CHILDREN'S ACQUISITION OF MANDARIN TONES IN CONTEXT

PING TANG

BA (Chinese Language & Literature), MA (Linguistics & Applied Linguistics)

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics

ARC Centre of Excellence in Cognition and its Disorders

Faculty of Human Sciences

Macquarie University

Sydney, Australia

TABLE OF CONTENTS

TABLE OF CONTENTS II
THESIS ABSTRACT VII
DECLARATION IX
ACKNOWLEDGEMENTS XI
CHAPTER ONE: THESIS INTRODUCTION 1
INTRODUCTION 2
Mandarin tone system 3
Lexical tones 3
Neutral tone 3
Tone sandhi 4
Acquisition of Mandarin tones 5
Typically developing children 5
Children with hearing impairment 6
Potential reasons for the late acquisition of contextual tones 7
Early input 7
Variable surface realization of contextual tones 8
Acquiring contextual tones in speech perception 8
Organization of the present thesis 9
REFERENCES 13
CHAPTER TWO: PHONETIC ENHANCEMENT OF MANDARIN VOWELS AND TONES: INFANT-DIRECTED SPEECH AND LOMBARD SPEECH 16
ABSTRACT 17
INTRODUCTION 18
METHOD 23
Participants 23
Stimuli 23
Procedure 26
Coding and Measurements 28
Vowels 28
Tones 29
Statistical analysis 31
RESULTS 33
Vowels 33
Vowel space area 33
Formant frequencies 35
Tones 40
Tone space area 40
Pitch height and tonal duration 43
DISCUSSION 49
CONCLUSION 55
ACKNOWLEDGEMENTS 55
REFERENCES 56
CHAPTER THREE: ACOUSTIC REALIZATION OF MANDARIN NEUTRAL TONE AND TONE SANDHI IN INFANT-DIRECTED SPEECH AND LOMBARD SPEECH 63
ABSTRACT 64
INTRODUCTION 65
Neutral tone 68
Tone sandhi 69

METHOD 73
Participants 73
Stimuli 74
Procedure 76
Coding and measurements 78
Neutral tone *78*
Tone sandhi *82*
Statistical analysis 82
RESULTS 83
Neutral tone 83
Tone sandhi 87
DISCUSSION 91
CONCLUSION 96
ACKNOWLEDGEMENTS 97
REFERENCES 98
APPENDIX 104
CHAPTER FOUR: ACQUISITION OF WEAK SYLLABLES IN TONAL LANGUAGES: ACOUSTIC EVIDENCE FROM NEUTRAL TONE IN MANDARIN CHINESE 113
ABSTRACT 114
INTRODUCTION 115
METHOD 122
Participants 122
Stimuli 123
Procedure 125
Coding and measurements 127
Statistical analysis 129
RESULTS 131
Pitch 131
Duration 138
DISCUSSION 144
CONCLUSION 148
REFERENCES 150
APPENDIX 155
CHAPTER FIVE: THE ACQUISITION OF PHONOLOGICAL ALTERNATIONS: THE CASE OF THE MANDARIN TONE SANDHI PROCESS 162
ABSTRACT 163
INTRODUCTION 164
Lexical tones 165
Tone sandhi application as a function of tonal context in disyllabic words 166
Tone sandhi application as a function of prosodic structure in trisyllabic words 167
Previous studies of tone sandhi acquisition 169
METHOD 173
Participants 173
Stimuli 174
Procedure 180
Coding and measurements 185
Statistical analysis 186
RESULTS 188

Tone sandhi in disyllabic words 188
Lexical tones *188*
Disyllabic tone sandhi compounds *190*
Tone sandhi in trisyllabic words 192
Control items: T3T3T1/2/4 compounds *193*
Target items: T3T3T3 compounds *194*
DISCUSSION 200
Adults: speaking rate 202
Frequency distribution of left-branching vs. right-branching compounds in Mandarin 204
Weak lexical representation of novel trisyllabic compounds 204
Children's T2T2T3 productions for the right-branching structure 206
CONCLUSION 208
REFERENCES 209
APPENDIX 215
CHAPTER SIX: THE ACQUISITION OF MANDARIN TONAL PROCESSES BY CHILDREN WITH COCHLEAR IMPLANTS 218
ABSTRACT 219
INTRODUCTION 221
GENERAL METHODS 227
Participants 228
Coding and measurements 230
Statistical analysis 231
EXPERIMENT 1: LEXICAL TONES 232
Stimuli 232
Procedure 232
Results 233
Group comparison *233*
Age at implantation 233
CI experience 235
Regression analysis *237*
Summary 237
EXPERIMENT 2: NEUTRAL TONE 237
Stimuli 238
Procedure 239
Results 240
Group comparison *240*
Age at implantation 240
CI experience 243
Regression analysis *245*
Summary 245
EXPERIMENT 3A: TONE SANDHI IN DISYLLABIC COMPOUNDS 246
Stimuli 246
Procedure 248
Results 249
Group comparison *249*
Age at implantation 249
CI experience 251
Regression analysis *253*
Summary 254
EXPERIMENT 3B: TONE SANDHI IN TRISYLLABIC COMPOUNDS 254
Stimuli 255

Procedure 255
Results 257
Group comparison *257*
Age at implantation 257
CI experience 259
Regression analysis *261*
Summary 261
GENERAL DISCUSSION 262
CONCLUSION 268
ACKNOWLEDGEMENTS 268
REFERENCES 270
SUPPLEMENTARY MATERIALS 277
CHAPTER SEVEN: THE REPRESENTATION OF ALLOPHONIC VARIANTS OF TONE SANDHI IN THE DEVELOPING LEXICON 367
ABSTRACT 368
INTRODUCTION 369
METHOD 372
Participants 372
Stimuli 373
Audio stimuli *374*
Visual stimuli *375*
Apparatus 376
Procedure 376
Data analysis 378
RESULTS 379
Difference ratio 379
Time course analysis 381
DISCUSSION 383
ACKNOWLEDGEMENTS 386
REFERENCES 387
APPENDIX 391
CHAPTER EIGHT: CHILDREN WITH COCHLEAR IMPLANTS' REPRESENTATION OF ALLOPHONIC VARIANTS OF TONE SANDHI 396
ABSTRACT 397
INTRODUCTION 398
METHODS 402
Participants 402
Stimuli 403
Audio stimuli *404*
Visual stimuli *405*
Apparatus 405
Procedure 406
Data analysis 407
RESULTS 407
Difference ratio 407
Time course analysis 410
DISCUSSION 412
REFERENCES 416
APPENDIX 419
CHAPTER NINE: THESIS DISCUSSION 423
DISCUSSION 424

STUDY LIMITATIONS AND FUTURE DIRECTIONS...430
REFERENCES 434

THESIS ABSTRACT

Mandarin is a tonal language with four lexical tones and two contextual tones, i.e., neutral tone and tone sandhi, exhibiting tonal variations across contexts. Previous studies found that Mandarin-learning children acquire lexical tones early (before 3 yrs.), but the neutral tone and tone sandhi are later acquired (after 4;6). Children with hearing impairment/cochlear implants (CIs) have problems in acquiring lexical tones unless implanted early or have long CI experience, since CIs do not transmit pitch information effectively; it is unclear how those implanted early perform on contextual tones. The general aim of this thesis was therefore to better understand the acquisition of contextual tones by both typically developing children and those with hearing impairment using acoustic, rather than perceptual measures.

This thesis consisted of seven studies. Firstly, we examined how contextual tones are realized in children's language input, i.e., infant-directed speech (IDS) and clear speech, where slower, hyperarticulated speech (as directed to children or possibly hearing impaired populations) might destroy the context for appropriate realization of contextual tones, thus potentially explaining the later acquisition of these tonal processes (Studies 1 & 2). However, our findings showed that the key features of contextual tones are well realized in both registers, suggesting that later acquisition is not due to the input. Next, we examined children's contextual tone productions in novel (rather than known) items, exploring when their knowledge of contextual tones becomes productive (Studies 3, 4 & 5). The results showed that (1) normal hearing 3-year-olds have already acquired productive knowledge of contextual tones, correctly producing tonal variations across contexts, though adult-like acoustic implementation is not fully mastered until age 5; (2) children with CIs face challenges

in producing correct contextual tones, but early implantation (before age 2) facilitates lexical and contextual tonal acquisition. Finally, we tested children's perception of lexical tones and tone sandhi in novel compounds using a mispronunciation/eye-tracking task (Studies 6 & 7). The results showed that, (1) all normal hearing participants could detect lexical tone mispronunciations, but none were sensitive to the novel compound tone sandhi mispronunciations; and (2) even detecting lexical tone mispronunciations was a challenge for the children with CIs.

Taken together, the findings of this thesis suggest that contextual tones are not later acquired by normal hearing children, and acquiring typical contextual tones is possible for children with hearing impairment as long as they receive CIs early.

DECLARATION

The research presented in this thesis is my original work and has not been submitted for a higher degree in any other institution. In addition, I certify that all information sources and literature used are indicated in the thesis. The research presented in this thesis has gained ethics approval from Macquarie University (5201600080, 5201600920 and 5201826823055).

Chapters **2, 3** and **4** in this thesis have been published as:

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(2), 493-503.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(5), 2823-2835.

Tang, P., Yuen I., Xu Rattanasone, N., Gao L., Demuth, K. (2018). Acquisition of weak syllables in tonal languages: acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, 1–27.

Chapters **5** and **6** in this thesis have been revised based on reviewers' comments and resubmitted as:

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., Demuth, K. (revision submitted). The acquisition of phonological alternations: the case of the Mandarin tone sandhi process. *Applied Psycholinguistics*.

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., Demuth, K. (revision submitted). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*.

I additionally certify that I was the first author of all the chapters of this thesis, and that all the experimental design, data collection, coding, analysis and writing of the articles have been performed and completed by me, in consultation with my principle supervisor Katherine Demuth, co-supervisor Nan Xu Rattanasone and advisor Ivan Yuen.

Signed: Ping Tang (Student Number: 44742509).

ACKNOWLEDGEMENTS

Time flies and now it has been three years since I started my Ph.D. The Ph.D. journey is short but I have learnt a lot things and I really enjoy it. Now I would like to give my grateful acknowledgement to everyone who had provided me with valuable help and encouragement during the past three years.

I would like to express my grateful thanks to my principal supervisor Distinguished Professor Katherine Demuth, who has provided me with tremendous help and encouragement during my entire Ph.D. journey. She is always available and provides fast responses to all of my queries and papers, greatly facilitating my thesis writing. She is an "artist" of academic writing, always reminding me to step back, zoom out and think about the big picture of my research, which I believe is critical for every researcher and substantially broadens my horizons of research. I would also thank my co-supervisor Nan Xu Rattanasone and my co-advisor Ivan Yuen, who spend numerous amounts of time and energy to help me in experiment designing, data analysis and paper writing. Without them, my Ph.D. journey would never be such smooth. I would also thank all member of the Child Language Lab for your assistance and accompany during the past years. Thank you to post-docs and researchers Carmen Kung, Laurence Bruggeman, Titia Benders and Elaine Schmidt. Thank you to Lab coordinators Katherine Revius, Nyari Marunda and Isabel O'Keeffe. Thank you to my colleagues and friends Julien Millasseau, Rebecca Holt, Thembi Dube, Ben Davies, Phillip Chen, Tunde Szalay, Louise Ratko, Jason Hollowel, Kyoji Iwamoto, Sujal Pokharel, Qandeel Hussain, Hui Chen, Tamara Schembri, Xuejing Lv and Valarie Pu. Thank you to researchers from the phonetic group Felicity Cox, Michael Proctor and

Anita Szakay. Thank you to Collette Ryan for providing tremendous help with my data collection in China.

I would like to thank the iMQRES for financial support, the Faculty of Human Sciences and the Hearing CRC for funding my research and the ARC Centre of Excellence for Cognition and its Disorders for providing advanced facilitates and such a wonderful work place.

I would like to thank all participants who have participated in my studies in Sydney and Beijing. I would like to thank my collaborators in Beijing to make these studies possible. Thank you Professor Liqun Gao for liaising with preschools and institutions. Thank you Xiaoling Rong for assistance in data collection. Thank you teachers in the affiliated Kindergarten of Beijing Language and Culture University, Piaoliangmama Rehabilitation centre, Xiaojudeng Rehabilitation centre and China Rehabilitation Research Centre for Hearing and Speech Impairment.

I would like to express my sincere thanks to my friends in Sydney who support and accompany me during my entire Ph.D. journey, making my life wonderful here. Thank you my best friends Lun, Care, Biubiu, Penny, Alex, Jim, Shan, Wei, ZZ and Ke Ma. I would also thank my best friends in Nanjing who keep staying in touch with me, providing me with remote supports, Lei Liu, Shanpeng Li and Haitao Guan.

Finally, I would express my grateful thanks to my beloved parents and my grandma for providing me with financial and spirit supports. Without your support, I would not conquer many difficulties that easily.

Chapter One: Thesis Introduction

Introduction

During early language acquisition, children have to develop a clear understanding of the speech sounds that can contrast word meanings in their native language and learn this sound system. In Germanic and Romance languages, word meanings are contrasted by segments such as consonants and vowels. However, most languages on the world are tonal languages, using tones in addition of segments to convey word meanings (Yip, 2002). Thus, children learning tonal languages have to develop a clear understanding of the tonal system of their native language. Mandarin Chinese is a tonal language spoken by a large population in mainland China. The Mandarin tonal system has four lexical tones. It has been suggested that, due to the importance in conveying word meanings, these four lexical tones are early acquired, i.e., before age 3 (Li & Thompson, 1977; Hua and Dodd, 2000; Hua, 2002). Apart from lexical tones, the Mandarin tonal system also has two types of contextual tones, i.e., neutral tone and tone sandhi, exhibiting contextually conditioned tonal realizations in connected speech. Relative to lexical tones, less attention has been focused on the acquisition of these contextual tones, especially for children with hearing impairment/cochlear implants (CIs). Among the few studies, it has been suggested that contextual tones are later acquired by typically-developing children, i.e., after 4;6, while the reason remains unclear (Hua & Dodd, 2000; Wang, 2011; Xu Rattanasone, et al., 2018). Moreover, it has been suggested that children with CIs show difficulties in acquiring lexical tones since the CI devices do not transmit pitch information effectively, but there has been no study of how they perform on contextual tones. The aim of the present thesis was therefore to better understand the acquisition of contextual tones by both typically developing children and children with CIs.

Mandarin tone system

Lexical tones

Mandarin is a tone language with four lexical tones, primarily contrasting in pitch contour, i.e., level Tone 1 (T1), rising Tone 2 (T2), dipping Tone 3 (T3) and falling Tone 4 (T4; also see Figure 1). Note that the four tones can also be described based on their phonological tonal features, such as High (H) for T1, Low-High (LH) for T2, Low (L) for T3 and High-Low (HL) for T4 (Yip, 2002). The four tones can contrast word meanings, i.e., "ma1" *mother*, "ma2" *hemp*, "ma3" *horse* and "ma4" *scold*.

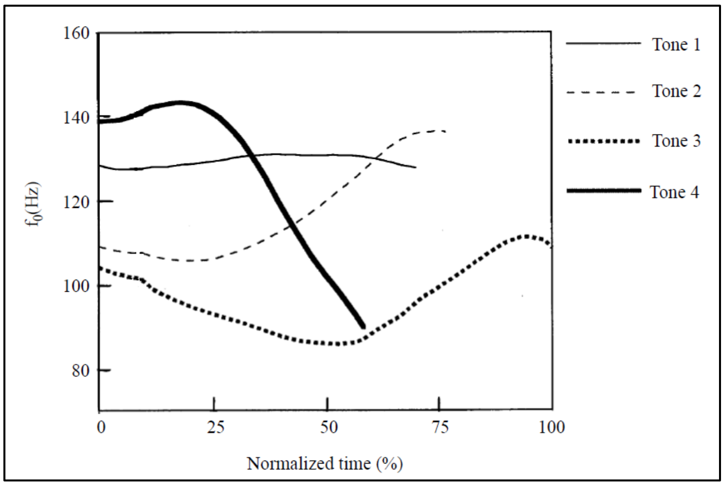


Figure 1. Pitch contour of Mandarin lexical tones, sourced from Xu (1997).

Neutral tone

Neutral tone is a special tone category, also referred as "T0", carried by weak syllables, occurring in connected speech (Yip, 2002). It does not have a fixed tonal pattern and its surface tonal realization is variable depending on the preceding lexical tones. Its pitch contour, for instance, is falling when following T1/2/4 syllables and rising/level when following T3 syllables (Cao, 1992). The duration of neutral tone syllables is generally shorter than that of full tone syllables, but varies following different lexical tones, i.e., shorter when following T1/2/4 (around 50% - 60% long as a full tone syllable) compared to T3 (around 70% as long as a full tone syllable; Tang, 2014).

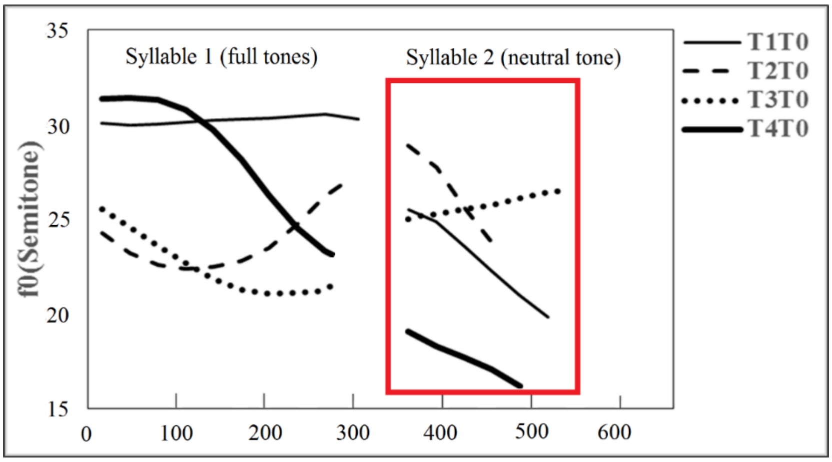


Figure 2. Pitch realization of neutral tone syllables (red box) across tonal contexts, from Tang (2014).

Tone sandhi

Tone sandhi is a tonal process occurring in connected speech, changing an underlying T3 from a dipping tone to **T2** (rising tone) or **half-T3** (low-dipping tone) in response to the following tonal context. When T3 is spoken before another T3 syllable, for instance, it becomes a rising T2: **T3** + T3 → **T2** + T3; when T3 is spoken before T1/2/4, it becomes a half-T3 with a low-dipping pitch: **T3** + T1/2/4 → **half-T3** + T1/2/4 (Chen, 2000). For multiple-T3 sequence such as trisyllabic T3T3T3 items, the application of the tone sandhi is more complex, and depends on the internal prosodic structure of the sequence, i.e., left-branching ((T3T3) T3) or right-branching (T3 (T3T3)) structures, resulting different surface outputs: left-branching ((T3T3) T3) → **T2T2**T3 and right-branching (T3 (T3T3)) → T3**T2**T3/**T2T2**T3 (Chen, 2000).

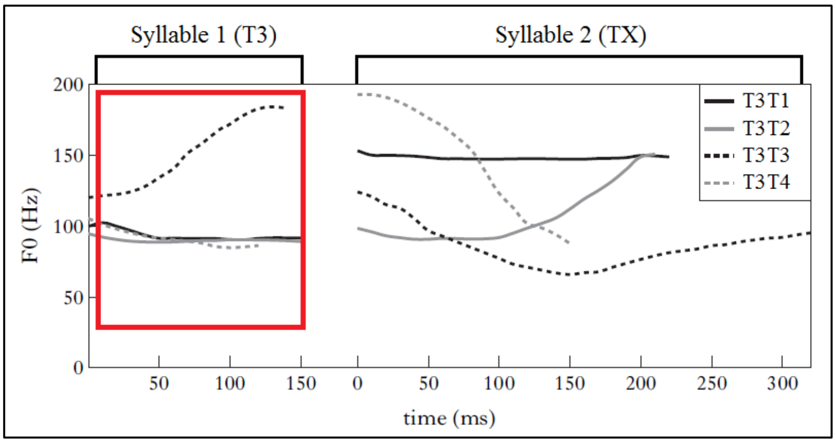


Figure 3. Pitch realization of tone sandhi syllables (red box) across different tonal contexts, sourced from Zhang and Lai (2010).

