type) pattern, on the basis that the Thai rising versus falling, and rising versus low phonemic tones are acoustically (e.g., F0 trajectory, vowel length) as well as gesturally distinct, and therefore assimilated to separate native (Chinese) toneme categories. That is, each tone of the contrast is a clear member of a different tone category.

There are several possible assimilation outcomes for the English-learning infants. Firstly, for the 6-month-old infants, discrimination of the [bǎ] vs. [bâ] lexical tone contrast was good, but discrimination of the rising versus low contrast was near chance levels. It is possible that the English-learning infants’ lower performance on the [bâ] vs. [bǎ] contrast is linked to their perception of similar phonetic- articulatory features of tones in the contrast. A logical possibility for the older 9-month-old English-learning infants with attenuated discrimination of the nonnative lexical tone contrasts is that they are demonstrating Category Goodness (CG type) assimilation for [bǎ] vs. [bâ], and Single Category (SC type) assimilation for [bǎ] vs. [bâ]. The CG discrimination of [bǎ] vs. [bâ] is likely to assimilate to the English syllable [ba], although both may be heard as atypical exemplars that differ in their discrepancy from the ideal native exemplar of [ba]. Whether [bǎ] or [bâ] is a better member of the English [ba] category cannot be ascertained from the results of the present experiments, but nevertheless it is an interesting question for future research. The assimilation of [bǎ] versus [bâ] is suggested to be SC because there is a tendency for poorer discrimination of SC than CG contrasts, and there was significantly poorer discrimination of [bǎ] vs. [bâ] than [bǎ] vs. [bâ]. It is possible that the tones of the nonnative [bǎ] vs. [bâ] pair were assimilated to a single native category [ba], that is, perceived as being equally similar to the native English exemplar of [ba].
In the above considerations it would be even better to try to base this choice of Single Category for rising and low Thai tones on their relative F0 contours or articulatory similarities with regard to Cantonese and Mandarin tones, rather than simply stating that [bǎ] vs. [bà] is SC because it is not discriminated as well. Experiments where Cantonese- and Mandarin-learning infants are tested for discrimination of native tone contrasts are essential to address this issue.

In conclusion, while application of the PAM to tone perception is promising, further research will be required to assess directly whether tone discrimination can be interpreted in terms of PAM assimilation patterns, and if so, what assimilation patterns account for the various tone distinctions.

10.4 More Questions and Future Directions

This chapter so far has attempted to account for the nature of English-learning infants initial state and subsequent age-related change in lexical tone discrimination, as well as the preservation of Chinese-learning infants’ tone perception abilities over age. The differences between tone perception in language, and nonspeech tone perception, have also been addressed and questions have been raised about various aspects of tone development. However, there are a number of questions pertaining to Chinese infants’ performance on the tasks that are even more challenging to resolve.

One surprising element of the Chinese data is that percent correct lexical tone discrimination did not exceed 60% for either the 6- or 9-month-old infants. For the English-learning infants, discrimination exceeded 70% correct (Easy contrast) for 6-month-old infants, and dropped to 60% correct at 9 months of age. Percent correct for Chinese infants was not as high as what might be expected for native listeners with attuned perception. Four possibilities are suggested for this lower than expected performance of the Chinese-learning infants:

1. For tone but not non-tone language-learning infants, tone is one extra speech sound in the input to learn concurrently with vowels and consonants, and perceptual attention to these features may be divided in the online
processing of the speech stream. Similarly in processing vowels, non-tone language learning infants must process F1, F2, F3 etc but tone language learning infants must also process F0\textsuperscript{75}. Therefore to discriminate lexical tone contrasts, Chinese-learning infants are possibly processing all the cues for tone including duration, length, and slope, while monitoring vowel features simultaneously, whereas English-learning infants are possibly attending only to changes in the general auditory dimensions of lexical tone, such as F0 height (and because this is not a tone-specific feature it is expected that it does not carry a high cognitive load).

2. Lower than optimal perception may be due to Chinese-learning infants’ discriminating nonnative lexical tone contrasts, that is, tones from Thai and not Chinese.

3. The results may be weakened by testing infants from Cantonese, Mandarin and bilingual Cantonese/Mandarin backgrounds on Thai tone discrimination. It is suggested for future research, that each language group be tested on their own tones and also on nonnative tone discrimination.

4. The task may not sufficiently tap real-world tone language processing and so Chinese-language learning infants’ discrimination of tone distinctions may be compromised.

The general complexity of tone perception calls into question whether the conditioned head-turn procedure is the most sensitive and appropriate method. The CHT does involve high learning, cognitive, and attentional demands and so it would be worthwhile employing an alternative discrimination procedure that requires less cognitive load, such as the habituation-dishabitation paradigm, to discover if the same perceptual abilities are uncovered.