Acquisition of Mandarin tones

Typically developing children

Previous studies generally agree that lexical tones appear early in acquisition, i.e., during the single-word stage of development, although confusion between T2 (the rising tone) and T3 (the dipping tone) continues into the 2/3-word stage of development (Li & Thompson, 1977). By the age of three, when children begin to combine words into longer sentences, all lexical tones are reportedly acquired (Li & Thompson, 1977). However, despite extensive research on lexical tone acquisition, relatively few studies have examined how and when contextual tones are acquired (Li & Thompson, 1977; Hua & Dodd, 2000; Wang, 2011; Xu Rattanasone et al., 2018). Among these studies, it is generally suggested that, relative to lexical tones, the acquisition of contextual tones is a protracted process. Hua and Dodd (2000), for instance, examined neutral tone productions from Mandarin-speaking children (1;6 to 4;6), finding that no 3-year-olds could correctly produce all neutral tone words, and only 36% 4-year-olds (4;6) could do so. Xu Rattanasone et al. (2018) and Wang (2011) examined children's tone sandhi productions in disyllabic and polysyllabic words, finding that even 5-year-olds' tone sandhi productions were not fully adult-like. However, most of these studies only conducted subjective perceptual coding, and it is unclear about the acoustic realization of children's productions of contextual tones across contexts. These raise many questions about why contextual tones are later acquired.

Children with hearing impairment

For children with CIs, acquiring typical lexical tones is challenging, since the CI devices do not transmit pitch information effectively, as a consequence of a restricted number of channels used to deliver a wide range of speech frequencies (Vandali & van Hoesel, 2012). Previous studies suggest that Mandarin-speaking children with CIs generally face challenges in acquiring lexical tones, with poorer performance in both tonal perception and production as compared to their hearing peers (Tan, Dowell & Vogel, 2016; Chen & Wong, 2017). However, it has also been suggested that early implantation and long CI experience benefit children with CIs in lexical tone acquisition, usually predicting better outcomes in both tone production and perception (Tan, Dowell & Vogel, 2016; Chen & Wong, 2017). However, since there has been no study examining the acquisition of contextual tones by children with CIs, it was unclear whether and when they acquire contextual tones, and whether early implantation and long CI experience would also benefit the acquisition of contextual tones. Given that people communicate with each other using connected speech rather than single words, it is necessary to extend previous studies on tonal acquisition by children with CIs from lexical tones in single words to contextual tones in connected speech, exploring the effect of early implantation and long CI experience in the acquisition of contextual tones.

Potential reasons for the late acquisition of contextual tones

Early input

One potential reason leading to the later acquisition of contextual tones might be related to the language input, i.e., infant-directed speech (IDS). On the one hand, IDS provides an important source of information from which children learn language (Kuhl et al., 1997; Burnham et al., 2002). On the other hand, it is also characterised by unique acoustic characteristics such as exaggerated pitch contours and slow speaking rate (Song, Demuth & Morgan, 2010). It has been suggested that, for Mandarin Chinese, the pitch contours of lexical tones are hyperarticulated in IDS, thus potentially facilitating children's acquisition of lexical tones (Liu, Tsao & Kuhl, 2007). However, given the unique acoustic features of IDS, especially the slow speaking rate, it was unclear whether the contexts triggering the correct realization of neutral tone and tone sandhi still exist in IDS. If not, children might not receive enough good exemplars of contextual tones and this could result in their later acquisition. To further explore the potential reasons leading to the later acquisition of contextual tones (Li & Thompson, 1977; Hua & Dodd, 2000), it was therefore necessary to examine the acoustic realization of contextual tones in IDS.

Variable surface realization of contextual tones

Another possible reason leading to the later acquisition of contextual tones might be related to the variable surface realization of neutral tone and tone sandhi across tonal contexts. This might pose potential challenges for children to abstract the phonological representation of contextual tones. However, as mentioned before, in analysing children's productions, previous studies primarily used human subjective evaluation rather than acoustic analysis, it was therefore unclear whether and when children are able to abstract the phonological representations of contextual tones and correctly implement them across tonal contexts. More importantly, previous studies primarily used familiar/lexicalized words as stimuli to elicit children's tonal productions, such as the neutral tone word "bi2 zi0" *nose* and tone sandhi word "shou3 zhi3" *finger* (Hua & Dodd, 2000). It is therefore unclear when children's knowledge of neutral tone (category) and tone sandhi (processes) become *productive*, i.e., correctly implementing tonal variation in *novel items* they have never heard before. This requires studies using both **novel** items to test children's productive knowledge of contextual tones and **acoustic** measurement to examine the fine-grained tonal realization of children's contextual tone productions across contexts.

Acquiring contextual tones in speech perception

So far most studies exploring the acquisition of contextual tones primarily conducted **production** experiments, analysing children's tonal outputs in speech production. Despite the fact that speech **perception** is an important component of daily speech communication, few studies have tested children's knowledge of contextual tones in speech perception. To our knowledge, there is only a study examining Mandarin-speaking 3-5-year-olds' perception of *tone sandhi* words, using familiar/lexicalized words as stimuli, such as "shou3 zhi3" *finger* (Wewalaarachchi & Singh, 2016). However, since familiar words are only realized with surface sandhi tones (i.e., T2T3) in children's language input, it is unclear whether children really understand the tone sandhi process, or simply remember the surface realization of these words in their language input. To better understand children's productive knowledge of the tone sandhi *process*, it is thus necessary to extend the study of Wewalaarachchi and Singh (2016) from lexicalized words to novel items that children have never heard before. Moreover, it is also necessary to test tone sandhi perception by children with CIs, to better examine the effect of early implantation on their acquisition of tone sandhi process in both production and perception processes.

Organization of the present thesis

Chapters 2 and 3 examine the acoustic realization of Mandarin tones in IDS, including lexical tones, neutral tone and tone sandhi, to explore the potential reason for the later acquisition of contextual tones from the perspective of early input. Tonal realizations were acoustically compared across three registers: adult-directed speech (ADS; serving as the baseline), IDS and Lombard speech (a speech style typically used in noisy environment). Lombard speech is another hyperarticulated speech register sharing similar acoustic features with IDS, belonging to a type of "clear speech" (Uchanski, 2005). Examining how tones are realized in Lombard speech could potentially inform the language input children with hearing impairment might receive. Given the acoustic features of IDS and Lombard speech, especially the slower speaking rate, it was hypothesized that the context for appropriate realization of contextual tones might be destroyed, resulting in atypical realizations of contextual tones in both registers. However, the results showed that, despite the fact that both registers exhibit hyperarticulation, the key features of the contextual tones were still preserved, suggesting that children's later acquisition of contextual tones is unlikely to be driven by insufficient input.

Chapters 4, 5 and 6 examined the contextual tone productions of typically developing children and children with CIs. To be specific, chapter 4 examined 3-5-year-old typically developing children's **neutral tone** productions in familiar and novel items across tonal contexts, acoustically comparing these with those from adult controls. Based on previous findings that neutral tone is later acquired, i.e., after 4;6 (Hua & Dodd, 2000), it was hypothesized that only 5-year-olds would be able to correctly produce the tonal variation of neutral tone syllables across tonal contexts. However, the results showed that children at age 3 can already correctly produce the tonal variation of neutral tones in both familiar and novel items, suggesting early arrived knowledge of the neutral tone category, though the full mastery of adult-like acoustic implementations of neutral tone is only acquired by age 5.

Chapter 5 then examined 3-5-year-old typically developing children's **tone sandhi** productions in novel disyllabic and trisyllabic compounds. We predicted that the full mastery of the tone sandhi process with novel items would be a protracted process. The results showed that children age 3 are already able to productively apply the tone sandhi process to novel disyllabic compounds in an adult-like way, suggesting the early arrived productive knowledge of tone sandhi, though mastering the application of the tone sandhi across different *trisyllabic* prosodic contexts appears to be a protracted process not yet completed by the age of 5.

Chapter 6 examined the tonal productions from **children with CIs**, including lexical tones and contextual tones. Based on previous studies (Tan, Dowell & Vogel, 2016; Chen & Wong, 2017), we predicted that both early implantation and long CI experience would facilitate children with CIs' acquisition of lexical tones and contextual tones. The results partly confirmed our prediction, showing that only early implanted (before age 2) children produced typical lexical tones and generally had contextual tones approximating those of the NH children, while other children, including those with a longer CI experience, still could not do so. This result reveals the importance of early implantation in the tonal development of children with CIs, including both lexical tones and contextual tones.

Chapters 7 and 8 examined the phonological representation of tone sandhi by both typically developing children (3-5-year-olds) and children with CIs. In Chapter 7, we examined children's perception of two surface allophonic variants of tone sandhi (rising and low-falling) in novel full sandhi compounds, using the tonal mispronunciation task similar to that used in Wewalaarachchi & Singh (2016). Based on Wewalaarachchi & Singh (2016), we predicted that children would reject the tonal mispronunciation between the two variants in this context. However, the results showed that both 3-5-year-olds and adults accepted both variants as possible surface realization for tone sandhi, indicating that children and adults represented both (allophonic) variants for tone sandhi realization in their mental lexicon. In Chapter 8, we conducted the same experiment on children with CIs. The results showed that children with CIs have difficulties recognizing novel compounds and detecting tonal changes even in the control condition, suggesting that novel-word learning and tonal perception are challenging for these children.

Finally, Chapter 9 summarizes the general findings of the entire thesis, arguing that the acquisition of contextual tones might not be as protracted process as suggested by previous studies. The chapter also highlights the implications and significance of the findings of the present thesis, outlining the limitations of the studies and suggesting future directions for further research.

References

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. *Science*, 296, 1435.

Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics*, 20(3), 49-50.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press.

Chen, Y., & Wong, L. L. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology*, 56(sup2), S7-S16.

Hua, Z. (2002). *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children*. Clevedon: Multilingual Matters Ltd.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3-42.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (**1997**). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684-686.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43, 912-917.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, 128, 389-400.

Tan, J., Dowell, R., & Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American Journal of Audiology*, 25(3), 246-256.

Tang, P. (2014). A study of prosodic errors of Chinese neutral tone by advanced Japanese students (Chinese version). *TCSOL Studies*, 56(4), 39-47. https://doi.org/10.16131/j.cnki.cn44-1669/g4.2014.04.005.

Uchanski, R. M. (2005). Clear speech, in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Publishers, Malden, MA), p. 207-235.

Vandali, A. E., & van Hoesel, R. J. (2012). Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking. *The Journal of the Acoustical Society of America*, 132(1), 392-402.

Wang, C. Y. (2011). *Children's Acquisition of Tone 3 Sandhi in Mandarin*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, 7, 627.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' Acoustic Realization of Mandarin Tone Sandhi and Lexical Tones in Context are Not Yet Fully Adult-like. *Frontiers in Psychology*, 9, 817.

Yip, M. (2002). *Tone*. Cambridge University Press.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153-201.

Chapter Two: Phonetic Enhancement of Mandarin Vowels and Tones: Infant-directed Speech and Lombard Speech

This chapter is based on the following published paper:

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(2), 493-503. https://doi.org/10.1121/1.4995998

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors) when needed.

Abstract

Speech units are reported to be hyperarticulated in both infant-directed speech (IDS) and Lombard speech. Since these two registers have typically been studied separately, it is unclear if the same speech units are hyperarticulated in the same manner between these registers. The aim of the present study was to compare the effect of register on vowel and tone modification in the tonal language Mandarin Chinese. Vowel and tone productions were produced by 15 Mandarin-speaking mothers during interactions with their 12-month-old infants during a play session (IDS), in conversation with a Mandarin-speaking adult in a 70 dBA 8-talker babble noise environment (Lombard speech), and in a quiet environment (adult-directed speech). Vowel space expansion was observed in IDS and Lombard speech; however, the patterns of vowel-shift were different between the two registers. IDS displayed tone space expansion only in the utterance-final position, whereas there was no tone space expansion in Lombard speech. The overall pitch increased for all tones in both registers. The tone-bearing vowel duration also increased in both registers, but only in utterance-final position. The difference in speech modifications between these two registers is discussed in light of speakers' different communicative needs.

Keywords: infant-directed speech; Lombard speech; Mandarin vowels; Mandarin tones.

Introduction

Speakers use different speech styles (i.e. registers) in different conditions, and each speech style has its own acoustic characteristics. For example, when talking to an infant, speakers tend to use an exaggerated speech style, resulting in a unique speech register known as "infant-directed speech" or IDS (Kuhl et al., 1997; Burnham, Kitamura & Vollmer-Conna, 2002). In IDS, suprasegmental and segmental characteristics are modified. Relative to adult-directed speech (ADS), IDS exhibits higher overall pitch (Fernald et al., 1989; Burnham, Kitamura & Vollmer-Conna, 2002; Fernald & Simon, 1984), larger pitch variability (Fernald et al., 1989), slower speaking rate (Fernald & Simon, 1984), and enhanced phonemic contrasts (Kuhl et al., 1997; Burnham, Kitamura & Vollmer-Conna, 2002; Liu, Kuhl & Tsao, 2003; Liu, Tsao & Kuhl, 2007). The general consensus is that speech modifications in IDS arise for the following reasons: to attract infants' attention, to communicate positive affect, and to facilitate the infants' language learning (cf. Song, Demuth & Morgan (2010), for discussion).

Another speech style occurs when people are talking in a noisy environment, which is known as Lombard speech (Lombard, 1911). This speech style also displays a set of acoustic enhancements. For example, relative to speech produced in quiet, Lombard speech exhibits higher pitch, greater intensity, longer vowel duration, higher first formant frequency (F1), and a tendency to enhance phonemic contrasts (see Junqua, 1996, for a review). It is suggested that talkers modify speech so as to maintain self-monitoring ability through the speech feedback loop and to improve the communicative effectiveness between speakers in a noisy environment (Zhao & Jurafsky, 2009).

It seems that these two speech registers undergo modifications with the common goal of accommodating to the communicative needs of the listener, and thus result in similar acoustic modification and speech hyperarticulation. However, to our knowledge, there is only one published study that has compared the acoustic modification between IDS and Lombard speech (Wassink, Wright & Franklin, 2007). In this study, Jamaican Creole and Jamaican English vowels were compared across three speech conditions: IDS, Lombard speech, and hyperspeech. Wassink et al. found similar pitch and intensity modifications in IDS and Lombard speech, while the two registers differed in how vowels and segmental duration were modified. It is unknown if, in a tonal language such as Mandarin Chinese which uses both vowels and lexical tones to convey linguistic meaning, these linguistic units also undergo different modifications across registers. This question was addressed in the current study by directly comparing vowel and lexical tone hyperarticulation in Mandarin IDS and Lombard speech.

In IDS, vowel contrasts are often enhanced, leading to vowel space expansion (e.g., Kuhl et al., 1997; Burnham, Kitamura & Vollmer-Conna, 2002). It has been argued that in this "stretched" vowel space, vowel categories are acoustically less overlapped and the vowel contrasts are maximized, which could provide infants with well-specified acoustic cues to help them develop vowel categories (Kuhl et al., 1997). It has also been claimed that the expanded vowel space is a universal characteristic employed to promote language learning (Kuhl et al., 1997), and that IDS reflects caregivers' teaching effort, including the expression of positive affect or attracting infants' attention (Uther, Knoll & Burnham, 2007). However, IDS studies are mixed in terms of the presence vs. absence of vowel space expansion (see Benders, 2013, for a review), with some studies actually reporting a surprisingly compressed vowel spaces in IDS (Englund & Behne, 2006; Benders, 2013). Therefore, whether vowel hyperarticulation is a universal feature of IDS and whether the nature of vowel hyperarticulation is a result of caregivers' teaching effort are still open questions.

IDS has also been shown to alter suprasegmental aspects of speech. Lexical tone languages, which use contrastive tones to indicate the meanings of words (Yip, 2002), exhibit tone hyperarticulation in IDS as well. Evidence from Taiwan Mandarin, Hakka and Cantonese has shown that, relative to ADS, acoustic differences among tones - pitch height differences, pitch range differences, and pitch duration differences - are made more distinct in IDS (Liu, Tsao & Kuhl, 2007; Xu Rattanasone, Burnham & Reilly, 2013; Cheng & Chang, 2014). It is therefore argued that, like vowel hyperarticulation, tone hyperarticulation in IDS could also increase infants' abilities to distinguish language-specific phonemic units and thus facilitate language learning (Liu, Tsao & Kuhl, 2007). It is also claimed that the larger pitch contours of IDS could improve infants' ability to discriminate vowels (Trainor & Desjardins, 2002). However, there is evidence showing that, in Thai IDS addressed to 3-12 month-olds, the tonal contour was less identifiable than in ADS (Kitamura et al., 2002). This divergence among studies challenges the universality of phonemic tone hyperarticulation in IDS, prompting further investigation of another tone language.

Although vowel hyperarticulation (Liu, Kuhl & Tsao, 2003) and tone hyperarticulation (Liu, Tsao & Kuhl, 2007) in IDS have been studied in Taiwan Mandarin, no study has explored IDS in Northern Mandarin, the dialect of Mandarin spoken in northern China (Li, 1973). Taiwan Mandarin and Northern Mandarin share the same vowel and tone inventory, but the phonetic realizations of vowels and tones are different across dialects. For instance, compared with Taiwan Mandarin, Northern Mandarin has a larger vowel space, a larger pitch range, and a higher mean pitch (Deng

& Shi 2006). This implies that, relative to Taiwan Mandarin, vowels and tones in Northern Mandarin might be harder to hyperarticulate in speech registers such as IDS and Lombard speech. Examining possible vowel and tone hyperarticulation in Northern Mandarin therefore provides an ideal opportunity to further examine the universality of phonemic hyperarticulation in IDS.

Similar to IDS, Lombard speech also modifies vowels (Junqua, 1996) and tones (Zhao & Jurafsky, 2009; Kasisopa, Attina & Burnham, 2014). However, studies of Lombard speech have provided conflicting evidence for vowel hyperarticulation, and there are few reports on tone hyperarticulation. For example, there is evidence showing that, relative to speech produced in a quiet environment, vowel space is expanded in Lombard speech (Bond, Moore & Gable, 1989), especially when an interlocutor was present (Cooke & Lu, 2010). However, Davis and Kim (2010) and Godoy, Koutsogiannaki and Stylianou (2014) did not observe vowel space expansion in Lombard speech. In spite of the mixed vowel space expansion results, these studies uniformly reported that formant frequencies, especially F1, are increased in Lombard speech. These results imply that the manifestation of vowel hyperarticulation might be language-dependent. Perhaps speakers of some languages choose to maximize vowel contrasts, whereas others prefer to improve communicative efficiency by reinforcing the speech signal through higher F1.

While most studies on Lombard speech have focused on vowel hyperarticulation, only two studies have examined tone hyperarticulation. On the one hand, Zhao and Jurafsky (2009) did not find enhanced tonal contrasts in Cantonese Lombard speech under 75 dB SPL white noise, although overall pitch was higher for all tones in the noise condition. On the other, Kasisopa, Attina and Burnham (2014) reported enhanced tonal contrast in Thai Lombard speech under the same noise condition as in Zhao and Jurafsky (2009). Although the use of tone hyperarticulation in Thai Lombard speech differed from that in Cantonese, Thai Lombard speech also displayed an overall increase in pitch for all tones in the noise condition, with enhanced tonal contrasts towards the later part of the tone contours. These conflicting findings suggest the need to examine another tone language to test whether tone hyperarticulation is an intrinsic feature of Lombard speech.

The present study was therefore aimed at investigating the phonetic modification of Northern Mandarin vowels and tones in IDS and Lombard speech. In particular, we asked whether vowels and tones would be hyperarticulated in both IDS and Lombard speech in Northern Mandarin. If so, would they be expanded in the same way? To explore these issues, IDS, Lombard speech, and ADS were elicited from the same speakers (15 Northern-Mandarin speaking mothers of 12-month-old infants). Their vowel and tone productions were compared across registers, where the ADS served as the control for comparison. We had the following predictions:

Hypothesis 1 (H1): Vowel space would be expanded in Northern Mandarin IDS and Lombard speech.

Hypothesis 2 (H2): Point vowels would be hyperarticulated in both IDS and Lombard speech, but the direction of the hyperarticulation would differ between the two speech registers: vowels in IDS might be made maximally distinct in the closed-open (F1) and/or in the front-back (F2) dimension, whereas the vowels in Lombard speech might move towards a higher F1.

Hypothesis 3 (H3): Tone space would be expanded in Northern Mandarin IDS as well as in Lombard speech.

Hypothesis 4 (H4): The overall pitch and syllable/word duration for all tones would be increased in both IDS and Lombard speech (Zhao & Jurafsky, 2009; Kasisopa, Attina & Burnham, 2014).