\textsuperscript{75} Of course in a supralinguistic sense Australian English infants process F0.
10.5 Conclusions

This thesis provides the first comprehensive investigation of tone perception by infants, and the first evidence that lexical tone perception is initially language-general and then becomes more language-specific as a function of age. The finding of such a decline for non-tone but not tone language-learning infants, and for speech but not nonspeech stimuli emphasises that the reorganisation of tone perception is specifically for speech.

The experiments conducted here show that infants attend selectively and efficiently to tone contrasts in the speech stream if they are phonologically relevant in their native language, and extend our existing knowledge about infants’ attunement to native segmental and suprasegmental properties to incorporate segmental/suprasegmental lexical tone.

These experiments represent the beginning of research into lexical tone perception by infants, and raise many exciting research questions. There are two main challenges that now face infant researchers. The first is to complete the puzzle of cross-language lexical tone perception development, and the second is to determine how existing models of speech perception development may accommodate the forgotten speech component - lexical tone. If they cannot, then they must be changed.
REFERENCES


REFERENCES


REFERENCES


Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong* (pp. 29-47).


274
REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES

distributional cues. *Journal of Memory and Language, 35*, 606-621.

Koninklijke Nederlandse Akademie van Wetenschappen, 43*, 991-991.

tonal and non-tonal language speakers*. Paper presented at the 14th Australian
Language and Speech Conference, Brisbane, Australia.


Language and Society, 3*(2), 186-189.

Stager, C. L., & Werker, J. (1997). Infants listen for more phonetic detail in speech


Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and


type theory of speech perception: A reply to Lane's critical review. *Psychological
Review, 77*, 234-249.

REFERENCES


REFERENCES


REFERENCES


APPENDIX A1

Recruiting Infants in Australia and Hong Kong
A number of approaches were employed to promote the MARCS BabyLab within the community and to recruit infants for the studies in this thesis. This section outlines these approaches.

A1.1 Recruiting English-learning Infants in Sydney, Australia

MARCS Auditory Laboratories is located at the University of Western Sydney, Bankstown Campus, approximately 20 kilometres South-West of Sydney in a predominately working/middle-class region of Sydney. Families from suburbs in the Bankstown area and the neighbouring St George and Sutherland regions (to the east and south-east of Bankstown) were identified as the target demographic for recruiting infants. These regions of Sydney are proximal to the University campus, are matched for socio-economic status and education level (secondary to tertiary level education), there is a large number of families with young children and consequently many community Early Childhood Health Centres. The means employed to promote the studies to these families are set out in the subsections below.

Print Media

A full-colour pamphlet was designed for recruiting infants for all infant-related projects in MARCS BabyLab. There have been two pamphlet designs over the life of this project, and these are presented in Appendix CD-A1-1 on the CD supplement. The pamphlets were distributed to Early Childhood Health Centres, maternity units of local hospitals, public libraries and Community Centres in the Bankstown, St George and Sutherland regions of Sydney. The list of Centres and Hospitals that displayed BabyLab pamphlets are shown in Table A1.1. Pamphlets were also displayed on notice-boards at the University’s Bankstown campus. The pamphlet includes a brief description of projects in the MARCS BabyLab, introduces the researchers, the research partners and program sponsors, and clearly states the compensation participants receive for their participation. The pamphlet includes an Expression of Interest form for parents to complete and return to the BabyLab in a reply paid envelope, and telephone numbers for parents to for more information about current projects and to register their infant over the phone. Details from registration forms and phone calls were entered into a Baby Register database to enable the researcher to contact parents of infants approaching the appropriate age for participation in a study.
Table A1.1 Early Childhood Health Centres and Public Hospitals in Bankstown, St George, and Sutherland regions of Sydney that displayed BabyLab pamphlets.

<table>
<thead>
<tr>
<th>Region of Sydney</th>
<th>Early Childhood Health Centre (ECHC)</th>
<th>Public Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>City of Bankstown</td>
<td>Bankstown ECHC</td>
<td>Bankstown</td>
</tr>
<tr>
<td></td>
<td>Chester Hill ECHC</td>
<td>Liverpool</td>
</tr>
<tr>
<td></td>
<td>Condell Park ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Georges Hall ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenacre ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Karitane Mothercraft Society</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paddstown ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Panania ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Regents Park ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revesby ECHC</td>
<td></td>
</tr>
<tr>
<td>St George District</td>
<td>Brighton-Le-Sands ECHC</td>
<td>St George</td>
</tr>
<tr>
<td></td>
<td>Hurstville ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ramsgate ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Riverwood ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rockdale ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roselands ECHC</td>
<td></td>
</tr>
<tr>
<td>Sutherland Shire</td>
<td>Caringbah ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Engadine ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Illawong ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Menai ECHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miranda ECHC</td>
<td></td>
</tr>
</tbody>
</table>

Families who contacted the lab received a letter acknowledging registration together with an information package including the pamphlet (if they had not already received one), an information sheet and an example of the BabyLab newsletter (see Appendix
CD-A1-1 on the CD supplement for copies of the acknowledgement letter, information sheet, and newsletter).

Full-colour posters (see example in Appendix CD-A1-1, CD supplement) of various sizes were printed for display around the University campus and at shopping-centre recruitment drives.

Advertisements were placed in local newspapers, ‘The Bankstown Torch’ and ‘The Leader’ (distributed to St George and Sutherland regions) and ‘Sydney’s Child’, a free monthly parenting publication. An example advertisement is shown in Appendix CD-A1-1 on the CD supplement. These advertisements attracted expressions of interest from families all over Sydney and from the neighbouring City of Wollongong.

**Radio**
The BabyLab has also received radio exposure; an interview for local radio station 2NBC during 2001.

**Meeting Parents**
BabyLab Promotion Days were held at two local Shopping Centres (shopping malls), Bankstown Square (in April 2001) and Westfield Shoppingtown Miranda (in October 2001). See Figures A1.1 and A1.2 for photographs of the BabyLab stall). The main purpose of these promotion days was to meet with parents of infants and inform them of current research projects, allow parents to ask questions about participating in the studies, and to disseminate information packages and scientific papers published by BabyLab members. A video recording of infants being tested at MARCS was played throughout the day to attract passers-by to the stall, and so that parents could observe a test session in action. Interested parents were given the opportunity to register at the stall, although many more opted to take the reading materials home and discuss the program with other family members before signing up for the projects. The promotion days were highly successful in recruiting participants.
Figure A1.1. MARCS BabyLab promotion day at Bankstown Square Shopping Centre. From left to right: Karen Mattock, Stephen Malloch, and Christine Kitamura.

Figure A1.2. MARCS BabyLab promotion day at Westfield Shoppingtown Miranda. From left to right: Karen Mattock and Iris-Corinna Schwarz.

World Wide Web
During 2002, the BabyLab site was established and launched on the World Wide Web. The web address was subsequently added to all promotional and advertising material as an alternative means by which parents, sponsors and members of the local and academic communities could obtain more information about the research. The website
introduces the research team and their projects, contains a sponsors’ and research partners page, and also includes photographs of and descriptions about the testing methods used in the lab. A registration form can be downloaded and sent via email or fax to the lab and a phone number of the BabyLab is provided for those who wish to register over the phone. The web address as at 14th June 2004 is http://marcs.uws.edu.au/research/babylab.

Talks and Word-of-Mouth

Another strategy that has proved fruitful in recruiting participants has been to speak about the studies at Mothers’ Groups76 that are convened at the Early Childhood Centres. In addition many mothers’ who participated in this study recommended participation to their friends with infants.

A1.2 Recruiting Chinese learning Infants in Hong Kong

This project was one of the first studies to be conducted in the new Infant Perception Laboratory run by Dr. Valter Ciocca at the Division of Speech and Hearing Sciences77, University of Hong Kong, and established in 2002. Therefore recruiting infants was a new challenge in the Division. The candidate spent a total of .5.5 weeks at the University of Hong Kong split across two visits, one in January 2002 and the other in September-October 2002. The main purpose of these visits was to discuss collaboration on the project, set up equipment for the Conditioned Head-turn procedure, train the Hong Kong Research Assistant in the testing method, recruiting infants and identifying sponsors for the project. Recruiting infants in Hong Kong was an extremely difficult task. The strategies used at the BabyLab in Sydney proved to be insufficient or inappropriate in Hong Kong for the following reasons:

1. The costs of advertising are incredibly high, with classified advertisements costing the equivalent of AUD $2000 Australian dollars per day.

76 At Mothers’ Groups, mothers meet weekly with the nurse for Health checks, development assessment, and for advice on breastfeeding, health-care issues and establishing a sleeping routine for baby. The author gratefully acknowledges and thanks the following Early Childhood Centres for the regular invitation to speak to their Mothers’ Groups: Caringbah Early Childhood Centre, Padstow Early Childhood Centre, Panania Early Childhood Centre, and Riverwood Early Childhood Centre.

77 Prior to October 2002, the Division was known as the Department of Speech and Hearing Sciences.
2. The public culture surrounding young babies is less visible in Hong Kong than in Sydney. There are no Early Childhood Centres in Hong Kong and mothers are rarely seen walking infants in prams because of the traffic congestion, crowded streets and dangerous pollution levels. Therefore it was not easy to determine how parents of young babies could be contacted.

3. There was wariness about research from potential participants regarding involvement in a project that was perceived to have ‘minimal returns’.