Method

Participants

Fifteen Mandarin-speaking mothers and their 12-month old infants (Mean = 12 months, SD = 0.99) were recruited in Sydney. All the mothers were born and raised in Mandarin-speaking families in Northern China (i.e., Beijing, Hebei Province, or North-eastern China). Their age ranged from 18 to 34 years (Mean = 24 years, SD = 4.93). At the time of study, their stay in Australia ranged from 1 to 8 years (Mean = 5 years, SD = 3.27). All mothers were the main caregivers of their infants and spoke only Mandarin Chinese to their infants at home. According to parental report, no infants tested had speech, hearing, or language problems

Stimuli

Six disyllabic Mandarin words were used to elicit the three target point vowels (/i/, /a/, and /u/) and the three target lexical tones (T1: level, T2: rising, and T4: falling[1]. All were nouns that could be illustrated with toys. The three point vowels were selected to examine the area of vowel space across registers. T1, T2, and T4 were selected because they have simple tone contours and differ from the complex contour tone (T3). The three simple contour tones are distinguished in terms of pitch at the onset and offset

[1] Six target words were selected and elicited in the present study, and additional eight T3, neutral tone and tone sandhi words were used for another study. All these stimuli were elicited in the same experiment.

of the syllable (T1: high pitch onset, high pitch offset; T2: mid pitch onset, high pitch offset; T4: high pitch onset, low pitch offset), which are used to define the tone space, in line with Barry and Blamey (2004) and Xu Rattanasone, Kitamura and Vollmer-Conna (2013). If tone hyperarticulation occurs, the acoustic tone space will also expand.

The three point vowels /i/, /a/, and /u/ occurred in V_1 position of three $C_1V_1.C_2V_2$ disyllabic Mandarin words (see Table I). The target vowels were preceded by a bilabial aspirated stop consonant /p^h/ to minimize co-articulation between C_1 and V_1, while keeping the place of articulation as similar as possible in the following C_2 (i.e. a retroflex aspirated alveo-palatal affricate /tʂ^h/, an aspirated alveolo-palatal affricate/tɕ^h/, and an aspirated alveolar stop /t^h/). Although C_2 for all was not identical across items, we think that this would have minimal effect on the focus of the investigation, namely, the effect of register on the vowel. Having controlled for the preceding consonant and place of articulation of the following consonants, we also controlled for the tone of the first syllable (i.e. T2).

The three target tones T1, T2, and T4 were carried by the second syllable of the test words (see Table 1). The tone of the first syllable was held constant to be T1 (i.e. level) so that the tonal coarticulation across the two syllables could be minimized, and the segment of the second syllable was held constant to be /tʂu/. See Table 1 for a list of all six test items.

Table 1. Target vowels (/i/, /a/, and /u/) and lexical tones (T1, T2, and T4) in disyllabic words. The columns from left to right illustrate the three target vowels and tones, the target words written in

Chinese characters, the corresponding meanings, and the phonetic transcriptions of the target words with the target syllables in bold.

Target vowels/tones	Target words	Meaning	Phonetic transcription
/i/	皮球	Ball	/**pʰi2** tɕʰiou2/
/a/	爬虫	Worm	/**pʰa2** tʂʰuŋ2/
/u/	菩提	Bodhi	/**pʰu2** tʰi2/
T1	珍珠	Pearl	/tʂən1 **tʂu1**/
T2	山竹	Mangosteen	/ʂan1 **tʂu2**/
T4	光柱	Light stick	/kuaŋ1 **tʂu4**/

Procedure

Prior to testing, mothers and their infants were invited into a sound-treated room to become familiar with the new environment. After about 10 minutes, the testing phase began, with each mother taking part in three speech production tasks in the same order: (1) IDS, (2) Lombard speech, (3) ADS. This helped ensure that data collection from the infants took place before they became fussy or tired.

In the IDS task, mothers were fitted with a head-mounted condenser microphone (AKG-C520) which was connected to a solid-state recorder (Marantz PMD661MKII) in a shoulder bag. Mothers wore the shoulder bag so that they were free to move around during the recording session. The same solid-state recorder and head-mounted microphone were used for recording in all three tasks. To elicit IDS, the mother and her infant engaged in a play session using a set of toys. Six toys corresponding to the six target words were provided, and each toy was labelled with the corresponding target word in written Chinese. These toys were randomly arranged into three cloth bags, to ensure counterbalancing of the order of the toys presented across mothers. The mothers were instructed to play with their infants using the toys as they normally would at home, and to use the written labels to refer to the toys when talking to the infant. Similar to the procedures used in previous IDS studies (i.e., Wassink, Wright & Franklin, 2007; Majornano, Rainieri & Corsano, 2012), a Mandarin-speaking experimenter (the first author) was present in the studio to ensure the experiment went successfully. At the same time, the experimenter kept count of the number of tokens the mother produced for each target word. When the mother had produced a minimum of 10 tokens for each target word, the experimenter provided the mother with the second bag of toys. This continued until all three bags of toys were used during the play session.

In the Lombard speech task, only the mother and the experimenter were present in the room. The mother was encouraged to describe to the experimenter her experience in using the labels to refer to the toys in the play session with her infant. The mother and the experimenter talked to one another while listening to a 70 dBA Chinese 8-talker babble noise via open-ear headphones (AKG-K612 PRO). The Digitech-QM1591 Decibel Meter was used to calibrate the sound level played via the headphones. Before the task, the decibel meter was positioned on the headphones to make sure that the sound level of the auditory output from the headphones stayed at 70 dBA. The experimenter was keeping count of tokens the mother produced for each target word while talking to the mother, and the conversation continued until the mother had produced a minimum of 8 repetitions for each toy. In cases when the mother did not produce at least 8 tokens for a toy, the experimenter would ask several questions to help elicit more tokens from the mother, such as "What was the color of X?" and "Did your baby like X?" etc.

In the ADS task, the procedure and the minimal number of repetitions were identical to those in the Lombard speech task. The only difference was that the mother and the experimenter talked to one another in a quiet environment. All the recordings were made with a sampling rate of 44.1 kHz and a 16-bit quantization.

Mothers of 12-month-olds were selected in order to compare our results with previous findings in Taiwan Mandarin IDS (Liu, Kuhl & Tsao, 2003, Liu, Tsao & Kuhl, 2007: 10-12-month-olds). The 70 dBA 8-talker babble noise was used in the present study as the noise mask to elicit Lombard speech in accord with the observation that N-talker babble noise resembles everyday speech noise (Simpson & Cooke, 2005) and with the mean intensity level (70 dBA) of young females' conversational speech (Morris & Brown, 1994).

Coding and Measurements

The data were coded in Praat (Boersma & Weenink, 2016) with the aid of spectrograms and waveforms, as outlined below. Ten percent of the items were re-coded by a second trained native speaker. Interrater reliability was 95.56% for target vowel carriers and 95.77% for target tone carriers, as measured by the correlation of vowel duration between raters.

Vowels

The three target point vowels (/i/, /a/, /u/) across the three register types (IDS, Lombard speech, ADS) were annotated. To identify the beginning and end of the target point vowels, clear F2 onset and offset, respectively, were used. On the basis of the annotated vowel intervals, the trajectories of F1 and F2 were extracted in Praat. Averaged values of the F1 and the F2 for each target point vowel were calculated from the middle portion (from 40% to 60%) of the vowel interval (Figure 1), as the vowel target (i.e. the most stable formants with minimal influence from the formant transitions) is typically reached towards the middle part of the vowel. The Burg method (Burg, 1975) was used in Praat to track the formant values, calculated over a range from 0 to 5500 Hz. The length of the analysis window was 25 ms, and frequencies above 50 Hz were pre-emphasized. Formant values which were mistracked were hand-corrected in Praat, although the influence of this mistracking was small (a correlation analysis of formant values extracted before and after hand-correction yielded high correlation coefficients for F1 and F2: 0.87 and 0.9). All the formant values extracted were then transformed to Bark values to match the scale of human perception, using the following formula (Zwicker & Fastl, 1980):

$$\text{Bark} = 13 * \arctan(0.76*\text{Hertz}) + 3.5 * \arctan(\text{Hertz}/7.5)^2$$

The averaged F1 and F2 values of /i/, /a/, and /u/ were used to derive the vowel space area for each mother/participant in the IDS, Lombard speech, and ADS conditions using the following formula (Liu, Kuhl & Tsao, 2003):

Vowel space area = |{[F1i*(F2a−F2u) + F1a*(F2u−F2i) + F1u*(F2i−F2a)]/2|, where F1i, F1a and F1u are the F1 values of /i/, /a/, and /u/ respectively; F2i, F2a, and F2u are the F2 values of /i/, /a/, and /u/ respectively.

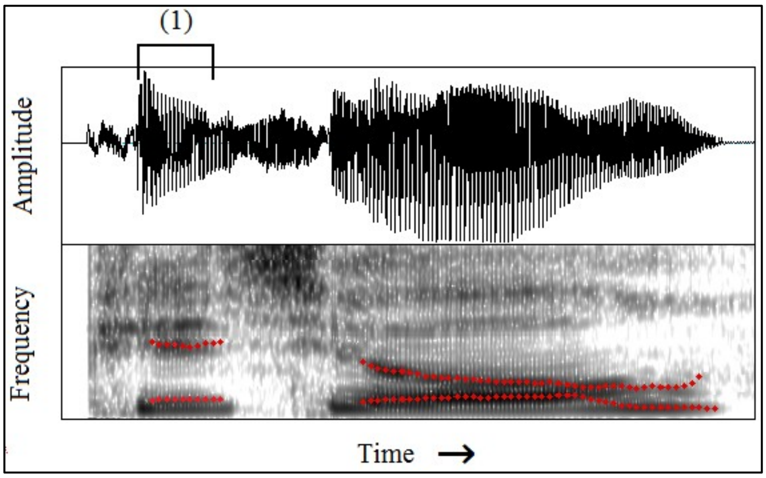


Figure 1. Waveform and spectrogram for the word /pʰi2 tɕʰiou2/. This token was produced by a subject M1 in Lombard speech condition. (1) illustrates the vowel portion of target syllable /pʰi2/. Red dotted lines from the bottom up represent F1 and F2 respectively. The averaged F1 and F2 values of target vowel /i/ were extracted from the 40-60% portion of (1).

Tones

The three target lexical tones (T1, T2 and T4) in the disyllabic words and the three register types (IDS, Lombard speech and ADS) were annotated. Since the realization of Mandarin tones is highly influenced by the tonal context, i.e., the preceding and the following tones (Xu, 1994), the positions of the target tones within utterances were also annotated. The target tones were always carried by the second syllable of the target disyllabic words, so the target tones appeared either in the “utterance-medial” position or the “utterance-final” position. To track the pitch contour of each tone, the onset and offset of the tone-bearing vowel were first identified on the basis of clear F2. Two pitch points (pitch onset and offset) were then measured from the vocalic portion of the target word, with the pitch onset extracted at 5% point and the pitch offset extracted at the 95% point (Figure 2) so that tonal coarticulation and pitch perturbation from neighbouring consonants could be minimised. Pitch was tracked using a short-term autocorrelation algorithm in Praat. Pitch values were checked and manually revised to correct for the “doubling” or “halving” errors in pitch tracking. In the analysis, the pitch values were transformed to semitones from observed Hertz values with 50 Hz as the reference to match the scale of human perception, using the following formula:

Semitone = 12 * log2 (target Hertz/50).

The averaged pitch onset and offset values of T1, T2, and T4 were then used to derive the tone space area in the IDS, Lombard speech, and ADS conditions for each mother/participant. The same formula for calculating the vowel space area was used to compute the tone space area, as below:

Tone space area = |{[T1onset * (T4offset−T2offset) + T2onset * (T1offset − T4offset) + T4onset * (T2offset − T1offset)]/2|. The tonal duration information was also extracted as the duration of the vowel interval (ms) in the target word.

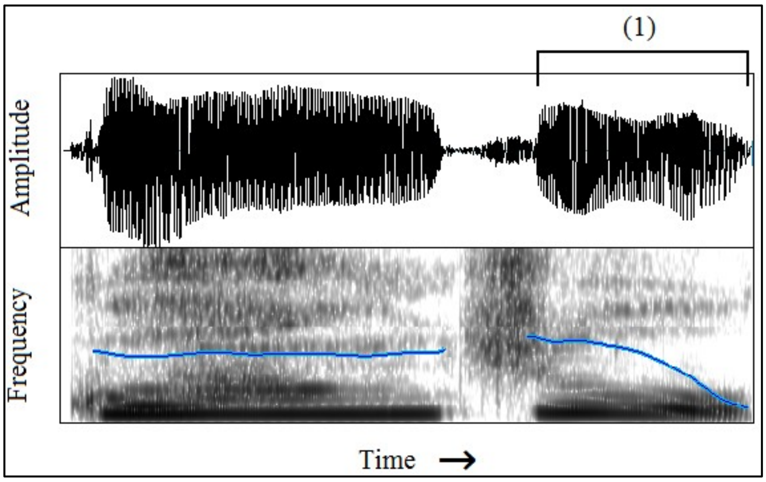


Figure 2. Waveform and spectrogram for the word /kuaŋ1 tʂu4/. This token was produced by subject M1 in the Lombard speech condition. (1) illustrates the vowel portion of target syllable /tʂu4/. Blue lines represent the pitch contours of two syllables. The pitch onset and offset of the target tone were extracted from the 5% and the 95% point of (1) respectively, and the tonal duration was measured across the duration of (1).

Statistical analysis

A total of 2828 tokens were included in the analysis, with 1431 vowel tokens and 1397 tone tokens, as illustrated in Table 2. An additional 141 tokens were excluded from the analysis for the following reasons: overlap with another sound, such as the infant's vocalization or noise made by toys or other environmental disturbance; the mother laughing or singing when producing the token; the token having been whispered; or mispronunciation.

Table 2. Number of vowel and tone tokens included in each register for further analysis. Vowel space and tone space were calculated based on the averaged data per participant.

Target vowels/tones	**Register**		
	IDS	**Lombard speech**	**ADS**
/a/	169	151	146
/i/	241	143	144
/u/	162	140	135
Total vowels	572	434	425
T1	199	152	139
T2	176	139	136
T4	182	135	139
Total tones	557	426	414

The data were analysed using R (R Core Team, 2016). A linear mixed effect model was adopted for the comparison of vowel and tone parameters across registers, using "lme4" package (Bates et al., 2015) and "lmerTest" package (Kuznetsova, 2013). When a significant main effect of a multi-level factor or a significant interaction effect was observed, Tukey-HSD post-hoc comparisons were then also performed on the multi-level factor, as well as interactions.

Results

Vowels

Vowel space area

The F1-F2 two-dimensional vowel space areas of the three registers are presented in Figure 3. To test H1, vowel space areas were compared across registers. A fixed factor "Register" (IDS, Lombard speech, and ADS) and a random factor "Subject" (15 subjects) were included in the model with a random intercept and a random slope for the fixed factor. The significance of the random slope in model fitting was first tested by a likelihood ratio test which showed that the exclusion of the random slope did not significantly affect the fit (χ^2 (5) = 2.571, p = 0.766). Therefore, to keep the model parsimonious, the random slope was excluded from the model in further analyses[2].

The results of the comparison are presented in Table 3. Consistent with H1, the vowel space was expanded in both IDS (Mean = 12.52 $Bark^2$, SD: 1.98; IDS – ADS: β = 1.23, SE = 0.48, t = 2.56, p < 0.05) and Lombard speech (Mean = 12.65 Bark2, SD = 1.43; Lombard speech – ADS: β = 1. 63, SE = 0.48, t = 2.82, p < 0.01) relative to ADS (Mean = 11.28 $Bark^2$, SD = 1.68).

[2] The R code of this model is: Vowel space area ~ Register + (1 | Subject).

Table 3. Results of the test of effects of the fixed factors "Register" (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison) on vowel space using a linear-mixed effects model. Estimated differences across registers, standard errors, degrees of freedom, t values, and *p* values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

	Estimate	**SE**	**df**	**t**	***p***
(intercept)	11.28	0.44	31.66	25.52	<0.001*
Register (IDS)	1.23	0.48	28	2.56	0.016*
Register (Lombard)	1.36	0.48	28	2.82	0.009*

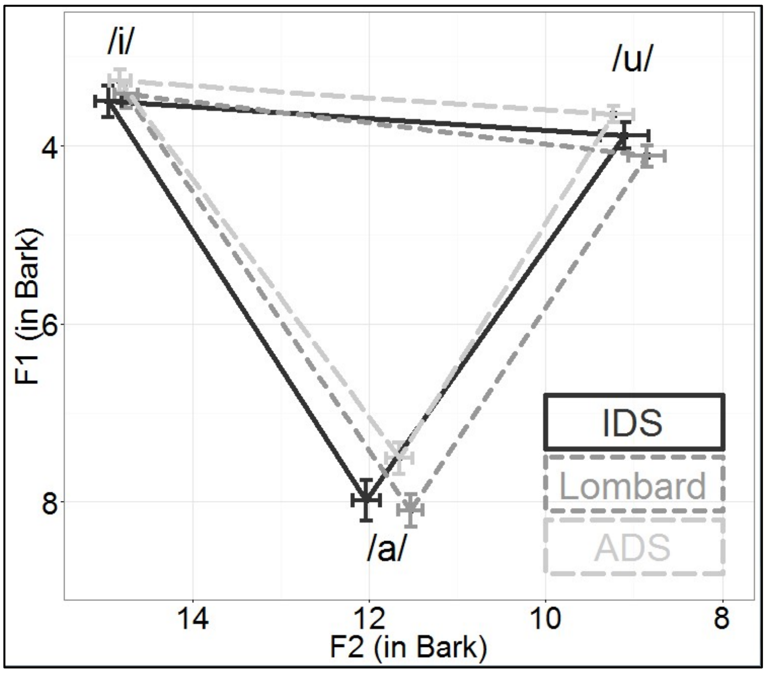


Figure 3. The vowel space, defined by F1 and F2 in Bark, with the three point vowels /i/, /a/ and /u/ in infant-directed speech (IDS), Lombard speech (Lombard) and adult-directed speech (ADS). The corners of each space represent the group means of averaged F1 and F2 values for /i/, /a/ and /u/ respectively, with error bars showing the standard error of the group means of F1 and F2 values.

Formant frequencies

Table 4 gives the F1 and F2 values across vowels and registers. To examine H2, F1 and F2 values of /i/, /a/, and /u/ were compared across the three registers. Two fixed factors "Register" and "Vowel" (/i/, /a/, and /u/) and the random factor "Subject" were included in models for F1 and F2 comparison, with random intercept and random slopes for the fixed factors. Since vowel duration has an effect on F1 and F2 (Gay, 1978), "Vowel Duration" (the duration of vowels, computed as the interval between vowel onset and offset) was also entered into the models as a covariate to minimise this effect. The random slopes were checked for significance by the likelihood ratio test for each model. In both models, the random slope of "Subject" for "Register" was excluded since it did not reach significance in model fitting (in the model for F1 comparison: χ^2 (9) = 16.03, p = 0.066; in the model for F2 comparison: χ^2 (9) = 10.087, p = 0.344)[3].

Table 4. Mean F1 and F2 (Bark) values (SD) across vowels (/a/, /i/, /u/) and registers (IDS, Lombard speech, ADS).

Register	**F1**			**F2**		
	/a/	**/i/**	**/u/**	**/a/**	**/i/**	**/u/**
IDS	7.92 (0.76)	3.41 (0.56)	3.92 (0.51)	11.98 (0.63)	14.88 (0.53)	9.10 (1.33)
Lombard speech	8.06 (0.63)	3.44 (0.45)	4.11 (0.40)	11.49 (0.44)	14.75 (0.44)	8.89 (0.90)
ADS	7.54 (0.75)	3.29 (0.42)	3.67 (0.34)	11.64 (0.62)	14.80 (0.39)	9.25 (1.05)

[3] The R code for these models is: F1/F2 ~ Register * Vowel * Vowel Duration + (1 + Vowel | Subject).

The results of the comparisons are presented in Table 5[4]. The main effect of "Register" was significant for both F1 and F2, which means that vowel formant frequencies differ across registers. Since a significant two-way interaction between "Register" and "Vowel" was observed, a Tukey-HSD post-hoc test was then performed on the model to compare the register difference in this interaction. The results of the post-hoc test are presented in Table 6.

[4] The main effect of the covariate "Vowel Duration" and its interactions with other factors were not presented since they were not the main interest of the current study.

Table 5. Results of the test of effects of the fixed factors "Register" (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison) and "Vowel" (/i/, /a/, and /u/, where /a/ was the reference level for comparison) on F1 and F2 using a linear-mixed effects model. Estimated differences, standard errors, degrees of freedom, t values, and *p* values are provided, where * indicates statistical significance with an alpha value of *p* = 0.05.