For these reasons in order to recruit infants in Hong Kong successfully, more time and a fresh approach was required.

**Print Media**

A press release (see Appendix CD-A1-2, CD supplement, for an English version of this press release) was circulated and picked up by Hong Kong-based parenting magazine, ‘Ours’. ‘Ours’ featured an article about the research in February 2003 issue of their magazine\(^{78}\) (see Appendix see Appendix CD-A1-2, CD supplement for English translation of feature). The magazine editors believed that the best means of recruiting participants was to advertise the research as a *competition*, with winners receiving participation in the study, a free hearing test for infants\(^{79}\), sponsors’ products, and HK$80 in return for participation. In addition to agreeing to publish a feature article, the magazine editors agreed to send letters to magazine subscribers inviting them to participate in the research. The company also mailed letters to businesses that held advertising contracts with the magazine, in an effort to seek sponsors for the research. An example letter is presented in Appendix CD-A1-2 (on CD supplement).

Posters were designed to supplement the recruitment of infants. An example poster is presented in Appendix CD-A1-2

---

\(^{78}\) The assistance of Associate Professor Valter Ciocca and Research Assistant, Ms Betty Lee, in negotiating the feature article with ‘Ours’ is greatly appreciated.

\(^{79}\) The Division of Speech and Hearing Sciences, University of Hong Kong has audiometric testing facilities and expert staff for conducting hearing tests.
Letters to Doctors
Letters were written to the Births and Deaths General Register Office; paediatric, obstetrical and gynaecological societies/colleges of Hong Kong; the Department of Paediatrics and Adolescent Medicine and the Department of Obstetrics and Gynaecology at the University of Hong Kong; hospitals from Hong Kong Island and Kowloon; midwifery services, and breast-feeding associations (see Table A1.2 for a full list). These letters requested support for the research and assistance with pamphlet distribution. For an example letter, see Appendix CD-A1-2.
Table A1.2. *The Hospitals, Universities, Government Organisations and Associations to Whom Assistance with Infant Recruiting was Requested.*

<table>
<thead>
<tr>
<th>Hospitals</th>
<th>University/Colleges</th>
<th>Government</th>
<th>Associations and Societies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsun Yuk Hospital</td>
<td>Department of Obstetrics &amp; Gynaecology HKU</td>
<td>Births &amp; Deaths General Register Office</td>
<td>The Obstetrical &amp; Gynaecological Association of Hong Kong</td>
</tr>
<tr>
<td>Queen Mary Hospital Obstetrics &amp; Gynaecology Department</td>
<td>Hong Kong College of Obstetricians &amp; Gynaecologists</td>
<td></td>
<td>Hong Kong Paediatric Society</td>
</tr>
<tr>
<td>Hong Kong Adventist Hospital</td>
<td>Department of Paediatrics &amp; Adolescent Medicine HKU</td>
<td></td>
<td>Hong Kong Breastfeeding Mothers’ Association</td>
</tr>
<tr>
<td>Hong Kong Baptist Hospital</td>
<td></td>
<td></td>
<td>La Leche League of Hong Kong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pan Asian Lactation Consultants Association</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mothers of Multiples Organisation</td>
</tr>
</tbody>
</table>
The Problem of SARS

Unfortunately the majority of infants recruited in Hong Kong could not participate in the research because of the outbreak of Severe Acute Respiratory Syndrome (SARS). SARS prevented the candidate from returning to Hong Kong in May 2003 as planned to commence testing for the study. The University of Hong Kong was officially closed for a period of time during the outbreak and The World Heath Organisation declared travel warnings for South East Asia that included China and Hong Kong (Special Administrative Region). Following this announcement, the Department of Foreign Affairs and Trading, Australia issued similar travel warnings to Australian residents and the University of Western Sydney did not support (withdrew all financial support and travel insurance) travel by staff and students to China and Hong Kong. The problem was compounded by the fact that the Division of Speech and Hearing Sciences, University of Hong Kong (where the project was to be conducted) is located at the Prince Philip Dental Hospital, next door to one of the biggest public hospitals in Hong Kong, and where many outbreaks of SARS were diagnosed. Unfortunately by the time SARS was contained and travel warnings were lifted at the end of June 2003, the majority of infants recruited for the study via the feature article in “Ours” were too old to participate and a new recruitment campaign was required.

A1.3 Recruiting Chinese-learning Infants in Sydney, Australia

Without being able to predict when the outbreak of SARS would be contained in Hong Kong, it was decided to move this part of the research project to Sydney.\textsuperscript{80} Fortunately, Sydney is a multicultural city with a large Chinese community so a Chinese promotion campaign was possible. A successful media release (Appendix CD-A1-3 for a copy of the media release) generated interest from the press\textsuperscript{81} and stimulated interest within the Chinese community.

\textsuperscript{80} Note that once the outbreak was contained and the ban lifted, there were a small number of infants of age from Hong Kong who could be tested.

\textsuperscript{81} The candidate wishes to thank Ms Angela McIntyre, Senior Media Officer, University of Western Sydney for writing an excellent media release and for dealing with the media regarding this research project.
Print Media
A native Cantonese speaker was employed as a research assistant to translate and interpret for the candidate. This included translating the pamphlet (see Appendix CD-A1-3) and press release into Chinese.

In May 2003 the project was featured on the front page of the ‘Higher Education Supplement’, of the National newspaper ‘The Australian’. The story was also covered by a national Chinese newspaper publication, the ‘Australian Chinese Daily’ (May 2003). See Appendix CD-A1-3 to view these newspaper articles.

Radio
The candidate and a research assistant82 were interviewed on the SBS Radio, Cantonese program. The twelve-minute interview went to air during Saturday evening prime time. A transcript of the interview is provided on the CD-Rom accompanying this thesis (Appendix CD-A1-3).

Television
The Sydney-based branch of Hong Kong television program TVB-J covered the project on their cable television program during June 2003. The story featured interviews with the candidate and the candidate’s thesis supervisor and included a mock testing session with an infant from a Cantonese speaking family, so that viewers could see and hear the task in action. A transcript of the television feature is presented in Appendix CD-A1-3 (on the CD supplement).

Pamphlet Displays
The pamphlets were displayed at Early Childhood Health Centres situated in suburbs of Sydney with large Cantonese speaking populations, in addition to being displayed at public libraries and in community organisations servicing the Chinese community see Table A1.3 for a full list). The candidate and research assistant visited cooperating Centres to speak to mothers about the study. Participants in the study were asked to recommend the study to Cantonese speaking friends with infants and this ‘word of mouth’ means of promotion was very successful.

82Sincerest thanks to Ms Joan Liang who was the research assistant for this project.
### APPENDICES

**Table A1.3**

*The Community and Health Organisations to Whom Assistance with Recruiting Infants was Requested.*

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Early Childhood Health Centres</th>
<th>Libraries &amp; Community Centres</th>
<th>Other Contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashfield</td>
<td>Ashfield</td>
<td>2AC Chinese Radio</td>
<td></td>
</tr>
<tr>
<td>Auburn</td>
<td>Auburn</td>
<td>2CR China Radio</td>
<td></td>
</tr>
<tr>
<td>Burwood</td>
<td>Bankstown</td>
<td>Australian Hokien Huay Kuan Assoc.</td>
<td></td>
</tr>
<tr>
<td>Cabramatta</td>
<td>Bexley</td>
<td>Australian Chinese Buddhist Society</td>
<td></td>
</tr>
<tr>
<td>Campbelltown</td>
<td>Bexley North</td>
<td>Australian Chinese Charity Foundation</td>
<td></td>
</tr>
<tr>
<td>Campsie</td>
<td>Brighton-le-Sands</td>
<td>Australian Chinese Community Association</td>
<td></td>
</tr>
<tr>
<td>Carlingford</td>
<td>Cabramatta</td>
<td>Australian Chinese Friendship Assoc.</td>
<td></td>
</tr>
<tr>
<td>Chatswood</td>
<td>Campsie</td>
<td>Australian Chinese Real Estate's Assoc.</td>
<td></td>
</tr>
<tr>
<td>Chester Hill</td>
<td>Casula</td>
<td>Australian Chinese Teo Chew Assoc.</td>
<td></td>
</tr>
<tr>
<td>Epping</td>
<td>Chatswood</td>
<td>Burwood Mother's Keep Fit Group</td>
<td></td>
</tr>
<tr>
<td>Fairfield</td>
<td>Chester Hill</td>
<td>Campsie Cultural Centre</td>
<td></td>
</tr>
<tr>
<td>Fairfield Heights</td>
<td>Dulwich Hill</td>
<td>CASS</td>
<td></td>
</tr>
<tr>
<td>Georges Hall</td>
<td>Earlwood</td>
<td>Chin Hwa University Alumni</td>
<td></td>
</tr>
<tr>
<td>Granville</td>
<td>Epping</td>
<td>Chinese Chamber of Commerce</td>
<td></td>
</tr>
<tr>
<td>Hurstville</td>
<td>Fairfield</td>
<td>Chinese Cooking Fraternity Association</td>
<td></td>
</tr>
<tr>
<td>Ingleburn</td>
<td>Granville</td>
<td>Chinese Language Education Council</td>
<td></td>
</tr>
<tr>
<td>Kingsgrove</td>
<td>Green Valley</td>
<td>Chinese Youth League of Australia</td>
<td></td>
</tr>
<tr>
<td>Lidcombe</td>
<td>Greenacre</td>
<td>Chung Chin Association</td>
<td></td>
</tr>
<tr>
<td>Liverpool</td>
<td>Haymarket</td>
<td>Chungshan Society of Australia</td>
<td></td>
</tr>
<tr>
<td>Marrickville</td>
<td>Homebush</td>
<td>Hong Kong Tourist Association</td>
<td></td>
</tr>
<tr>
<td>Merrylands</td>
<td>Hurstville</td>
<td>Indochina Chinese Association</td>
<td></td>
</tr>
<tr>
<td>Newtown</td>
<td>Ingleburn</td>
<td>Jin Wu Koon</td>
<td></td>
</tr>
<tr>
<td>North Ryde</td>
<td>Kings Cross</td>
<td>N.S.W Hinh Linh Community Services</td>
<td></td>
</tr>
<tr>
<td>North Sydney</td>
<td>Kogarah</td>
<td>SBS Radio</td>
<td></td>
</tr>
<tr>
<td>Padstow</td>
<td>Lakemba</td>
<td>Sun Yat-Sen University Alumni Assoc.</td>
<td></td>
</tr>
<tr>
<td>Panania</td>
<td>Leichard</td>
<td>Sydney Chinese Community Centre</td>
<td></td>
</tr>
<tr>
<td>Parramatta</td>
<td>Lidcombe</td>
<td>TVB</td>
<td>Writer's Association</td>
</tr>
<tr>
<td>Petersham</td>
<td>Liverpool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punchbowl</td>
<td>Macquarie Fields</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Redfern</td>
<td>Marrickville</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regents Park</td>
<td>Merrylands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revesby</td>
<td>Miller</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riverwood</td>
<td>Moorebank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rockdale</td>
<td>Oatley</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ryde</td>
<td>Padstow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Hurstville</td>
<td>Parramatta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer Hill</td>
<td>Regents Park</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sutherland</td>
<td>Riverwood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yagoona</td>
<td>Rockdale</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sans Souci</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>South Hurstville</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>St Peters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strathfield</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Advertising