	F1					F2				
	Estimate	**SE**	**df**	**t**	***p***	**Estimate**	**SE**	**df**	**t**	***p***
(intercept)	**6.38**	**0.16**	**96**	**40.04**	**<0.001*** **	**11.82**	**0.17**	**176**	**70.63**	**<0.001*** **
Register (IDS)	**1.12**	**0.14**	**1396**	**7.79**	**<0.001*** **	**0.40**	**0.16**	**1392**	**2.43**	**<0.05***
Register (Lombard)	**0.86**	**0.16**	**1389**	**5.33**	**<0.001*** **	-0.25	0.19	1397	-1.34	0.182
Vowel (/i/)	**-3.12**	**0.21**	**166**	**-15.02**	**<0.001*** **	**2.70**	**0.21**	**730**	**13.04**	**<0.001*** **
Vowel (/u/)	**-2.79**	**0.19**	**112**	**-14.48**	**<0.001*** **	**-0.93**	**0.21**	**172**	**-4.43**	**<0.001*** **
Register (IDS) : Vowel (/i/)	**-1.01**	**0.20**	**1402**	**-5.14**	**<0.001*** **	-0.16	0.22	1340	-0.72	0.470
Register (IDS) : Vowel (/u/)	**-0.74**	**0.18**	**1400**	**-4.00**	**<0.001*** **	**-0.60**	**0.21**	**1403**	**-2.85**	**<0.01** **
Register (Lombard) : Vowel (/i/)	**-1.04**	**0.23**	**1392**	**-4.51**	**<0.001*** **	0.20	0.27	1399	0.77	0.442
Register (Lombard) : Vowel (/u/)	**-0.49**	**0.22**	**1396**	**-2.27**	**<0.05***	0.17	0.25	1396	0.68	0.500

According to Table 6, IDS showed higher F1 and F2 values relative to ADS for /a/ and /u/, thus leading to a downward and forward shift of the acoustic vowel space.

This was inconsistent with H2, which predicted vowels in IDS could be made maximally distinct in the closed-open (F1) and/or the front-back (F2) dimensions. In contrast, Lombard speech showed a higher F1 value for all three vowels, which led to a downward shift of the acoustic vowel space, consistent with H2.

Table 6. Tukey-HSD pairwise comparisons of the different levels of "Register" (IDS, Lombard speech, and ADS) and the two-way interaction effect between "Register" and "Vowel" (/i/, /a/, and /u/) on F1 and F2. The estimated difference between the two levels of "Register", standard error, t values, and HSD-adjusted *p* values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

Register difference	**F1**				**F2**			
	Estimate	**SE**	**t**	***p***	**Estimate**	**SE**	**t**	***p***
				/i/				
IDS-ADS	0.14	0.06	2.17	0.077	0.01	0.07	0.17	0.984
Lombard-ADS	**0.19**	**0.07**	**2.86**	**<0.05***	-0.12	0.08	-1.48	0.299
IDS-Lombard	-0.06	0.05	-1.17	0.471	0.13	0.06	2.25	0.063
				/a/				
IDS-ADS	**0.47**	**0.05**	**8.66**	**<0.001*** **	**0.39**	**0.06**	**6.23**	**<0.001*** **
Lombard-ADS	**0.45**	**0.06**	**7.53**	**<0.001*** **	0.13	0.07	1.83	0.160
IDS-Lombard	-0.01	0.06	-0.19	0.981	**0.51**	**0.07**	**7.57**	**<0.001*** **
				/u/				
IDS-ADS	**0.24**	**0.06**	**4.21**	**<0.001*** **	**0.23**	**0.07**	**3.42**	**<0.01** **
Lombard-ADS	**0.47**	**0.06**	**7.89**	**<0.001*** **	-0.02	0.07	-0.31	0.949
IDS-Lombard	**-0.22**	**0.05**	**-4.32**	**<0.001*** **	**0.21**	**0.06**	**3.43**	**<0.01** **

Tones

Tone space area

The pitch onset-offset two-dimensional tone space areas in different positions (utterance-medial and utterance-final) across the three registers are presented in Figure 4. To test H3, the tone space areas were compared across registers. Two fixed factors "Register" and "Position" (medial and final) and the random factor "Subject" were included in the model, where the random factor had its own intercept as well as random slopes for fixed factors. The significance of the random slopes was first checked and showed that all slopes were significant in model fitting. Therefore, all random slopes were included in the model[5].

The results of the comparisons are presented in Table 7. The results showed that, relative to ADS, tone space was expanded in the utterance-final position of IDS. To further examine the tone space expansion across registers and positions, a Tukey-HSD post-hoc test was also performed on the model, and the results of the post-hoc test are presented in Table 8.

Partially consistent with H3, tone space expansion was only observed in IDS (utterance-medial: Mean = 2.17 St^2, SD = 0.58; utterance-final: Mean = 8.18 St^2, SD = 2.55) in the utterance-final position compared with ADS (utterance-medial: Mean = 1.10 St^2, SD = 0.17; utterance-final: Mean = 2.29 St^2, SD = 0.52). Unexpectedly, Lombard speech (utterance-medial: Mean = 1.33 St^2, SD = 0.26; utterance-final: Mean

[5] The R code for this model is: Tone space area ~ Register * Position + (1 + Register + Position | Subject).

= 4.03 St^2, SD = 0.64) did not show tone space expansion relative to ADS in either position.

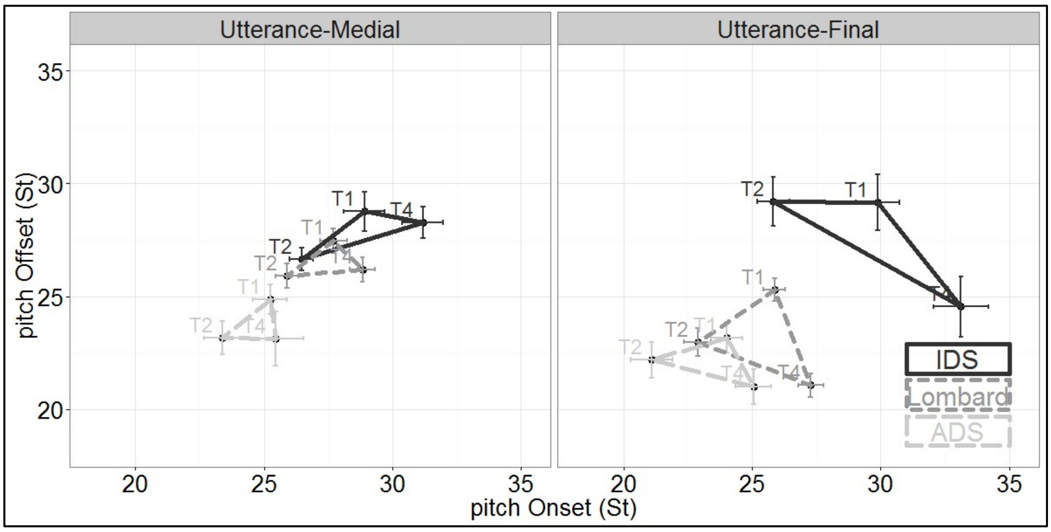


Figure 4. The tone space, defined by pitch onset and offset in semitones, with the three lexical tones T1, T2 and T4 in utterance-medial and -final positions, in IDS, Lombard (speech) and ADS. The corners of each space represent the group means of averaged pitch onset and offset values for T1, T2 and T4 respectively, with error bars showing the standard errors of the group means of pitch onset and offset values.

Table 7. Results of the test of the effects of "Register" (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison) and "Position" (utterance-medial and utterance-final, where the utterance-medial was the reference level) on the tone space area (St^2) using a linear-mixed effects model. Estimated differences, standard errors, degrees of freedom, t values, and p values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

	Estimate	**SE**	**df**	**t**	***p***
(intercept)	1.87	2.09	40.75	0.89	0.377
Register (IDS)	2.42	3.88	26.37	0.62	0.538
Register (Lombard)	0.53	2.68	74.10	0.20	0.845
Position (final)	2.79	3.15	48.78	0.89	0.380
Register (IDS) : Position (final)	**13.33**	**3.78**	**75.00**	**3.53**	**<0.001*** **
Register (Lombard) : Position (final)	2.76	3.78	75.00	0.73	0.467

Table 8. Tukey-HSD pairwise comparisons on different levels of "Register" (IDS, Lombard speech, and ADS) in the two-way interaction effect between "Register" and "Position" (utterance-medial and utterance-final) on tone space area. The estimated difference between the two levels of "Register", standard error, t values, and HSD-adjusted *p* values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

Register difference	Utterance-medial				Utterance-final			
	Estimate	**SE**	**t**	***p***	**Estimate**	**SE**	**t**	***p***
IDS-ADS	2.42	4.02	0.62	0.808	**15.75**	**3.88**	**4.06**	**<0.001*** **
Lombard-ADS	0.53	2.68	0.20	0.979	3.30	2.68	1.23	0.440
IDS-Lombard	1.90	3.76	0.51	0.870	**12.46**	**3.76**	**3.32**	**<0.01** **

Pitch height and tonal duration

Table 9 gives the mean pitch and duration of the tone-bearing vowel across tones and registers. To test H4, the mean pitch and duration were compared across registers. Two models were built for the comparison of mean pitch and tone-bearing vowel duration respectively. In both models, three fixed factors "Register", "Position" and "Tone" (T1, T2, and T4) and the random factor "Subject" were included, where the random factor had its own intercept as well as random slopes for fixed factors. The random slopes were checked for significance by the likelihood ratio test for each model. In the model for mean pitch comparison, the random slope of "Subject" for "Position" was excluded since it did not reach significance in model fitting (χ^2 (6) = 4.857, p =

0.562)[6]; in the model for tone-bearing vowel duration, the random slope of "Subject" was excluded for "Tone" since it did not reach significance in model fitting (χ^2 (11) = 17.833, p = 0.086)[7].

Table 9. Mean pitch height (St) and tonal duration (s) values (SD) for tones (T1, T2, T4), positions (utterance-medial and utterance-final), and registers (IDS, Lombard speech, ADS).

Register	**Position**	**Pitch height**			**Tonal duration**		
		T1	**T2**	**T4**	**T1**	**T2**	**T4**
IDS	Medial	28.57 (4.45)	26.33 (3.68)	29.43 (4.56)	0.17 (0.08)	0.16 (0.05)	0.18 (0.07)
	Final	29.53 (4.97)	26.92 (5.06)	29.46 (6.29)	0.31 (0.15)	0.30 (0.15)	0.30 (0.16)
Lombard speech	Medial	27.74 (2.43)	25.74 (2.17)	27.49 (2.29)	0.16 (0.06)	0.17 (0.08)	0.18 (0.07)
	Final	25.27 (3.15)	22.54 (3.25)	24.10 (3.34)	0.29 (0.06)	0.31 (0.07)	0.26 (0.06)
ADS	Medial	24.97 (3.34)	23.32 (3.47)	24.37 (5.25)	0.16 (0.05)	0.16 (0.06)	0.17 (0.06)
	Final	23.71 (4.36)	20.93 (4.56)	23.19 (4.47)	0.26 (0.06)	0.29 (0.07)	0.22 (0.07)

[6] The R code for this model is: Pitch height ~ Register * Position * Tone + (1 + Register + Tone | Subject).

[7] The R code for this model is: Tonal duration ~ Register * Position * Tone + (1 + Register + Position | Subject).

The results of the comparisons are presented in Table 10. Since a significant two-way interaction between "Register" and "Position" was observed in both models, a Tukey-HSD post-hoc test was performed on the two models to further compare the register difference across positions. The results are presented in Table 10I and Table 12. Other interactions were not further examined since they were not the focus of the present study.

Partially consistent with H4, the mean pitch of tones was highest in IDS, followed by Lombard speech, and then the lowest for ADS. The only exception was observed in utterance-medial position, when there was only a trend for the mean pitch to be higher in IDS than Lombard speech ($p = 0.068$). The tone-bearing vowel duration was also longer in both IDS and Lombard speech than in ADS, but only in utterance-final position.

Table 10. Results of the test of the effects of "Register" (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison), "Position" (utterance-medial and utterance-final, where utterance-medial was the reference level for comparison) and "Tone" (T1, T2, and T4, where T1 was the reference level for comparison) on pitch height and tonal duration using a linear-mixed effects model. Estimated differences, standard errors, degrees of freedom, t values, and p values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

	Pitch height					**Tonal duration**				
	Estimate	**SE**	**df**	**t**	***p***	**Estimate**	**SE**	**df**	**t**	***p***
(intercept)	**25.08**	**0.80**	**23**	**31.21**	**<0.001*** **	**0.16**	**0.01**	**141**	**15.57**	**<0.001*** **
Register (IDS)	**3.60**	**0.82**	**37**	**4.40**	**<0.001*** **	0.01	0.01	114	0.90	0.373
Register (Lombard)	**2.51**	**0.64**	**169**	**3.92**	**<0.001*** **	0.01	0.01	400	0.48	0.630
Position (final)	**-1.68**	**0.64**	**378**	**-2.61**	**<0.01** **	**0.10**	**0.02**	**109**	**6.24**	**<0.001*** **
Tone (T2)	**-2.06**	**0.64**	**342**	**-3.24**	**<0.001*** **	0.00	0.01	376	-0.27	0.784
Tone (T4)	-0.84	0.67	142	-1.26	0.211	0.01	0.02	137	0.80	0.423
Register (IDS) : Position (final)	**2.37**	**0.81**	**1339**	**2.93**	**<0.01** **	0.03	0.02	1343	2.08	<0.05*
Register (Lombard) : Position (final)	-0.50	0.86	1333	-0.59	0.556	0.02	0.02	1345	0.85	0.395
Register (IDS) : Tone (T2)	-0.40	0.79	1331	-0.50	0.616	-0.01	0.02	1333	-0.61	0.540
Register (Lombard) : Tone (T2)	0.06	0.84	1334	0.07	0.944	0.01	0.02	1339	0.47	0.639

Register (IDS) : Tone (T4)	1.79	0.79	1333	1.16	0.261	-0.01	0.02	1340	-0.40	0.690
Register (Lombard) : Tone (T4)	0.68	0.86	1336	0.79	0.428	0.00	0.02	1343	0.12	0.908
Position (final) : Tone (T2)	-0.45	0.87	1336	-0.52	0.604	0.03	0.02	1337	1.65	0.099
Position (final) : Tone (T4)	0.26	0.87	1335	0.30	0.768	**-0.05**	**0.02**	**1337**	**-2.80**	**<0.01****
Register (IDS) : Position (final) : Tone (T2)	0.48	1.16	1341	0.42	0.677	-0.01	0.03	1342	-0.38	0.708
Register (Lombard) : Position (final): Tone (T2)	-0.49	1.23	1335	-0.40	0.691	-0.01	0.03	1346	-0.38	0.704
Register (IDS): Position (final) : Tone (T4)	-1.13	1.15	1339	-0.98	0.325	0.04	0.03	1343	1.46	0.143
Register (Lombard) : Position (final) : Tone (T4)	-1.30	1.23	1337	-1.06	0.290	0.01	0.03	1346	0.48	0.628

Table 11. Tukey-HSD pairwise comparisons on different levels of "Register" (IDS, Lombard speech, and ADS) in the two-way interaction effect between "Register" and "Position" (utterance-medial and utterance-final) on pitch height. The estimated difference between two levels of "Register", standard error, t values, and HSD-adjusted *p* values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

Register difference	**Utterance-medial**				**Utterance-final**			
	Estimate	**SE**	**t**	***p***	**Estimate**	**SE**	**t**	***p***
IDS-ADS	**4.06**	**0.70**	**5.77**	**<0.001*** **	**6.22**	**0.71**	**8.73**	**<0.001*** **
Lombard-ADS	**2.78**	**0.44**	**6.28**	**<0.001*** **	**1.68**	**0.45**	**3.73**	**<0.001*** **
IDS-Lombard	1.28	0.55	2.33	0.068	**4.54**	**0.56**	**8.10**	**<0.001*** **

Table 12. Tukey-HSD pairwise comparisons on different levels of "Register" (IDS, Lombard speech, and ADS) in the two-way interaction between "Register" and "Position" (utterance-medial and utterance-final) on tonal duration. The estimated difference between two levels of "Register", standard error, t values, and HSD-adjusted *p* values are provided, where * indicates statistical significance with an alpha value of $p = 0.05$.

Register difference	**Utterance-medial**				**Utterance-final**			
	Estimate	**SE**	**t**	***p***	**Estimate**	**SE**	**t**	***p***
IDS-ADS	0.01	0.01	0.66	0.786	**0.05**	**0.01**	**4.65**	**<0.001*** **
Lombard-ADS	0.01	0.01	1.14	0.496	**0.03**	**0.01**	**3.01**	**<0.01** **
IDS-Lombard	-0.01	0.01	-0.27	0.961	0.02	0.01	1.73	0.209

Discussion

The results of this study showed that, in Northern Mandarin, vowels were hyperarticulated in both IDS and Lombard speech in terms of vowel space expansion, while the patterns of vowel-shift were different between the two registers: IDS showed an increase in both F1 and F2 of /a/ and /u/, and Lombard speech showed an increase in F1 for all three point vowels. IDS displayed tone space expansion in utterance-final position but not utterance-medial position, whereas Lombard speech showed no tone space expansion. In spite of the difference in tone space expansion between the two registers, the overall pitch increased for all tones in both registers. The tone-bearing vowel duration also increased in both registers but only in utterance-final position. These results indicate that IDS and Lombard speech modify vowels and tones in somewhat different ways.

The expanded vowel space in IDS and Lombard speech found in the current study is consistent with previous studies on IDS (e.g., Kuhl et al., 1997; Liu, Kuhl & Tsao, 2003) and Lombard speech (e.g., Bond, 1989). Our results show that, although Northern Mandarin has a larger vowel space than Taiwan Mandarin, the vowel space still expands in Northern Mandarin IDS, suggesting that the extent of vowel contrast enhancement in IDS is independent of vowel inventory size. This is similar to a previous study of clear speech (Smiljanić & Bradlow, 2005) which observed equivalent vowel space expansion in English and Croatian clear speech, despite the difference in vowel inventory size between the two languages. In the case of Lombard speech, the expanded vowel space also suggests that vowels are adjusted to be maximally contrastive so as to provide listeners with more distinctive vowel information and improve the speech intelligibility when the communicative environment is noisy.

Although the vowel space expands in both registers, the way this is achieved differs. In IDS, F1 and F2 increased for /a/ and /u/, but remained unchanged for /i/. This resulted in vowel space rotation around /i/, with a downward and forward shift (higher F1 and F2 respectively). Since /a/ moved further than /u/ during the vowel space rotation, this led to an expanded vowel space. In Lombard speech, F1 increased in all three point vowels, whereas F2 remained unchanged. The vowel space was therefore shifted downward in Lombard speech, relative to ADS. Since the extent of F1-increase differed across the three vowels, i.e., /a/ = /u/ > /i/, with a larger F1 shift in /a/ and /u/ than /i/, this also resulted in vowel space expansion (see Figure 3).

In spite of the enhanced vowel contrasts in both IDS and Lombard speech, the different vowel-shift patterns observed for the two styles suggest that the mechanisms underlying these vowel shifts are different. The vowel modifications in IDS may be driven by mothers' articulatory adjustment to express positive affect to infants during the mother-infant interaction. The vowel modifications in Lombard speech, in contrast, may be the result of the speakers' articulatory adjustment to increase intensity in a noisy environment.

Infants are likely to be attracted to positive affect as expressed in IDS, such as joy (Singh, Morgan & Best, 2002). There are studies suggesting that, during mother-infant interaction, positive affect is not only expressed through the mothers' voices but also their facial expressions (Stern, 1974; Chong et al., 2003). It is claimed that two of the predominant facial expressions in IDS to convey positive affect are an open 'surprised' mouth and a joyful smile, often with a slightly opened mouth (Benders, 2013). A more open month in IDS could result in jaw-lowering, thus leading to the increase in F1, especially for the low and back vowels (in our results, the F1 of /i/ in IDS is also affected by the jaw-lowering, although the effect is a trend: $p = 0.077$). The other facial expression is ‘spread lips’ in IDS, which could lead to F2 increase, especially for /a/ and /u/, since the lips are already spread when producing /i/, at least in the Northern Mandarin in this study. The phonetic outcomes of these articulatory modifications, such as the increased F1 and F2 values for the low and back vowels in the current study, are also consistent with the acoustic features of happy speech (Kienast & Sendlmeier, 2000).