An advertisement (in English) was placed in ‘Sydney’s Child’, a parenting publication, calling for infants from Cantonese speaking families to participate in studies of tone language development. A classified advertisement was also placed in a number of Chinese newspapers distributed in Sydney.

The Inclusion of Mandarin-learning Infants

Many respondents to the advertising campaign were Chinese families who speak Mandarin rather than Cantonese and therefore the decision was made to also include infants from Mandarin-speaking families.

A1.4 Sponsorship

As set out in Section 4.2, the recruitment campaign targeted the suburbs surrounding the University of Western Sydney, Bankstown campus. Initially small incentives for participation such as a certificate and t-shirt (as offered by many other infant researchers) did not attract sufficient numbers of participants. To combat this problem, sponsors for the BabyLab were sought, and a system of gifts to parents instigated.\(^3\)

Letters were posted to companies identified as potential sponsors, such as baby product companies, book companies, baby food companies, photographic companies and toy companies (see Appendix CD-A1-4 for an example letter). The letters emphasised that this was a cause-related marketing opportunity that would generate company exposure to their target market, that is, mothers of infants and young children. In return for product donation, potential sponsors were informed that their name and logo would appear on the BabyLab pamphlet, posters, website, newsletters, and on conference presentations. Follow-up calls and ongoing correspondence with the companies resulted in substantial product donation for the project. These products were included in a ‘goodie bag’ for participants and proved vital to the recruitment campaign. The list of sponsors and the products they donated are detailed in Figure A1.3.

\(^3\) The assistance of MARCS BabyLab research assistant Kate Harle in liaising with potential sponsors is greatly appreciated.
The same approach was taken for the Hong Kong- based project. A copy of the letter to sponsors is presented in Appendix CD-A1-4 on CD supplement.

Figure A1.3. The BabyLab sponsors. These organisations donated baby products to supplement participant reimbursement.

**Gifts to Participants**

Participants in Sydney received AUD $20 travel reimbursement, a bag of sponsors’ baby products, an ‘I’ve been to uni, you can go too!’ t-shirt or feeding bib, and a Young Scientist Award. The t-shirt/bib design and the Young Scientist Award are presented in Appendix CD-A1-4 on CD supplement). Infants who returned for a second test session received a new collection of baby products, AUD$20, t-shirt or bib (i.e., whichever they did not receive on the first visit) and another Young Scientist Award.