In contrast, when talking in a noisy environment, speakers usually speak more loudly to maintain self-monitoring ability and to transmit the speech signals more effectively (Zhao & Jurafsky, 2009). To increase loudness, speakers adjust their articulatory gestures with acoustic consequences. It is suggested that one of the most commonly used articulatory adjustments made to speak more loudly in noisy environments is to increase the degree of jaw-lowering (Schulman, 1989), which can lead to an overall increase in F1. Godoy, Mayo and Stylianou (2013) also suggested that, to increase loudness, another possible articulatory adjustment is to contract the vocal muscles and to move the vocal tract resonances closer together, which results in decreased separation between formant frequencies, i.e., a decrease in F1-F2 and F2-F3. The phonetic outcome of such articulatory adjustments is largely consistent with our results (i.e., an increased F1 and a decreased F1-F2, which was a result of the increased F1 and the unchanged F2), which implies that vowel modifications in Lombard speech are presumably the result of articulatory adjustments made to increase intensity in a noisy environment.

With respect to our results in tone hyperarticulation, it is surprising that the tone space of IDS only expands in utterance-final position. Note that previous Taiwan Mandarin and Cantonese IDS studies did not take the positional effect into consideration when reporting enhancement of tonal contrasts. There are two possible reasons for the position effect in tone hyperarticulation. First, as reported in previous studies, pitch information is one of the most important carriers of affect (Pell et al., 2009). In a tonal language, such as Mandarin, when speakers express 'happy' affect, overall pitch is increased, and the slope of the pitch contour of the final syllable of each prosodic word is exaggerated, especially for syllables in utterance-final position (Wang, Li & Fang, 2005). These acoustic features of Mandarin 'happy' speech are highly consistent with our results. The tone hyperarticulation and vowel hyperarticulation findings converge, suggesting that speech modifications in IDS are mainly driven by the mothers' communicative goal to convey positive affect.

Second, tonal contrasts may not have been enhanced in utterance-medial position because of the tone context. It is well-known that Mandarin tone realization is heavily influenced by adjacent tones, as reported in the tone carry-over effect and the tone anticipatory effect (Xu, 1997). The tone contours of T2 and T4 are prone to modification when they occur in a "conflicting tone context" (Xu, 1994), leading to a possible 'neutralization' of the tone contour. For instance, the conflicting context for T2 (rising tone) refers to a situation where the T2 syllable is preceded by a syllable with a high pitch offset and followed by a syllable with a low pitch onset; for T4 (falling tone), this refers to a context where a T4 syllable is preceded by a syllable with a low pitch offset and followed by a syllable with a high pitch onset. In the present study, the tone preceding the target tones was controlled (i.e. T1). Since we could not control for the tone following the targets in utterance-medial position, this led to conflicting tone contexts in some cases, such as when a target T2 is followed by a T2. This will minimize the tonal contrast between the target T2 and T1. As a result, the tone space was less distinct between the registers when the target tones occurred in the utterance-medial position.

It is worth noting that Trainor and Desjardins (2002) observed that, in English IDS, the exaggerated pitch contour could aid infants' acquisition of vowel categories, while the increased pitch height could serve affective or attentional functions. Therefore, it is also possible that the expanded tone space in the present study is related to mothers' didactic intention to help infants in learning vowels.

It is also surprising to find no tone space expansion in Lombard speech. It should be noted that the task of Lombard speech in our experiment is different from previous Lombard speech studies of Cantonese (Zhao & Jurafsky, 2009) and Thai (Kasisopa, Attina & Burnham, 2014), which only adopted a reading task with one speaker. The present study, in contrast, used a communicative task where two speakers are engaged in a conversation. The task difference may explain the inconsistent results between the present study and previous studies such as Kasisopa, Attina and Burnham (2014), which observed tone space expansion in Lombard speech. Perhaps, in the noisy environment, speakers increase their effort, and this manifests in terms of increasing the overall pitch of tones, rather than exaggerating the tonal contrast, to convey tone units more effectively. Alternatively, it is also possible that the exaggerated pitch contour found in IDS in the present study was simply a result of mothers expressing positive affect, which is expected in IDS but not Lombard speech.

The increased overall pitch in IDS and Lombard speech found in our study is consistent with other findings in the literature (e.g., Fernald et al., 1989; Junqua, 1996). It is generally agreed that the increased pitch height in IDS is associated with the communication of positive affect to infants (see Xu Ratanasone, et al., 2013, for a review), while it is claimed the increased overall pitch in Lombard speech is merely an irrelevant by-product of an increased vocal effort that serves to increase the relative intensities of the higher-frequency components in the speech spectrum (Uchanski, 2005).

Although there was no difference in the overall tone-bearing word duration between the two registers, the tone-bearing word duration was longer in IDS and Lombard speech than in ADS in utterance-final position. This duration pattern resembles the well-documented final-lengthening phenomenon found in IDS (Morgan, Meier & Newport, 1987). Perhaps mothers exaggerate word duration in utterance-final position in IDS to facilitate word segmentation for their infants, in line with the prosodic bootstrapping hypothesis (Morgan, Meier & Newport, 1987). Similarly, the increase in utterance-final word duration in Lombard speech could also aid listeners in identifying word units in a noisy environment and thus improve communicative effectiveness.

However, it should be acknowledged that there are several potential limitations of the present study. First, the use of one token for each target vowel and target tone could potentially limit the extent of generalizability. We did our best to identify words in Mandarin Chinese that could meet the criterion of our stimuli: picturable disyllabic triads that only differed in vowels or tones. Second, the order of the three tasks was fixed: IDS, Lombard speech, and ADS. Keeping the ADS last may potentially increase the contrast between ADS and other registers due to previous mention or fatigue effects. However, most of the present results are consistent with previous findings regarding vowel and tone hyperarticulation. In addition, the focus of the present study was to examine the dimensions of vowel and tone hyperarticulation across registers rather than the degree of hyperarticulation.

Conclusion

Speech units such as vowels and tones are reported to be hyperarticulated in both IDS and Lombard speech, however it has been unclear if the same acoustic parameters might be hyperarticulated in a similar manner between these two registers. The results of this study show that vowels and tones were hyperarticulated in different ways in Northern Mandarin between IDS and Lombard speech. Vowels and tones in IDS are modified to express positive affect, whereas vowel and tone modification in Lombard speech is mainly aimed at transmitting speech signals more effectively in noisy environments. According to Hazan and Baker (2011), speech production is listener-focused, and talkers modulate their speech according to their interlocutors' needs. Our results support this view and also reveal that speakers modify their speech to cater for different communicative needs of the addressee, thus providing a more in-depth understanding of the nature of these speech registers and speech enhancement more generally.

Acknowledgements

We thank Titia Benders, Carmen Kung, the Child Language Lab, the Phonetics Lab, and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and supports. We thank Xin Cheng for helping with the reliability check. We also acknowledge the comments and suggestions from the editor and anonymous reviewers which significantly improved the manuscript. This research was supported, in part, by a Macquarie University iMQRES scholarship to the first author, the following grants: ARC CE110001021, ARC FL130100014. The equipment was funded by MQSIS 9201501719.

References

Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *The Journal of the Acoustic Society of America*, 116, 1739-1748.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package, version 1.1–9, https://CRAN.Rproject.org/package=lme4.

Benders, T. (2013). Mommy is only happy! Dutch mothers realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behaviour and Development*, 36, 847–862.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from http://www.praat.org/.

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic–phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *The Journal of the Acoustic Society of America*, 85, 907–912.

Burg, J. P. (1975). *Maximum Entropy Spectrum Analysis*, Ph.D. thesis, Stanford University, Stanford, CA.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. *Science*, 296, 1435.

Cheng, M.-C., & Chang, K.-C. (2014). Tones in Hakka infant-directed speech: An acoustic perspective. *Language and Linguistics*, 15, 341-390.

Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, 12, 211–232.

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustic Society of America*, 128, 2059–2069.

Davis, C., & Kim, J. (2010). Is speech produced in noise more distinct and/or consistent? in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, (Sydney, Australia), pp.46–49.

Deng, D & Shi, F. (2006) The comparison on phonetic system between Putonghua and Taiwan Mandarin, in *Mountain Lofty, River Long: Festschrift in Honor of Professor Pang-hsin Ting on His Seventieth Birthday*, edited by D. A. He (Academia Sinica Press, Taipei, Taiwan), pp. 371–393. (in Chinese)

Englund, K., & Behne, D. M. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development*, 15, 139–160.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477–501.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustic Society of America*, 63, 223–230.

Godoy, E., Mayo, C., & Stylianou, Y. (2013). Linking loudness increases in normal and lombard speech to decreasing vowel formant separation, In *Proceedings of INTERSPEECH*, Lyon, France, pp. 133–137.

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. *Computer Speech and Language*, 28, 629–647.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustic Society of America*, 130, 2139-2152.

Junqua, J.-C. (1996). The influence of acoustics on speech production: a noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, 20, 13–22.

Kasisopa, B., Attina, V., & Burnham, D. (2014). The Lombard effect with Thai lexical tones: an acoustic analysis of articulatory modifications in noise, *In Proceedings of INTERSPEECH*, Singapore, pp. 1717-1721.

Kienast, M., & Sendlmeier, W. F. (2000). Acoustical analysis of spectral and temporal changes in emotional speech, in *Proceedings of the ISCA Workshop on Emotion and Speech*, Belfast, Northern Ireland, pp. 92–94.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behaviour and Development*, 24, 372–392.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684–686.

Kuznetsova, A. (2013). Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package), R Package, version 2.0-29, pp. 1–17. Available at https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf (Last viewed 5/4/2016).

Li, F. K. (1973). Languages and dialects of China. *Journal of Chinese Linguistics*, 1, 1–13.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6, F1–F10.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43, 912–917.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Annals Maladiers Oreille, Larynx, Nez, Pharynx*, 37, 101–119.

Majorano, M., Rainieri, C., & Corsano, P. (2013). Parents' child-directed communication and child language development: A longitudinal study with Italian toddlers. *Journal of Child Language*, 40, 836–859.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498–550.

Morris, R. J., & Brown Jr. W. S. (1994). Age-related differences in speech intensity among adult females. *Folia Phoniatrica et Logopaedica*, 46, 64–69.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37, 417–435.

R Core Team (2016). R: A language and environment for statistical computing [Computer program]. Version 3.3.1, retrieved 21 June 2016 from https://www.R-project.org/

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustic Society of America*, 85, 295–312.

Simpson, S., & Cooke, M. P. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustic Society of America*, 118, 2775–2778.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk?. *Infancy*, 3, 365–394.

Smiljanić, R., & Bradlow, A. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustic Society of America*, 116, 2627-2627.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustic Society of America*, 128, 389–400.

Stern, D. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal and gaze behaviors, In *The Effect of The Iinfant on its Caregiver*, edited by M. Lewis, and L. Rosenblum (Wiley, New York), p. 187-232.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review*, 9, 335–340.

Uchanski, R. M. (2005). Clear speech, in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Publishers, Malden, MA), p. 207–235.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? Similarities and differences in speech to foreigners and infants. *Speech Communication*, 49, 1–7.

Wang, H., Li, A., & Fang, Q. (2005). F0 contour of prosodic word in happy speech of Mandarin, in *Proceedings* of *International Conference on Affective Computing and Intelligent Interaction*, edited by J. Tao, T, Tan and R. W. Picard (Springer, Berlin Heidelberg), pp. 433–440.

Wassink, A. B., Wright, R. A., & Franklin, A. D. (2007). Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics*, 35, 363–379.

Xu Rattanasone, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog-and infant-directed speech. *Anthrozoös*, 26, 373-380.

Xu Rattanasone, N., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, 41, 332–343.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustic Society of America*, 95, 2240–2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61–83.

Yip, M. (2002). *Tone* (Cambridge University Press, Cambridge), pp. 1.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37, 231–247.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical band rate and critical bandwidths as a function of frequency. *The Journal of the Acoustic Society of America*, 68, 1523–1525.

Chapter Three: Acoustic Realization of Mandarin Neutral Tone and Tone Sandhi in Infant-directed speech and Lombard speech

This chapter is based on the following published paper:

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(5), 2823-2835. https://doi.org/10.1121/1.5008372

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors) when needed.

Abstract

Mandarin lexical tones are modified in both infant-directed speech (IDS) and Lombard speech, resulting in tone hyperarticulation. However, it is unclear if these registers also alter contextual tones (neutral tone and tone sandhi) and if such phonetic modification might affect acquisition of these tones. This study therefore examined how neutral tone and tone sandhi are realized in IDS, and how their acoustic manifestations compare with those in Lombard speech, where the communicative needs of listeners differ. Neutral tone and tone sandhi productions were elicited from 15 Mandarin-speaking mothers during 1) interactions with their 12-month-old infants (IDS), 2) in conversation with a Mandarin-speaking adult in a noisy environment (Lombard speech), and 3) in conversation with a Mandarin-speaking adult in a quiet environment (adult-directed speech (ADS)). The results showed that, although both contextual tones were modified in IDS and Lombard speech, their key tone features were maintained. In addition, IDS and Lombard speech modified these tones differently: IDS increased pitch height and modified pitch contour, while Lombard speech increased pitch height only. The realization of neutral tone and tone sandhi across registers is discussed with reference to listeners' different communicative needs.

Keywords: infant-directed speech; Lombard speech; Mandarin neutral tone; Mandarin tone sandhi.

Introduction

Speakers modify their speech to cater to different communicative needs of the addressee, which result in various speech styles. For example, when talking to an infant, speech is usually hyperarticulated, resulting in a unique speech register known as infant-directed speech or IDS (Kuhl et al., 1997; Burnham, Kitamura & Vollmer-Conna, 2002). Speech hyperarticulation in IDS typically includes a series of suprasegmental and segmental modifications. Relative to adult-directed speech (ADS), IDS exhibits higher pitch (Fernald et al., 1989; Burnham, Kitamura & Vollmer-Conna, 2002; Fernald & Simon, 1984), larger pitch variability (Fernald et al., 1989), slower speaking rate (Fernald & Simon, 1984) and enhanced vowel contrasts (Kuhl et al., 1997; Burnham, Kitamura & Vollmer-Conna, 2002; Liu et al., 2003). It is generally agreed that these modifications may help attract and maintain infants' attention, express positive affect, and possibly facilitate infants' language learning (Cooper et al., 1997; Grieser & Kuhl, 1988; Singh, Morgan & Best, 2002).

Another register exhibiting similar speech characteristics is known as Lombard speech, a speech style that people use when talking in a noisy environment (Lombard, 1911). Relative to speech produced in quiet conditions, Lombard speech typically exhibits higher pitch, greater intensity, longer vowel duration and increased formant frequencies (Summers et al., 1988; Junqua, 1996). It is claimed that Lombard speech is used to maintain self-monitoring ability and transmission of speech signals more effectively in a noisy environment (Zhao & Jurafsky, 2009).

In tonal languages such as Mandarin Chinese, IDS and Lombard speech also modify lexical tones, resulting in tone hyperarticulation (Liu, Tsao & Kuhl, 2007; Tang et al., 2017). Tone hyperarticulation can enhance the acoustic features of lexical tones and thus benefit listeners. However, in connected speech, there are many tone variants, such as tone reduction or tone change. Two well-reported types of contextual variation are neutral tone and tone sandhi (Yip, 2002). Although a large body of work has explored the realization of neutral tone and tone sandhi under normal speech conditions, no study has yet directly examined how these contextual tones are realized in IDS and Lombard speech. It is therefore not known whether these contextual tones will also undergo modification in IDS and Lombard speech. It has been claimed that IDS can benefit young children in lexical tone acquisition by means of tone hyperarticulation (Liu, Tsao & Kuhl, 2007; Xu Rattanasone, Burnham & Reilly, 2013). Indeed, some acquisition evidence has shown that Mandarin-learning children acquire lexical tones quite early, i.e., before age 3 (Li & Thompson, 1977; Zhu & Dodd, 2000). However, relative to lexical tones, contextual tones are reported to be much later acquired by children, i.e., only by the age of 6 (Wang, 2011), but the reason remains unclear. Later acquisition of tone sandhi processes compared to lexical tones is also reported in Bantu languages (e.g., Demuth, 1993), suggesting that learning tone sandhi processes may take longer to master in many tone languages. It is also possible that these contextual tones might not undergo phonetic exaggeration in IDS to the same extent that lexical tones do, making them harder to learn. Thus, the primary goal of the present study was to examine the acoustic realization of these contextual tones in IDS to shed light on the possible reasons for the late acquisition of contextual tones.

Although lexical tones have been found to be hyperarticulated in both IDS and Lombard speech registers (e.g., Tang et al., 2017), the phonetic dimensions of tone hyperarticulation manifest differently across the two registers. For example, although pitch height is increased in both registers, the pitch contour is modified only in IDS

(Liu, Tsao & Kuhl, 2007; Zhao & Jurafsky, 2009; Tang et al., 2017). This subtle difference appears to be driven by the different communicative goals of these two registers. The increased pitch in Lombard speech, for example, might be a by-product of the increased vocal effort needed in order to speak more loudly, as proposed by Uchanski (2005). Similar proposals are suggested by Schulman (1989), who found that vocal effort increased dramatically with loud speech, and, as a consequence of increase of vocal effort, the pitch was also increased. In contrast, acoustic modifications of pitch height and pitch contour in IDS have been associated with attentional, affective and didactic functions (Fernald & Kuhl, 1987; Kitamura & Burnham, 2003; Trainor & Desjardins, 2002). For example, Fernald and Kuhl (1987) found that infants prefer to listen to IDS when it differs from ADS in terms of pitch (both pitch height and pitch range), but not when it differed in amplitude or duration. Similarly, when adults were asked to rate their impression of IDS, Kitamura and Burnham (2003) found that pitch height was associated with affective and attentional scales, but pitch range correlated only with attentional scales. However, Trainor and Desjardins (2002) provided some evidence suggesting that exaggerated pitch contours might help with language learning. In their study, 6-7-month-old infants were reported to better discriminate the vowels /i/ vs /ɪ/ with exaggerated pitch contours compared to monotone speech. However, since monotone prosody is quite unnatural, the observed effect of the exaggerated pitch contour might also be attributed to attracting the infants' attention.

Contextual tones might also undergo pitch modification in both IDS and Lombard speech in order to achieve certain communicative aims. If so, the modifications might be different between the two registers, i.e., pitch height might be increased in both registers, while the pitch range might be expanded in IDS only.

Therefore, the second goal of the present study was to compare directly the role of register on the realization of contextual tones, to better understand the nature of tonal enhancement across these two registers.

Neutral tone

In Mandarin, neutral tone is called the "fifth tone" and often referred to as "toneless" or T0. Neutral tone has short duration, and usually occurs in a final unstressed syllable of a disyllabic word. Depending on the preceding lexical tone, the pitch contour of the neutral tone varies (Cao, 1992). For example, neutral tone exhibits a falling pitch when preceded by syllables containing T1 (tone 1, the level tone), T2 (tone 2, the rising tone) and T4 (tone 4, the falling tone); however, its pitch contour rises when it is preceded by a syllable with T3 (tone 3, the dipping tone) (Yip, 2002). Neutral tones are thus often found in (1) certain monosyllabic particles (i.e., the diminutive particle /tsi0/, the nominalizer or possessive particle /tɤ0/ and the classifier particle /kɤ0/); (2) the second syllable of a disyllabic reduplicated word, such as /ma1 ma0/ *(mother*); or (3) the second syllable of some disyllabic words, such as /t^{h}ou2 fa0/ (*hair*) (Li & Thompson, 1977; Zhu & Dodd, 2000).

It has been reported in previous studies that children do not master neutral tone productions until around 4;6. The most common error that children make is to replace a neutral tone with a full lexical tone, i.e., using /tsi3/ for /tsi0/ (Li & Thompson, 1977; Zhu & Dodd, 2000). According to Zhu and Dodd (2000), these tone substitution errors are often found when using the diminutive particle /tsi0/ and in the final syllable of disyllabic words such as /t^{h}ou2 fa0/ (*hair*).

Tone sandhi

Tone sandhi (also known as tone 3 sandhi) refers to a phonological process in which a lexically-specified tone 3 syllable (T3) is realised as a rising tone (T2: full sandhi) when followed by another T3 syllable or realised as a low-falling tone (half sandhi) when followed by different tones: T1, T2 or T4 (Yip, 2002). The (full) sandhi rule can also be applied recursively to multiple T3 words, and it depends on the prosodic structure of the phrase (Shih, 1997). Consider the underlying forms /ɕiau3 ma3 ji3/ (*small ant*) and /ma3 ji3 tɕiau3/ (*ant's feet*). The former has a right-branching prosodic structure [σ [σσ]] and the latter has a left-branching prosodic structure [[σσ] σ]. These structures trigger rightward parsing [see (1)] and leftward parsing [see (2)] respectively, leading to two different surface realizations of the underlying tone sequence "/ma3 ji3/" → T2T3 vs. T2T2 respectively (Shih, 1997).

(1) A right-branching trisyllabic noun phrase [σ [σσ]]

[ɕiau [ma ji]]	(*Small ant*)
T3 T3 T3	Underlying tone
T3 (T2 T3)	Tone sandhi applies within prosodic domain
(T3 T2 T3)	Incorporation, no additional tone sandhi rule applied; surface tone

(2) A left-branching trisyllabic noun phrase [[σσ] σ]

[[ma ji] tɕiau] (*Ant's feet*)

T3 T3 T3 Underlying tone

(T2 T3) T3 Tone sandhi applies within prosodic domain

(T2 T2 T3) Incorporation, tone sandhi rule applied again; surface tone

The tone sandhi rule is challenging for children to acquire, with children at age 6 still not achieving productive knowledge of adult-like tone sandhi in production (Wang, 2011). It requires children to have knowledge of both the tonal and prosodic context in which the rule applies (Wang, 2011).

As overall prosodic features are modified in IDS and Lombard speech, these modifications, especially the slow speaking rate, might have additional repercussions for the acoustic representation of both neutral tone and tone sandhi as input to children. For example, neutral tone, realized on an unstressed syllable, is mainly characterized by short duration. Thus the need to preserve short duration for neutral tone will go against the need to speak slowly (with increased duration) in both registers. The conflicting demands between speech register and the acoustic characteristics of neutral tone might thus result in production of ambiguous tone. If so, neutral tone might be realized more like a full lexical tone, i.e., /tsi0/ being realized as /tsi3/, similar to the type of error that children tend to make in producing neutral tone syllables (Zhu & Dodd, 2000). A recent study found greater utterance-final lengthening in IDS and Lombard speech relative to ADS (Tang et al., 2017). Since neutral tone syllables mainly occur in the final syllable of a word, IDS and Lombard speech could also potentially lengthen neutral tone productions, especially in the utterance-final position, resulting in tone ambiguity.

Similar to neutral tone, tone sandhi rule application may also be affected by the slow speaking rate of IDS and Lombard speech, especially for some polysyllabic noun phrases. For example, as mentioned above, in normal speaking rate, the surface tones of a T3 right-branching phrase [σ [σσ]] vs. a left-branching phrase [[σσ] σ] would be T3T2T3 and T2T2T3, respectively. However, in slow speaking rate, a prosodic boundary could be inserted between the disyllabic word [σσ] and the monosyllabic word [σ] in both cases. Although an inserted boundary would not affect tone sandhi application of the right-branching phrase [see (3)], it would influence tone sandhi application in the left-branching case, resulting in reduced tone sandhi application [see (4)], and the potential for different meanings. In other words, slow speech in IDS and Lombard speech could undermine the didactic aim of IDS and the clarification goal of Lombard speech. Furthermore, it has recently been found that different speech registers can affect both the pitch height and pitch contour of lexical tones differently (Tang et al., 2017). This raises the possibility that certain phonetic aspects of contextual tones (neutral tone and tone sandhi) might be differentially modified to suit the didactic vs. clarification aims of IDS vs. Lombard speech respectively, resulting different types modifications for these two registers.

(3)	[σ [σσ]]	right-branching phrase
	T3T3T3	Underlying tone
	T3#T3T3	Boundary inserted
	T3#T2T3	Tone sandhi applies; surface tone

(4)	[σσ] σ]	left-branching phrase
	T3T3T3	Underlying tone
	T3T3#T3	Boundary inserted
	T2T3#T3	Tone sandhi applies; surface tone

It is also possible that the resulting speech modifications on the neutral tone and the tone sandhi rule in IDS and Lombard speech may distort their acoustic/linguistic characteristics and thus be a potential hindrance for the listeners. If so, this might lead to insufficient input for children, and therefore help explain the reported later acquisition of these contextually determined tones. It may also decrease communicative effectiveness by making it more difficult for listeners to perceive these tones in noisy environments. These outcomes would then go against the communicative aims of these registers, i.e., the language-teaching aim of IDS and the clarification aim of Lombard speech.

The aim of the present study was therefore to investigate the phonetic modification of Mandarin contextual tones - neutral tone and tone sandhi - in IDS and Lombard speech. In particular, we asked whether IDS and Lombard speech would modify neutral tone and tone sandhi productions, and if so, whether the two speech registers would differ in the acoustic implementation of these changes. To explore these issues, neutral tone and tone sandhi productions were elicited across three conditions: IDS, Lombard speech (while listening to babble noise) and ADS (in quiet, as a control). Neutral tone vs. full tone minimal pairs were employed to investigate the acoustic realization of neutral tone using the full lexical tone as a baseline control; tone sandhi productions were elicited in three different contexts (a disyllabic word, a right-branching disyllabic noun phrase and a left-branching disyllabic noun phrase) to examine the tone sandhi realization across different prosodic contexts. Neutral tone and tone sandhi productions were also compared between registers to explore the register effect on these tones. We predicted that:

Hypothesis 1 (H1): Neutral tone syllables would be modified in terms of duration and pitch and produced as full tone in IDS and Lombard speech.

Hypothesis 2 (H2): Tone sandhi syllables would be modified in terms of duration and pitch and produced with the underlying tone (T3) in IDS and Lombard speech.

Hypothesis 3 (H3): Relative to ADS, IDS would increase the pitch height and modify the pitch contour of neutral tone and tone sandhi syllables, while Lombard speech would increase the pitch height only, as previously observed for lexical tones in IDS (Liu, Tsao & Kuhl, 2007) and Lombard speech (Zhao & Jurafsky, 2009).

Method

Participants

Fifteen mothers and their 12-month-old infants (Mean = 12 months, standard deviation (SD) = 0.99) were recruited from Sydney, Australia. Mothers of 12-month-olds were selected to compare our results to previous IDS studies of (Taiwan and Northern) Mandarin lexical tones, which used similar age groups (Liu, Tsao & Kuhl, 2007; Tang et al., 2017).

All mothers were raised in Northern-Mandarin speaking families (e.g., Beijing, Hebei province and North-eastern China) before 18 yrs of age, had been in Australia from 1 to 8 yrs (Mean = 5 yrs, SD = 3.27), and spoke both Mandarin and

English. Their age ranged from 18 to 34 yrs (Mean = 24 yrs, SD = 4.93). All were the main caregiver of their infants, speaking only Mandarin to their infants at home.

Stimuli

Seven Mandarin nouns were chosen as target stimuli, all illustrated with toys. For the neutral tone conditions, two neutral tone vs. full tone minimal pairs were selected (see Table 1). All were disyllabic words where the second syllable was the target syllable, which carried either the neutral tone or the full tone counterpart. The neutral tone-bearing unit was the Chinese word /t^{h}ou/ (*head*) in pair 1 and a diminutive particle /tsi/ (*piece*) in pair 2. These two types of neutral tone were selected because they have been reported to be acquired last by Mandarin-learning children, compared to the other neutral tone semantic categories (e.g., the nominalizer or possessive particle /tɤ0/ and the classifier particle /kɤ0/). The neutral tone syllables in these two types of words are easily confusable with their full tone counterparts in children's productions, i.e., /ʂɤ2 t^{h}ou0/ (*tongue*) confused with /ʂɤ2 t^{h}ou2/ (*snake's head*) and /tɕ^{h}i2 tsi0/ (*flag*) confused with /tɕ^{h}i2 tsi3/ (*chess piece*) (Li & Thompson, 1977; Zhu & Dodd, 2000). Using minimal pairs thus allowed for better control in comparing neutral tone with full tone productions (as a baseline control).

Table 1. Stimuli for eliciting neutral tone productions, including two neutral vs. full tone minimal pairs.

Pair	Type	Stimuli	Target syllable	Target tone
Pair 1	Neutral tone	/ʂɤ2 t^{h}ou0/ (*tongue*)	/t^{h}ou0/	T0
	Full tone	/ʂɤ2 t^{h}ou2/ (*snake's head*)	/t^{h}ou2/	T2
Pair 2	Neutral tone	/$tɕ^{h}$i2 tsi0/ (*flag*)	/tsi0/	T0
	Full tone	/$tɕ^{h}$i2 tsi3/ (*chess piece*)	/tsi3/	T3

For the tone sandhi comparisons, a disyllabic T3T3 word "/ma3 ji3/" (*ant*) and two trisyllabic T3T3T3 noun phrases were selected (see Table 2). These two noun phrases shared the same syllables "/ma3 ji3/" (*ant*) but differed in their prosodic structures, i.e., right-branching /ɕiau3 **ma3 ji3**/ (*little ant*) and left-branching /**ma3 ji3** tɕiau3/ *(ant's feet*). The syllable /ma3/ and syllable /ji3/ were treated as the target syllables. If the tone sandhi rule is correctly applied, the expected surface tone of /ma3 ji3/ would be T2T3 in the disyllabic word and the right-branching noun phrase, and T2T2 in the left-branching noun phrase (Shih, 1997).

Table 2. Stimuli for eliciting tone sandhi productions, including a lexical disyllabic tone sandhi word, a right-branching trisyllabic noun phrase and a left-branching trisyllabic noun phrase. The underlying tone and the surface tone of the target words are also provided.

Stimuli	Type	Target syllables	Underlying tone	Surface tone
/ma3 ji3/ (*ant*)	Disyllabic	/ma3 ji3/	T3T3	T2T3
/ɕiau3 ma3 ji3/ (*little ant*)	Right-branching	/ma3 ji3/	T3T3	T2T3
/ma3 ji3 tɕiau3/ (*ant's feet*)	Left-branching	/ma3 ji3/	T3T3	T2T2

Procedure

Every mother-infant dyad was tested in a sound-attenuated room. During the familiarization phase, the mother was instructed to wear a head-mounted microphone (AKG-C520) which was connected to a solid-state recorder (Marantz PMD661MKII). The recorder was placed in a shoulder bag to allow for the mother’s free movement in the test phrase. The recording was made at a sampling rate of 44.1 kHz and a 16-bit quantization. In the test phase, three speech production tasks were conducted to elicit the three registers: (1) IDS, (2) Lombard speech, (3) ADS. The order of these tasks was consistent across participants: IDS first, followed by Lombard speech, and then ADS. This order ensured that IDS data were collected before the infant became fussy.

In the IDS task, the mother and her infant were engaged in a play session. The seven target stimuli were labelled on the corresponding toys using Chinese characters, and the toys were randomly allocated to three cloth bags, minimizing noise. These bags were adopted to counterbalance the order of presentation of the toys/test items across participants. The mother was provided with one bag at a time and asked to play with the toys while interacting with her infant as they normally would at home. From the control room, a Mandarin-speaking experimenter (P.T.) kept count of the tokens that the mother produced (at least ten repetitions) for each toy in each bag.

In the Lombard speech task, the mother was asked to converse with the experimenter about her experience in the play session with her infant. The mother was instructed to use the same written labels to refer to the toys. The mother and the experimenter both wore open-ear headphones (AKG-K612 PRO), through which a 70-dBA Chinese 8-talker babble noise was played during the Lombard speech sessions. Open-ear headphones were adopted since they allowed the participants to hear both the babble noise played via headphone and the speech produced by the interlocutor. This was done to ensure that babble noise did not interfere with the sound recordings of the participant's speech. A Digitech QM1591 Decibel Meter was used to calibrate the sound level played via headphones. Before testing, the decibel meter was placed on the headphone to make sure the sound level of the auditory output was at 70-dBA. The conversation continued until the mother produced at least 8 repetitions for each toy (compared to ten repetitions for each toy in IDS, where we expected a higher exclusion rate due to overlap with infant vocalizations). In case the mothers failed to produce the minimal eight repetitions for a toy, the researcher would prompt them to produce more repetitions by asking questions such as "what is the color of X?" or "does your infant like X?," etc.

In the ADS task, the procedure and the minimal number of repetitions were identical to that for the Lombard speech task. The only difference was that the conversation in the ADS task took place in a quiet environment.

Coding and measurements

Mothers' production data were coded by a trained native speaker of Mandarin Chinese (P.T.) in Praat (Boersma & Weenink, 2016), with the aid of spectrograms and waveforms. All the target words were first identified and segmented from the raw speech file for further analysis. Vowel onset and offset of the target words were annotated based on clear F2 regardless of utterance position, though position information was also labelled (see the section of "Neutral tone" below). Ten percent of the tokens were recoded by another trained native Mandarin Chinese speaker for a reliability check to ensure that the annotated vowel interval is consistent across different coders. This is important because the duration and pitch information were extracted automatically within this annotated interval. Correlations were conducted between coders for the annotated vowel duration, resulting in a Pearson's correlation of 0.90 for neutral tone words and 0.89 for tone sandhi words.

Neutral tone

Target syllables (the second syllable of each target word) across the two neutral tone vs. full tone minimal pairs (pair 1: /ʂɤ2 **tʰou0**/ vs. /ʂɤ2 **tʰou2**/; /tɕʰi2 **tsi0**/ vs. /tɕʰi2 **tsi3**/) and three registers (IDS, Lombard speech, ADS) were annotated. The utterance position of the target syllable was also annotated. Since the neutral tone and its full tone counterpart were always carried by the second syllable of a target disyllabic word, the target tones appeared either in utterance-medial position (when medial in an utterance) or in utterance-final position (when final in the utterance). The absolute duration (in ms) of both syllables (first and second syllable) of each target word were measured and the normalized duration was computed for the target syllable (second syllable), using the following formula:

(1) Normalized duration = 2nd vowel duration / (1st vowel duration + 2nd vowel duration).

In addition, four fundamental frequency parameters were measured from the vocalic part of the target 2^{nd} syllable: f0 onset, f0 offset, f0 minimum and mean f0. F0 onset and f0 offset were measured as the f0 values (in Hertz) of the 5% and the 95% points from the vocalic portion so that tonal coarticulation and micro prosodic perturbation from neighbouring consonants could be minimised; the f0 minimum and mean f0 were then measured from 5% to 95% of the vocalic portion, respectively (see Figure 1).

F0 points measured by Praat were checked and mis-tracked points were manually revised, to correct for the "doubling" or "halving" errors in pitch tracking. In the analysis, f0 values were transformed to semitones from observed Hertz values with 50 Hz as the reference, using the following formula:

(2) Semitone = 12 * log2 (target Hertz/50).

Extracted f0 onset, f0 offset and f0 minimum values were then used to compute Δ onset (changes in tone onset) and Δ offset (changes in tone onset), using the following formulas (see Figure 1):

(3) Δ Onset = f0 onset – f0 minimum; Δ offset = f0 offset - f0 minimum.

Δ onset and Δ offset are two parameters that can effectively quantify the pitch contour of Mandarin tones (Shen, Lin & Yan, 1993; Moore & Jongman, 1997). Δ

Onset indicates the size of the falling component of the pitch contour, and Δ offset indicates the size of the rising component. For example, according to Yip (2002), neutral tone exhibits a falling pitch when following a T2 syllable (as in /ʂɤ2 **tʰou0**/ and /tɕʰi2 **tsi0**/ of our stimuli), so the f0 minimum was the same as f0 offset of the pitch contour, resulting in a larger Δ onset and a smaller Δ offset. A full set of the description of Δ onset and Δ offset patterns across tones is presented in Table 3.

Among all these parameters, four duration and fundamental frequency parameters were adopted in the later analysis: normalized duration, mean f0, Δ onset and Δ offset. These four parameters can quantify both the duration and the pitch contour of tone production, which are two of the most important dimensions in distinguishing between neutral tone and full tone productions (Li et al., 2014).

Table 3. Pitch contours of lexical tones and neutral tone (following T2) and their corresponding Δ onset and Δ offset patterns: "+" indicates a relative large value and "-" indicates that the value is relatively small or close to zero.

Tone	Pitch contour	Δ onset	Δ offset
T1	level	-	-
T2	rising	-	+
T3	dipping	+	+
T4	falling	+	-
T0 (following T2)	falling	+	-

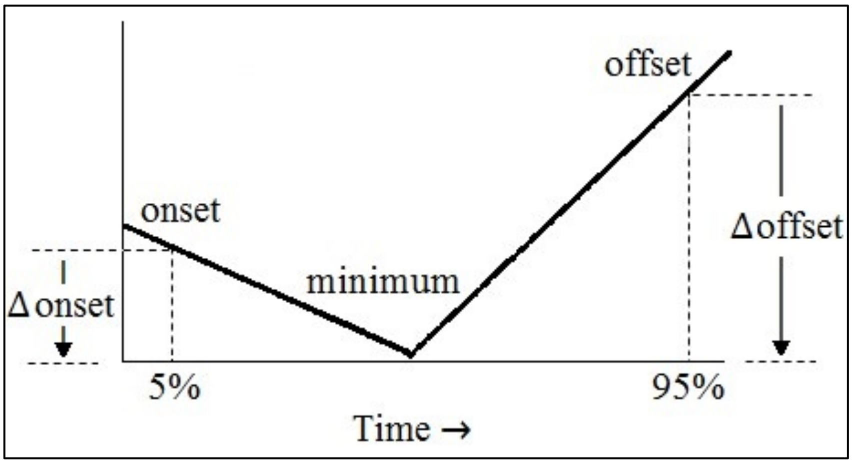


Figure 1. Pitch parameters schematized for a contour tone (T3), including Δ onset and Δ offset. Δ onset and Δ offset were computed as the differences between f0 onset/offset and the f0 minimum. F0 onset and offset were measured as the f0 values (in Hertz) of the 5% and the 95% points from the vocalic portion; f0 minimum was measured as the f0 minimum value from 5% to 95% of the vocalic portion.

Tone sandhi

Target tone sandhi syllables (/ma3 ji3/) across three contexts (disyllabic word, right-branching noun phrase and left-branching noun phrase) and three registers (IDS, Lombard speech, ADS) were annotated. Three fundamental frequency parameters were derived from tone sandhi target syllables for later analysis: f0 mean, Δ onset and Δ offset, using the same method as described above. These parameters can quantify the pitch contour of tone production, which is the most important dimension in distinguishing between the underlying tone (T3, with a dipping pitch) and the surface tone (T2, with a rising pitch) of tone sandhi syllables (Moore & Jongman, 1997).

Statistical analysis

A total of 3285 tokens were included in the analysis, including 1821 neutral tone vs. full tone productions and 1464 tone sandhi productions. The 1821 neutral tone vs. full tone productions included 926 neutral tone syllables and 895 full tone counterparts. The 1464 tone sandhi productions included 565 trisyllabic tone sandhi words, 442 right-branching noun phrases and 457 left-branching noun phrases. An additional 254 neutral tone productions and 111 tone sandhi productions were excluded from the analysis for the following reasons: overlap with another sound, such as the infant's vocalization or noise made by toys or other environmental disturbance; the mother laughing or singing when producing the token; the token produced in whisper; mispronunciation.

The data were analyzed using R (R Core Team, 2016). A linear mixed-effects model was performed to compare the acoustic characteristics across tone categories and across registers, using the "lme4" package (Bates et al., 2015) and the "lmerTest" package (Kuznetsova, Brockhoff & Christensen, 2013). The *post hoc* test was performed on these models to conduct pairwise comparisons, using the “lsmeans” package (Lenth, 2016).

Results

Neutral tone

Figure 2 and figure 3 summarize the pitch contours and the acoustic characteristics (normalized duration, mean f0, Δ onset and Δ offset) of the neutral tone vs. full tone syllables across the two utterance positions (medial and final), the three registers (IDS, Lombard speech, ADS) and the two minimal pairs (pair 1: T0 vs. T2; pair 2: T0 vs. T3). Means and SDs for these parameters are also provided (see Appendix 1).

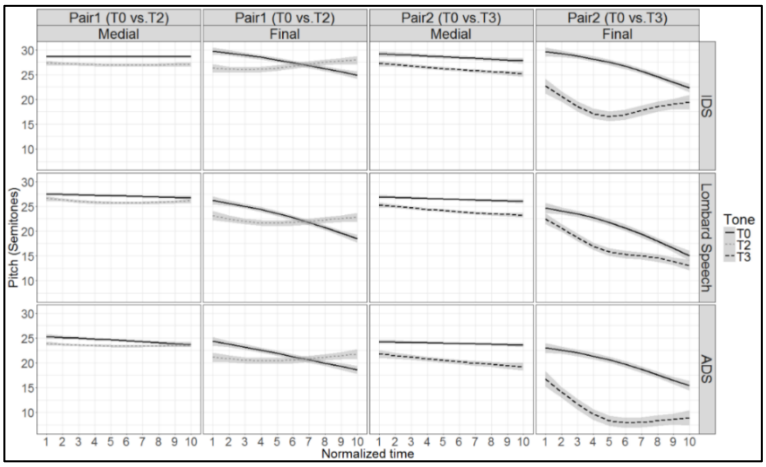


Figure 2. Pitch contours of neutral tone and full tone syllables across three registers (IDS, Lombard speech and ADS) and two utterance positions (medial and final) in two neutral tone vs. full tone minimal pairs (pair 1: T0 vs. T2; pair 2: T0 vs. T3).