Participants in Hong Kong received $80HK travel reimbursement, a bag of sponsors’ baby products, a toy koala and a t-shirt printed with ‘I’ve been to uni, you can go too!’.
APPENDIX A2

The Lexical Tones and the Making of the Nonspeech Tones
APPENDICES

The nonspeech stimuli (used for Experiments 2, 4, 6 and 8) were created from the Thai lexical tone stimuli (used for Experiments 1, 3, 5 and 7) by tracking the F0 of each Thai rising, falling, and low lexical tone exemplar, and mapping the F0 of each tone onto a keyboard generated violin sound (the template) so the F0 and violin sounds are combined to make a new sound. In doing this, phonetic information in the lexical tones is removed and only F0 and duration cues (i.e., length) remain. The violin sound was chosen as the template because musical glides can be produced on this instrument. Synthesising F0 onto an instrument where glides are possible was assumed to be a more realistic template for the smooth mapping of F0 changes in the speech sounds\(^84\). The A below middle C was the chosen note for the Violin sound and was recorded for 800ms, that is, longer than the duration of all the lexical tone tokens. The lexical tone were spoken by a female native speaker of Thai and so the note A below middle C, (i.e., 220Hz) approximating the fundamental frequency of the female voice (usually 200Hz plus, see section 1.2.1.5), was the chosen note for recording the violin template.

A2.1. Analysing the Lexical Tones for F0

The Speech Transformation & Representation using Adaptive Interpolation of weighted spectrogram (STRAIGHT) program developed by Kawahara et al. (1999)\(^85\) was used to analyse a source for F0, that is, to extract F0 from each lexical tone exemplar.

The STRAIGHT program is launched in Matlab version 6 (The Mathworks Inc, 1999) using a command line interface. A control panel for STRAIGHT then appears. The following steps must be performed on each lexical tone exemplar (in .wav format) separately:

1. The approximate F0 range of the speech file should be entered to assist the pitch tracker and to speed up the analysis. The program default F0 is 800Hz. This setting allows the program to search for F0 with values in this

\(^84\) Other instruments like the piano for example have discrete notes and no glides, and would not be the best choice for mapping onto F0 changes.

\(^85\) Thank you to Hideki Kawahara for permission to use STRAIGHT. Note that this program is not commercially available and so only a general description of how to extract F0 from an existing audio file is given here. These details are not sufficient for reproducing the F0 pitch plots of lexical tones.
range. For the present purpose, this was changed to 200Hz to be approximately equivalent to the F0 in Hz of the female voice.

2. A frame-length of one millisecond was specified, and the sampling rate was checked (STRAIGHT reads this direct from the audio file).

3. After analysing a source for F0 a window appears containing the candidate pitch plots of the .wav file and the one that STRAIGHT selects is highlighted. STRAIGHT will select the F0 candidate with the highest carrier-to-noise ratio at each time frame (in this case, every 1ms). If a F0 plot is not extracted it is recommended to widen the F0 search range.

Morphing the Pitch and Violin Template Without Changing F0

The Auditory Perception Toolbox (APT, Haszard Morris et al., 2002; see also website http://marcs.uws.edu.au/research/software) was used to map the extracted pitch to an audio violin sound from specifications in .mxs file form. APT is controlled using a command line interface in Matlab version 6 (The Mathworks Inc, 1999). The following steps were taken to make the nonspeech stimuli:

1. The extracted F0 variable from STRAIGHT was saved by APT in .mxs file format with file name and file duration specified.

2. The resultant .mxs file structure contains important information for mapping the pitch onto the violin template using APT, in particular the frequency and amplitude of the pitch plots at the start of each frame, the length of each analysis frame (i.e., 1 ms.), the start time of each analysis frame, and the overall amplitude of each analysis frame.

3. The ‘synthaudio’ command file was used for the synthesis with a text array specifying the file name and path name of the .wav file to be used as the template for synthesis (i.e., the violin sound), and the synthesised data was scaled to the amplitude contour of the pitch plot (specified in the .mxs file format from STRAIGHT).
A2.2 MXS File Data
The .mxs file structures created and subsequently used by APT, store fundamental frequency, frame duration, absolute duration, and amplitude values of the lexical tone exemplars. For example a data section from a bâ exemplar .mxs file contained the following information:

129.20  233.112  .8   .2   .001
130.2   233.339  .8   .2   .001

The first column represents the time snapshot at 129.20 and 130.2 ms respectively. The second is the F0 value at each time frame (i.e., 233.112Hz at 130ms into the file, and 233.339Hz at 131ms into the file). The third column specifies RMS amplitude-data, and the fourth column specifies frame length data in seconds (i.e., 1 ms.)