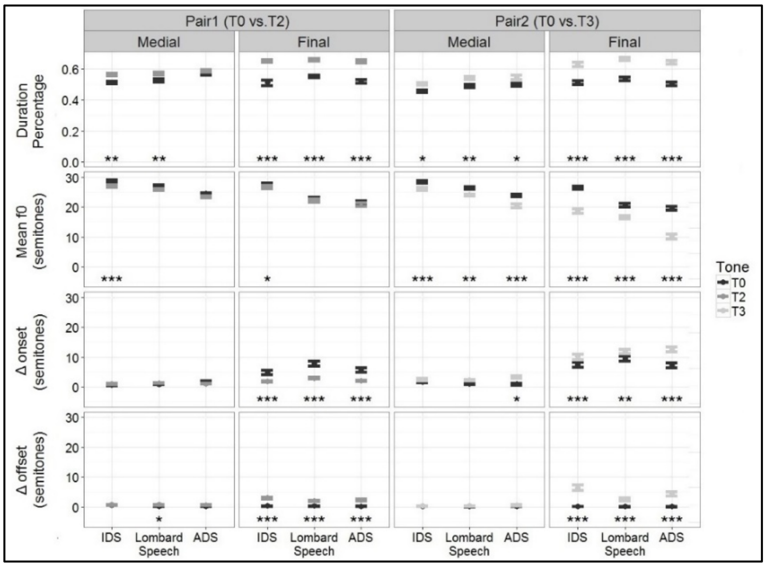


Figure 3. Acoustic characteristics of neutral tone and full tone syllables across three registers (IDS, Lombard speech and ADS) and two utterance positions (medial and final) in two neutral tone vs. full tone minimal pairs (pair 1: T0 vs. T2; pair 2: T0 vs. T3). Four acoustic parameters are included: (1) normalized duration, (2) mean f0, (3) Δ onset and (4) Δ offset. Results of significant pairwise comparisons between neutral tone and full tone syllables are illustrated by asterisks, and the number of which indicates the level of statistical significance between neutral tone and full tone syllables: $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***.

To test H1 regarding the acoustic realization of neutral tone syllables, these acoustic parameters were compared between neutral tone and full tone productions across registers in the two minimal pairs. Two separate linear mixed-effect models for pair 1 and pair 2 were performed on these parameters with two fixed factors: Tone (neutral tone and full tone), Position (medial and final) and Register (IDS, Lombard speech and ADS). A random factor was also included: Subject (15 subjects). To keep the model optimal for generalizing the data, random slopes of Subject for main effects of all fixed factors were included (Barr et al., 2013), i.e., random slopes of Subject on Tone, Position and Register[8]. The results are summarized in Appendix 2.

Our results showed that, for both pairs, the main effect of Tone was significant for all the four acoustic parameters, the two-way interaction of Tone × Register was not significant for most parameters except for mean f0, Δ onset and Δ offset of syllables in pair 2, and the three-way interaction of Tone × Position × Register was not significant for most parameters except for mean f0 of syllables in pair 2 and Δ offset of syllables in both pairs. Pairwise comparisons were adopted to further compare the neutral tone and full tone syllables across positions and registers (see Figures 2 and 3). The results showed that, for IDS, neutral tone and full tone syllables exhibited significant duration (normalized duration) and pitch height (mean pitch) differences in both medial and final positions, and they exhibited significant pitch contour (Δ onset and Δ offset) differences in the final position; for Lombard speech, neutral tone and full tone syllables showed significant duration (normalized duration) differences in both medial and final positions, and they exhibited significant pitch contour differences in the final position.

We also ran a separate analysis to compare the absolute raw duration (in milliseconds) of both neutral tone and full tone syllables across registers. The results showed that, relative to ADS, IDS exhibited longer absolute duration (in ms) for both the neutral tone syllable ($\beta = 29.99$, standard error (SE) = 9.64, $t = 3.11$, $p < 0.05$) and its preceding full tone syllable ($\beta = 33.82$, $SE = 9.64$, $t = 3.51$, $p < 0.01$). Lombard speech, in contrast, did not show this pattern. Relative to ADS, neither the neutral

---

[8] The R code of this model is: Normalized duration/ Mean f0/ Δ Onset/ Δ Offset ~ Tone * Position * Register + (1 + Tone + Position + Register | Subject).

tone syllable nor its preceding full tone syllable (neutral tone: $\beta = 17.93$, $SE = 7.86$, $t = 2.28$, $p = 0.07$; full tone: $\beta = 11.48$, $SE = 7.85$, $t = 1.46$, $p = 0.32$) was lengthened. This result indicated that, in IDS, speakers lengthened both neutral tone and full tone syllables while maintaining their durational difference; in Lombard speech, in contrast, speakers did not change duration of either neutral tone syllables or full tone syllables, and the durational difference between neutral tone and full tone syllables was maintained as well.

These results did not support H1, which predicted that, relative to ADS, neutral tone syllables would be acoustically realized as full tone syllables (i.e., their full tone counterparts) in IDS and Lombard speech. Rather, our results demonstrated that, in both registers, neutral tone and the full tone counterparts were distinct in terms of normalized duration, pitch height and pitch contour, and this distinction was exaggerated in the final position.

To test H3 regarding the registers' difference in modifying neutral tone syllables, we performed linear mixed-effects models on the acoustic parameters of the two neutral tone syllables only (t^hou0/ and /tsi0/). Three fixed factors Syllable (/t^hou0/ and /tsi0/), Position (medial and final) and Register (IDS, Lombard speech and ADS) and the random factor Subject (15 subjects) were included. Random slopes of Subject for main effects of all fixed factors were included and were kept in the model as well[9]. The results are summarized in Appendix 3.

The results showed that the main effect of Register was significant for mean f0 and Δ offset, and the interaction of Position × Register was significant for these

[9] The R code of the models is: Normalized duration/ Mean f0/ Δ Onset/ Δ Offset ~ Syllable * Position * Register + (1 + Syllable + Position + Register | Subject).

parameters as well. This interaction indicates neutral tone syllables were modified differently across registers in terms of mean f0 and Δ offset. Pairwise comparisons on register differences in the interaction of Position × Register indicated that: (1) IDS exhibited higher mean f0 for neutral tone syllables than Lombard speech and ADS in both positions, while Lombard speech exhibited higher mean f0 than ADS only in the medial position ($\beta = 2.5$, $SE = 0.58$, $t = 4.34$, $p < 0.001$) rather than the final position ($\beta = 1.07$, $SE = 0.58$, $t = 1.86$, $p = 0.16$); (2) IDS exhibited a larger Δ offset of neutral tone syllables in the medial position than Lombard speech ($\beta = 0.27$, $SE = 0.08$, $t = 3.37$, $p < 0.01$) and ADS ($\beta = 0.28$, $SE = 0.08$, $t = 3.35$, $p < 0.01$). These results supported H3, which predicted that IDS will modify both pitch height and pitch contour of neutral tone syllables, while Lombard speech will modify the pitch height only.

Tone sandhi

Figure 4 and figure 5 summarize the pitch contours and the acoustic characteristics (mean f0, Δ onset and Δ offset) of tone sandhi syllables /ma/ and /ji/ across contexts (disyllabic word, right-branching noun phrase and left-branching noun phrase) and registers (IDS, Lombard speech, ADS). Means and SDs for these parameters are provided (see Appendix 4).

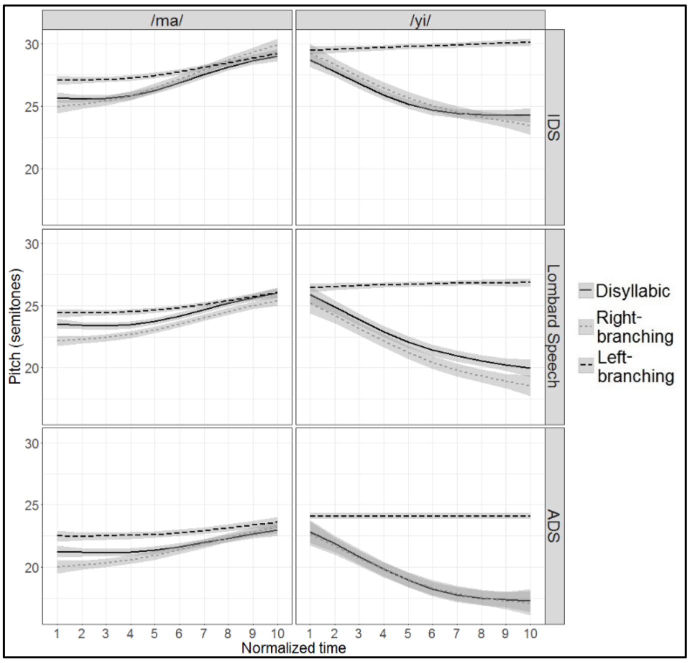


Figure 4. Pitch contours of tone sandhi syllables /ma/ and /ji/ across three contexts (disyllabic word, right-branching noun phrase and left-branching noun phrase) and three registers (IDS, Lombard speech and ADS).

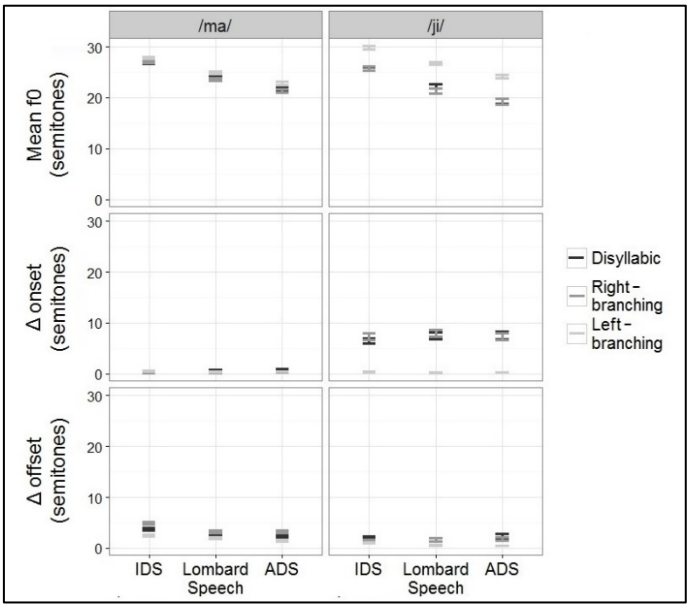


Figure 5. Acoustic characteristics of tone sandhi syllables /ma/ and /ji/ across three contexts (disyllabic word, right-branching noun phrase and left-branching noun phrase) and three registers (IDS, Lombard speech and ADS). Three acoustic parameters are included: (1) mean f0, (2) Δ onset and (3) Δ offset.

To test H2 regarding the acoustic realization of the tone sandhi syllables, the acoustic parameters of /ma/ and /ji/ were compared across contexts and registers. Two separate linear mixed-effect models for /ma/ and /ji/ were performed on these parameters with two fixed factors: Context (disyllabic word, right-branching noun phrase and left-branching noun phrase) and Register (IDS, Lombard speech and ADS). The random factor Subject was also included: (15 subjects). All random slopes were included in the model to keep the model optimal generalizing the data[10]. The results are summarized in Appendix 5.

The results showed that, for both /ma/ and /ji/, the main effect of Context was significant for all parameters, and the interaction of Context × Register was not significant. These results indicate that, although the pitch height and pitch contour of tone sandhi syllables varied across contexts, this variation did not change as a function of register.

A Tukey HSD *post hoc* test was then performed on the main effect of Context to determine if tone sandhi rule was applied for tone sandhi syllables in correct contexts (/ma/ for all three contexts and /ji/ for the left-branching noun phrase). The results are presented in Appendix 6, which shows that, (1) although Δ onset and Δ offset values of the syllable /ma/ differed across contexts, the syllable /ma/ exhibited a rising pitch contour in all contexts, as illustrated in figure 4. It indicates that speakers applied the tone sandhi rule on this syllable and produced it with the surface tone (T2); (2) the syllable /ji/ exhibited a higher mean f0 and a smaller Δ onset and a smaller Δ offset for the left-branching noun phrase than other contexts, which

[10] The R code of this model is: Mean f0/ Δ Onset/ Δ Offset ~ Context * Register + (1 + Context + Register | Subject).

indicates that it exhibited a high-level tone in the left-branching noun phrase and a low dipping tone in other contexts (also see Figure 4). It indicates that speakers applied the tone sandhi rule on the syllable /ji/ according to the prosodic structure, i.e., tone sandhi rule applied for /ji/ for the left-branching noun phrase, whereas the high-level pitch contour for /ji/ in this case was a result of coarticulation effect that changed the surface tone (T2) of /ji/ to a high level tone (T1)[11].

These results did not support H2, which predicted that, relative to ADS, tone sandhi syllables would be modified and realized with the underlying T3 in IDS and Lombard speech. Thus, our results show that, although the acoustic realization of tone sandhi syllables varied across contexts, this variation did not change as a function of register, and speakers correctly applied tone sandhi across registers.

We also observed a significant main effect of Register on (1) mean f0 for syllable /ma/ and /ji/ and (2) Δ offset for syllable /ma/. To test H3 regarding differences in the modification of tone sandhi syllables /ma/ and /ji/ across registers, we performed a *post hoc* test on these effects. The results showed that (1) for both /ma/ and /ji/, IDS exhibited higher mean f0 than Lombard speech (/ma/: $\beta = 3.05$, $SE = 0.52$, $t = 5.82$, $p < 0.001$; /ji/: $\beta = 3.62$, $SE = 0.3$, $t = 12.27$, $p < 0.001$) and ADS (/ma/: $\beta = 5.4$, $SE = 0.57$, $t = 9.46$, $p < 0.001$; /ji/: $\beta = 6.26$, $SE = 0.3$, $t = 20.63$, $p < 0.001$), and Lombard speech exhibited higher mean f0 than ADS (/ma/: $\beta = 2.35$, $SE = 0.3$, $t = 7.72$, $p < 0.001$; /ji/: $\beta = 2.64$, $SE = 0.33$, $t = 7.92$, $p < 0.001$); (2) for syllable /ma/, IDS exhibited larger Δ offset than ADS ($\beta = 1.71$, $SE = 0.36$, $t = 4.75$, $p <$

[11] According to Xu (1994), a T2 syllable will be realized as a high-level tone (T1) when it is preceded by a syllable with a high pitch offset (i.e., T2 or T4) and followed by another syllable. This is the case for the left-branching noun phrase (the surface tone is T2T2T3) in the present study, and this coarticulation effect leads to the surface tone T2T2T3 of left-branching noun phrase being realized as T2T1T3.

0.001) and Lombard speech ($\beta = 1.76$, $SE = 0.35$, $t = 5$, $p < 0.001$). These results indicate that both IDS and Lombard speech increased the pitch height of tone sandhi syllables, and IDS exaggerated the rising component (Δ offset) of tone sandhi syllable /ma/, which had a rising pitch. These results supported H3, which predicted that IDS would modify both pitch height and pitch contour of the tone sandhi syllables, while Lombard speech would modify pitch height only.

Discussion

This study compared the acoustic characteristics of Mandarin neutral tone and tone sandhi productions in IDS, Lombard speech and ADS. The results showed that, relative to ADS (in quiet, as a control), neutral tone and tone sandhi syllables were modified in IDS and Lombard speech to some extent, but the key tonal features of these contextual tones were maintained in both registers, including the short duration of the neutral tone and the rising pitch contour of the tone sandhi syllable. Additional register differences were found in the extent to which these contextual tones are modified. Specifically, IDS increased the pitch height and exaggerated the pitch contour of neutral tone (in utterance-medial position) and tone sandhi syllables, while Lombard speech increased the pitch height only.

The production of neutral tone and tone sandhi syllables in IDS and Lombard speech did not support our hypotheses (H1 and H2), which predicted that the slow speaking rate in the two registers might distort the two contextual tones to the extent that they would be realized as full tones. Thus, our results demonstrated that, in all registers, neutral tone is acoustically different from the full tone counterpart (controls) in tonal duration, pitch height (mean f0), and pitch contour (Δ onset and Δ offset). The tone sandhi rule was also correctly applied across registers. These results extend previous studies on the realization of Mandarin lexical tones in IDS (Liu, Tsao & Kuhl, 2007) and Lombard speech (Tang et al., 2017), which observed that the register modifies the prosodic features of tones, but not at the expense of tonal contrast.

The durational difference between neutral tone and full tone syllables was maintained across registers, as reflected by the similar normalized duration (the duration proportion of neutral syllable in a disyllabic word). We hypothesized that, in both IDS and Lombard speech, the slow speaking rate might lengthen the neutral tone syllables, especially in the utterance-final position, and therefore distort the durational distinction between neutral tone and full tone syllables. However, our results showed that speakers maintain this durational distinction across positions and registers, though the strategies adopted were different: in Lombard speech, the raw/absolute duration of neutral tone syllables and full tone syllables were unchanged, and therefore the durational distinction was maintained; in IDS, in contrast, the raw/absolute duration of both neutral tone and full tone syllables was lengthened, but the durational difference was preserved. This (raw/absolute) durational difference between IDS and Lombard speech might be associated with different modifications on pitch contour across registers. In other words, it is possible that the exaggerated duration of IDS might be related to the need to implement its exaggerated pitch contour, i.e., the exaggerated contour in IDS might need more time to be implemented. This assumption is supported by a significant correlation found between the raw/absolute duration and pitch range (f0 maximum- f0 minimum) of the neutral tone and full tone syllables across registers (IDS: $r = 0.49$, n = 1520, $p < 0.001$; Lombard speech: $r = 0.44$, n = 1146, $p < 0.001$; ADS: $r = 0.38$, n = 1030, $p < 0.001$).

Similarly, the slow speaking rate in IDS and Lombard speech did not alter the realization of tone sandhi syllables. This result is not consistent with Speer, Shih and

Slowiaczek (1989), who claimed that the size (number of syllables) of the tone sandhi domain (the domain where the tone sandhi rule is applied) is one of the most important factors in determining how tone sandhi is realized. According to this view, the slow speaking rate is likely to induce boundaries between words and therefore to break up the prosodic domain in which the tone sandhi rule applies. However, our results indicate that speaking rate does not seem to influence the tone sandhi domain, or at least that the slow speaking rate of IDS and Lombard speech does not necessarily affect the size of tone sandhi domain. This interpretation is in line with the findings in Kuo, Xu and Yip (2007). They compared tone sandhi productions in slow, normal, and fast speech rate, and found that the size of the prosodic domain is larger in the slow speech rate than other speech rates. However, since the length of tone sandhi words in the present study were relatively short, i.e., either disyllabic or trisyllabic, it is possible that the effect of the speaking rate is more likely to influence the size of the tone sandhi domain at the sentence level. This is an issue that could be addressed in future studies.

Regarding the reason for children's late acquisition of neutral tone and tone sandhi, it seems unlikely that this is related to the exposure to acoustically distorted exemplars for these contextual tones in the language input learners hear. An alternative reason for their late acquisition may relate to the frequency of these contextual tones in the input. It has been shown that phoneme frequency in children's ambient language plays an important role in phonological development. For example, de Boysson-Bardies and Vihman (1991) conducted a cross-language study on French, English, Swedish and Japanese infants' babbling and found a clear correlation between infants' babbling patterns and the distribution of consonantal place and manner categories of infants' ambient languages. Ingram (1988) also found that, relative to English-learning infants, word-initial /v/ is acquired much earlier by Swedish-, Estonian- and Bulgarian-learning children, as it plays a more prominent role in the lexicon of these languages. This evidence suggests a close relationship between language acquisition and phoneme frequency in the infant's ambient language. Therefore, it is possible that neutral tone and tone sandhi words do not occur very frequently in children's language input, resulting in the later acquisition of these tones. Some evidence suggests that the frequency of tone sandhi syllables in IDS is below 5%, and most of these syllables are disyllabic lexicalized items where productive application of sandhi rule might not be applied (Wang, 2011). Therefore, children may not receive enough input of the right type to allow them to learn the tone sandhi rule early. However, little is known about the frequency of neutral tone in children's language input and the relationship between tone input and children's tone acquisition: this needs to be further explored.