A2.3 Stimulus Details
The total soundfile duration (length), and duration of the consonant and periodic segments of each lexical and nonspeech tone is shown below for rising (Table A2.1), falling (Table A2.2), and low (Table A2.3) respectively. Audio files of the stimuli (.wav format) are presented in Appendix CD-A2 on the CD supplement. The total duration of each nonspeech tone is shorter than the lexical tone it was created from for two reasons: (a) F0 information is carried in the periodic (vowel) portion of a speech sound and only the periodic portion of the syllables contributed to the F0 tracking - importantly the periodic portion of each nonspeech analogue and its original lexical tone are identical in length -and (b) some (but not all) speech soundfiles contain prevoicing or creaky voice (a phonetic feature of Thai, see section 1.3.3.2) and the amplitude of this prevoicing was too low for accurate pitch tracking. Therefore the difference in duration of the speech and its nonspeech tone analogue is due to difficulties tracking F0 of the consonant and prevoicing segments. F0 plots of a falling, low, and rising tone exemplar are shown in Figures A2.1, A2.2, and A2.3 respectively.
Table A2.1.

_Periodic, consonant, and total soundfile duration of the rising lexical and nonspeech tone stimuli. Difference column represents discrepancy in duration of lexical and nonspeech tokens following mapping of F0 to violin sound._

<table>
<thead>
<tr>
<th>Rising Tone</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Periodic</td>
</tr>
<tr>
<td>bā1 rising1</td>
<td>559</td>
</tr>
<tr>
<td>bā2 rising2</td>
<td>562</td>
</tr>
<tr>
<td>bā3 rising3</td>
<td>628</td>
</tr>
<tr>
<td>bā4 rising4</td>
<td>558</td>
</tr>
<tr>
<td>bā5 rising5</td>
<td>589</td>
</tr>
</tbody>
</table>

314
Table A2.3

Periodic, consonant, and total soundfile duration of the falling lexical and nonspeech tone stimuli. Difference column represents discrepancy in duration of lexical and nonspeech tokens following mapping of F0 to violin sound.

<table>
<thead>
<tr>
<th>Falling Tone</th>
<th>Periodic</th>
<th>Consonant</th>
<th>Total</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>bâ1 falling1</td>
<td>562</td>
<td>89</td>
<td>651</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>88</td>
<td></td>
<td>650</td>
<td></td>
</tr>
<tr>
<td>bâ2 falling2</td>
<td>572</td>
<td>88</td>
<td>660</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td></td>
<td>614</td>
<td></td>
</tr>
<tr>
<td>bâ3 falling3</td>
<td>612</td>
<td>139</td>
<td>751</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td></td>
<td>657</td>
<td></td>
</tr>
<tr>
<td>bâ4 falling4</td>
<td>573</td>
<td>133</td>
<td>706</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td></td>
<td>629</td>
<td></td>
</tr>
<tr>
<td>bâ5 falling5</td>
<td>575</td>
<td>72</td>
<td>647</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td></td>
<td>608</td>
<td></td>
</tr>
</tbody>
</table>
Table A2.3

*Periodic, consonant, and total soundfile duration of the low lexical and nonspeech tone stimuli. Difference column represents discrepancy in duration of lexical and nonspeech tokens following resynthesis.*

<table>
<thead>
<tr>
<th>Low Tone</th>
<th>Periodic</th>
<th>Consonant</th>
<th>Total</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>bà1</td>
<td>549</td>
<td>59</td>
<td>608</td>
<td>8</td>
</tr>
<tr>
<td>low1</td>
<td></td>
<td>51</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>bà2</td>
<td>530</td>
<td>85</td>
<td>615</td>
<td>1</td>
</tr>
<tr>
<td>low2</td>
<td></td>
<td>84</td>
<td>614</td>
<td></td>
</tr>
<tr>
<td>bà3</td>
<td>578</td>
<td>97</td>
<td>675</td>
<td>26</td>
</tr>
<tr>
<td>low3</td>
<td></td>
<td>71</td>
<td>649</td>
<td></td>
</tr>
<tr>
<td>bà4</td>
<td>592</td>
<td>103</td>
<td>695</td>
<td>3</td>
</tr>
<tr>
<td>low4</td>
<td></td>
<td>100</td>
<td>692</td>
<td></td>
</tr>
<tr>
<td>bà5</td>
<td>571</td>
<td>81</td>
<td>652</td>
<td>1</td>
</tr>
<tr>
<td>low5</td>
<td></td>
<td>80</td>
<td>651</td>
<td></td>
</tr>
</tbody>
</table>
Figure A1.1. F0 plot of a Thai falling tone on the syllable [ba]. Duration of exemplar was .651ms. Duration of the other exmaplars was 660ms, 751ms, 706ms, and 647ms.

Figure A1.2. F0 plot of a Thai low tone on the syllable [ba]. Duration of exemplar was .616ms. Duration of the other exemplars was 608ms, 675ms, 695ms, and 652ms.

Figure A2.3. F0 plot of a Thai rising tone on the syllable [ba]. Duration of this exemplar was 702ms. Duration of the other exemplars was 621ms, 630ms, 664ms, and 741ms.