Another potential reason for the late acquisition of these contextual tones could be attributed to the challenge in learning the prosodic contexts that trigger neutral tone and tone sandhi. For example, the pitch contours of neutral tone syllables differ depending on the preceding lexical tones (Cao, 1992). Tone sandhi involves syllables being realized as a rising tone (full sandhi), a low falling tone (half sandhi) or a dipping tone (T3) in different tonal and prosodic contexts (Yip, 2002). These variations might be challenging for young learners, and require sufficient exposure to these different types of contexts for learning to take place.

With respect to the hypothesis that there would be register differences in the realization of neutral tone and tone sandhi (H3), neutral tone and tone sandhi syllables in IDS showed an overall raised pitch and an exaggerated pitch contour as compared with ADS. In contrast, neutral tone and tone sandhi syllables in Lombard speech exhibited an overall higher pitch only relative to ADS. These results are consistent with previous findings on lexical tones in IDS (Kitamura et al., 2002; Liu, Tsao & Kuhl, 2007; Xu Rattanasone, Burnham & Reilly, 2013) and Lombard speech (Zhao & Jurafsky, 2009; Boontham et al., 2016).

The increased pitch height and exaggerated pitch contour of IDS observed in the present study are consistent with previous IDS studies which found similar modifications for lexical tones (i.e., Liu, Tsao & Kuhl, 2007; Tang et al., 2017). It has been claimed that IDS mainly serves three functions: attracting and maintaining infants' attention, communicating positive emotion or affect between a caregiver and an infant, and (possibly) facilitating language acquisition (Song et al., 2010). It is generally agreed that the increased pitch height in IDS is mainly associated with attentional and affective functions (Fernald & Kuhl, 1987; Kitamura & Burnham, 2003), and the exaggerated pitch contour in IDS is likely to be related to attentional and didactic functions (Trainor & Desjardins, 2002; Kitamura & Burnham, 2003; Uther, Knoll& Burnham, 2007). This is in line with our observed effect of expanded pitch contour in IDS, but not in Lombard speech. It follows that this phonetic dimension of pitch was manipulated to cater to the needs of the addressee. That is, the pitch contour was exaggerated to facilitate tone learning by drawing the learners' attention through the overall pitch increase to the pitch movement of the contextual tones in IDS. The increased pitch height in Lombard speech, on the contrary, is consistent with the suggestion of Uchanski (2005) that it might be merely a by-product of an increased vocal effort. Evidence from Liénard and Di Benedetto (1999) also indicates that pitch (f0) increases with vocal effort as defined in terms of the physical distance between two persons. A similar argument has also been put forward in Uther et al. (2007) who compared the acoustic parameters across IDS, foreignerdirected speech (FDS, to adults), and ADS. The authors found that, on the one hand, vowel space was expanded in both IDS and FDS but not ADS, reflecting the didactic purposes of both registers; on the other, pitch was higher in IDS than FDS and ADS, and IDS was rated the highest with a positive effect. Since pitch increase can be associated with both attentional/ affective (in IDS) and vocal effort (in Lombard speech) in the present study, it would be interesting to compare a Lombard version of IDS to Lombard speech in future studies to examine the effects of different communicative aims on pitch increase in the two registers. It is also possible that the pitch increase in the current study might be related to the involvement of the experimenter in eliciting Lombard speech. Perhaps the speaker might unconsciously increase her pitch in response to the experimenter. Therefore, future studies could do better by using a coopted confederate to eliminate the potential influence from the experimenter.

Conclusion

Mandarin lexical tones are reported to be hyperarticulated in IDS and Lombard speech to benefit listeners. However, it has been unclear how contextual tones such as neutral tone and tone sandhi are realized in these registers, which might influence children's acquisition of these tones. The results of this study show that, although these contextual tones undergo certain modifications in both IDS and Lombard speech, these modifications do not distort their key tonal features, and they are still well-realized in these registers. These findings shed light on how Mandarin contextual tones are realized in the early language input to infants, and provide further insight into why these contextual tones might be later acquired. Registers differ in their modifications of these contextual tones between IDS and Lombard speech, but these appear to primarily reflect differences in addressees and communicative situations.

Acknowledgements

We thank Titia Benders, Carmen Kung, the Child Language Lab, the Phonetics Lab, and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and supports. We thank Xin Cheng for helping with the reliability check. We also acknowledge the comments and suggestions from the editor and anonymous reviewers which significantly improved and strengthened this manuscript. This research was supported, in part, by a Macquarie University iMQRES scholarship to P.T., and the following grants: Grant Nos. ARC CE110001021 and ARC FL130100014 (K.D.). The equipment was funded by MQSIS 9201501719.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package, version 1.1-9, https://CRAN.Rproject.org/package=lme4.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from http://www.praat.org/.

Boontham, C., Onsuwan, C., Saimai, T., & Tantibundhit, C. (2016). An Analysis of Lombard Effect on Thai Lexical Tones: The Role of Communicative Aspect, in *Proceedings of the 16th Speech Science and Technology Conference*, (Sydney, Australia), pp. 149-152.

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). "What's new pussycat? On talking to babies and animals," *Science*, 296, 1435.

Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics*, 20(3), 49-50.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, 20(4), 477-488. Demuth, K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language*, 20, 275-301.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297-319.

Demuth, K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language*, 20, 275–301.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10, 279-293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1), 104.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477-501.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, 24, 14–20.

Ingram, D. (1988). The acquisition of word-initial [v]. *Language and Speech*, 31, 77-85.

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, 20, 13-22.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, 24, 372-392.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4, 85-110.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684-686.

Kuo, Y., Xu, Y, & Yip, M. (2007). The phonetics and phonology of apparent cases of iterative tone change in Standard Chinese, In *Experimental Studies in Word and Sentence Prosody*, edited bt Gussenhoven, C. and Riad, T. (Mouton de Gruyter, Berlin, Germany), pp. 211-237.

Kuznetsova, A. (2013). Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package), R Package, version 2.0-29, pp. 1-17. https://cran.r-project.org/web/packages/lmerTest/lmerTest. pdf (Last viewed 5/4/2016).

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistic Software*, 69, 1-33.

Li, A., Gao, J., Jia, Y., & Wang, Y. (2014). Pitch and duration as cues in perception of neutral tone under different Contexts in Standard Chinese, In *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Angkor Wat, Cambodia, pp. 1-6.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199.

Liénard, J. S., & Di Benedetto, M. G. (1999). Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America*, 106(1), 411-422.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6(3), F1-F10.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43(4), 912.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Annals Maladiers Oreille, Larynx, Nez, Pharynx*, 37, 101-119.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, 102, 1864-1877.

R Core Team (2016). R: A Language and Environment for Statistical Computing [Computer program]. Version 3.3.1, retrieved 21 June 2016 from https://www.R-project.org/.

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America*, 85, 295-312.

Shen, X., Lin, M., & Yan, J. (1993). F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3. *The Journal of the Acoustical Society of America*, 93, 2241-2243.

Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, 20, 81-124.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk?. *Infancy*, 3, 365–394.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, 128(1), 389-400.

Speer, S. R., Shih, C. L., & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech*, 32(4), 337-354.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84(3), 917-928.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142, 493-503.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, 9(2), 335-340.

Uchanski, R. M. (2005). Clear speech, in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Publishers, Malden, MA), p. 207-235.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication*, 49, 2-7.

Wang, C. Y. (2011). *Children's Acquisition of Tone 3 Sandhi in Mandarin*, Ph.D. dissertation, Michigan State University, East Lansing, MI, pp. 1-336.

Xu Rattanasone, N., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, 41, 332-343.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95, 2240-2253.

Yip, M. (2002). *Tone* (Cambridge University Press, Cambridge), pp. 181.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37, 231-247.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, 27(1), 3-42.

Appendix

Appendix 1. Means and standard deviations (in bracket) for normalized duration, mean f0, Δ onset and Δ offset values of neutral tone syllables (T0) and full tone counterparts (T2 and T3) across registers (IDS, Lombard speech, ADS) and utterance positions (medial and final). The units are percentage for normalized duration and semitone for mean f0, Δ onset and Δ offset.

Pair	Register	Tone	Utterance positions	Parameters			
				Normalized Duration	Mean f0	Δ onset	Δ offset
Pair1	IDS	T0	Medial	0.51 (0.08)	28.60 (4.38)	0.58 (1.12)	0.65 (1.03)
			Final	0.51 (0.17)	27.54 (5.05)	4.86 (6.79)	0.35 (0.65)
		T2	Medial	0.57 (0.07)	27.03 (3.15)	1.12 (1.67)	0.79 (0.99)
			Final	0.65 (0.09)	26.76 (4.89)	1.94 (2.18)	3.03 (3.43)
	Lombard	T0	Medial	0.53 (0.09)	27.10 (3.00)	0.93 (1.30)	0.24 (0.42)
			Final	0.55 (0.08)	22.79 (3.90)	7.85 (6.42)	0.33 (1.34)
		T2	Medial	0.57 (0.08)	25.95 (2.41)	1.36 (1.11)	0.78 (0.86)
			Final	0.66 (0.06)	22.19 (4.45)	2.99 (2.92)	2.07 (1.82)
	ADS	T0	Medial	0.57 (0.06)	24.52 (2.82)	1.61 (3.08)	0.15 (0.35)
			Final	0.52 (0.09)	21.52 (4.71)	5.69 (6.00)	0.29 (0.77)
		T2	Medial	0.59 (0.08)	23.49 (2.25)	1.09 (1.01)	0.65 (0.96)
			Final	0.65 (0.07)	20.87 (5.11)	2.05 (1.83)	2.37 (2.33)
Pair2	IDS	T2	Medial	0.46 (0.09)	28.51 (4.67)	1.73 (3.10)	0.26 (0.60)

		Final	0.51 (0.13)	26.59 (5.55)	7.46 (8.45)	0.22 (0.78)
	T0	Medial	0.50 (0.09)	26.15 (4.33)	2.47 (3.40)	0.32 (1.14)
		Final	0.63 (0.13)	18.72 (6.89)	10.03 (7.92)	6.44 (8.35)
Lombard	T2	Medial	0.49 (0.08)	26.44 (2.90)	1.01 (1.72)	0.12 (0.27)
		Final	0.54 (0.11)	20.61 (5.03)	9.47 (6.75)	0.14 (0.51)
	T0	Medial	0.54 (0.09)	24.10 (2.93)	2.39 (1.78)	0.37 (0.95)
		Final	0.67 (0.08)	16.63 (4.07)	11.91 (6.61)	2.58 (4.29)
ADS	T2	Medial	0.50 (0.09)	23.93 (2.89)	0.98 (2.58)	0.18 (0.59)
		Final	0.50 (0.10)	19.72 (5.03)	7.32 (6.57)	0.15 (0.63)
	T0	Medial	0.55 (0.11)	20.40 (4.90)	3.31 (3.69)	0.59 (1.64)
		Final	0.65 (0.09)	10.19 (5.99)	12.67 (6.46)	4.48 (5.24)

Appendix 2. Results of linear mixed-effects models on normalized duration, mean f0, Δ onset and Δ offset in two neutral tone vs. full tone minimal pairs (pair 1: T0 vs. T2; pair 2: T0 vs. T3). Three fixed factors were included in the model: Tone (neutral tone and full tone), Position (medial and final) and Register (IDS, Lombard speech and ADS). Asterisks indicate the level of statistical significance: $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***. The units are percentage for normalized duration and semitone for mean f0, Δ onset and Δ offset.

Pair	**Factors**	**DF**	**Acoustic parameters**							
			Normalized Duration		**Mean f0**		**Δ onset**		**Δ offset**	
			F	***p***	**F**	***p***	**F**	***p***	**F**	***p***
Pair 1	Tone	1	**71.35**	**<0.001*** **	**13.31**	**<0.01** **	**40.51**	**<0.001*** **	**54.27**	**<0.001*** **
	Position	1	**11.15**	**<0.01** **	**39.32**	**<0.001*** **	**49.98**	**<0.001*** **	**18.25**	**<0.001*** **
	Register	2	**5.15**	**<0.05* **	**41.25**	**<0.001*** **	**8.97**	**<0.001*** **	**3.81**	**<0.05* **
	Tone × Position	1	**57.9**	**<0.001*** **	1.51	0.22	**73.83**	**<0.001*** **	**84.76**	**<0.001*** **
	Tone × Register	2	0.71	0.49	1.17	0.31	1.51	0.22	1.09	0.34
	Position × Register	2	**5.41**	**<0.01** **	**19.93**	**<0.001*** **	**5.99**	**<0.01** **	1.88	0.15
	Tone × Position × Register	2	1.19	0.31	0.22	0.8	1.71	0.18	**6.47**	**<0.01** **
Pair 2	Tone	1	**50.98**	**<0.001*** **	**112.09**	**<0.001*** **	**33.1**	**<0.001*** **	**43.03**	**<0.001*** **

Position	1	**36.9**	**<0.001*** **	**324.63**	**<0.001*** **	**112.77**	**<0.001*** **	**30.55**	**<0.001*** **
Register	2	**5.25**	**<0.05***	**73.34**	**<0.001*** **	1.59	0.2	**9.28**	**<0.001*** **
Tone × Position	1	**47.85**	**<0.001*** **	**46.78**	**<0.01** **	**9.93**	**<0.01** **	**101.24**	**<0.001*** **
Tone × Register	2	0.28	0.75	**10.9**	**<0.001*** **	**3.57**	**<0.05***	**7.62**	**<0.001*** **
Position × Register	2	**3.84**	**<0.01** **	**10.38**	**<0.001*** **	**3.86**	**<0.05***	**8.72**	**<0.001*** **
Tone × Position × Register	2	0.3	0.74	**4.71**	**<0.01** **	0.74	0.48	**8.79**	**<0.001*** **

Appendix 3. Results of linear mixed-effects models on normalized duration, mean f0, Δ onset and Δ offset for two neutral tone syllables (/t^hou0/ and /tsi0/). Three fixed factors were included in the model: Syllable (/t^hou0/ and /tsi0/), Position (medial and final) and Register (IDS, Lombard speech and ADS). Asterisks indicate the statistical significance: $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***. The units are percentage for normalized duration and semitone for mean f0, Δ onset and Δ offset.

Factors	DF	Acoustic parameters							
		Normalized Duration		Mean f0		Δ onset		Δ offset	
		F	*p*	F	*p*	F	*p*	F	*p*
Syllable	1	**8.95**	**<0.01****	**7.47**	**<0.05***	**7.53**	**<0.01****	**6.29**	**<0.05***
Position	1	0.66	0.43	**76.81**	**<0.001*****	**83.33**	**<0.001*****	0.15	0.71
Register	2	2.06	0.06	**61.09**	**<0.001*****	2.01	0.13	**4.63**	**<0.05***
Syllable × Position	1	**10.82**	**<0.01****	**4.35**	**<0.05***	**5.26**	**<0.05***	0.01	0.95
Syllable × Register	2	0.83	0.44	1.31	0.27	1.3	0.27	2.16	0.12
Position × Register	2	1.69	0.29	**16.88**	**<0.001*****	2.2	0.12	**3.43**	**<0.05***
Syllable × Position × Register	2	0.53	0.59	0.37	0.69	0.39	0.72	1.6	0.2

Appendix 4. Means (and standard deviations) for mean f0, Δ onset and Δ offset values of /ma/ and /ji/ across contexts (disyllabic word, right-branching noun phrase and left-branching noun phrase) and registers (IDS, Lombard speech, ADS). The unit is semitone for mean f0, Δ onset and Δ offset.

Register	Syllable	Context	Acoustic parameters		
			Mean f0	**Δ onset**	**Δ offset**
IDS	/ma/	Disyllabic	26.92 (3.94)	0.59 (1.17)	3.73 (3.57)
		Right-branching	27.09 (4.36)	0.23 (0.58)	4.87 (4.02)
		Left-branching	27.86 (3.13)	0.51 (0.81)	2.51 (2.12)
	/ji/	Disyllabic	25.63 (5.21)	6.46 (7.54)	2.10 (4.79)
		Right-branching	25.80 (6.01)	7.32 (8.21)	1.42 (3.64)
		Left-branching	29.82 (4.07)	0.41 (0.80)	1.01 (1.51)
Lombard speech	/ma/	Disyllabic	24.32 (2.82)	0.64 (1.44)	2.92 (2.60)
		Right-branching	23.50 (2.92)	0.23 (0.58)	3.24 (2.03)
		Left-branching	24.95 (2.55)	0.38 (0.60)	1.95 (1.38)
	/ji/	Disyllabic	22.27 (4.67)	7.46 (7.48)	1.67 (3.67)
		Right-branching	21.30 (5.71)	7.96 (7.75)	1.66 (3.45)
		Left-branching	26.70 (3.21)	0.23 (0.46)	0.53 (0.64)
ADS	/ma/	Disyllabic	21.77 (3.47)	0.86 (1.57)	2.33 (2.19)
		Right-branching	21.38 (3.67)	0.34 (0.93)	3.25 (2.56)
		Left-branching	22.85 (3.40)	0.48 (0.67)	1.51 (1.30)

	Disyllabic	19.26 (5.75)	7.59 (8.54)	2.42 (4.99)
/ji/	Right-branching	19.21 (6.11)	7.30 (7.85)	1.83 (4.18)
	Left-branching	24.11 (3.89)	0.28 (0.60)	0.50 (0.64)

Appendix 5. Results of the linear mixed-effects model on mean f0, Δ onset and Δ offset of two tone sandhi syllables /ma/ and /ji/ (presented in the upper and lower tables). Two fixed factors were included in these models: Context (disyllabic word, right-branching noun phrase and left-branching noun phrase) and Register (IDS, Lombard speech and ADS). Asterisks indicate the level of statistical significance: $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***. The unit is semitone for mean f0, Δ onset and Δ offset.

Syllable	Factors	DF	Acoustic parameters					
			Mean f0		Δ onset		Δ offset	
			F	*p*	F	*p*	F	*p*
/ma/	Context	2	**11.58**	**<0.001*** **	**14.31**	**<0.001*** **	**25.94**	**<0.001*** **
	Register	2	**53.68**	**<0.001*** **	1.62	0.23	**14.59**	**<0.001*** **
	Context × Register	4	1.96	0.1	1.25	0.29	2.36	0.05
/ji/	Context	2	**78.64**	**<0.001*** **	**58.29**	**<0.001*** **	**4.95**	**<0.05***
	Register	2	**225.6**	**<0.001*** **	0.35	0.71	0.86	0.44
	Context × Register	4	1.53	0.19	1.15	0.33	1.21	0.3

Appendix 6. Tukey HSD pairwise comparison for the acoustic parameters of syllables /ma/ and /ji/ across contexts (1: disyllabic word; 2: right-branching noun phrase; 3: left-branching noun phrase). Asterisks indicate the level of statistical significance: $p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***. The unit is semitone for mean f0, Δ onset and Δ offset.

Parameter	**Syllable**	**Group difference**	***β***	***SE***	**t**	***p***
Mean f0	/ma/	1-2	0.45	0.34	1.32	0.4
		1-3	**-0.79**	**0.28**	**-2.86**	**<0.05***
		2-3	**-1.25**	**0.28**	**-4.42**	**<0.001*****
	/ji/	1-2	0.26	0.4	0.65	0.79
		1-3	**-4.5**	**0.38**	**-11.82**	**<0.001*****
		2-3	**-4.76**	**0.47**	**-10.06**	**<0.001*****
Δ onset	/ma/	**1-2**	**0.42**	**0.08**	**5.1**	**<0.001*****
		1-3	0.22	0.1	2.29	0.08
		2-3	**-0.19**	**0.07**	**-2.6**	**<0.05***
	/ji/	1-2	-0.33	0.54	-0.61	0.82
		1-3	**6.96**	**0.66**	**10.54**	**<0.001*****
		2-3	**7.29**	**0.79**	**9.23**	**<0.001*****
Δ offset	/ma/	**1-2**	**-0.9**	**0.23**	**-3.87**	**<0.01****
		1-3	**0.94**	**0.18**	**5.29**	**<0.001*****
		2-3	**1.84**	**0.27**	**6.85**	**<0.001*****

	1-2	0.48	0.25	1.89	0.17
/ji/	**1-3**	**1.42**	**0.46**	**3.05**	**<0.05***
	2-3	0.94	0.43	2.16	0.1

Chapter Four: Acquisition of Weak Syllables in Tonal Languages: Acoustic Evidence from Neutral Tone in Mandarin Chinese

This chapter is based on the following published paper:

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., Demuth, K. (2018). Acquisition of weak syllables in tonal languages: acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, 1–27. https://doi.org/10.1017/S0305000918000296

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors and collaborators) when needed.

Abstract

Weak syllables in Germanic and Romance languages have been reported to be challenging for young children, with syllable omission and/or incomplete reduction persisting till age 5. In Mandarin Chinese, neutral tone (T0) involves a weak syllable with varied pitch realizations across (preceding) tonal contexts and short duration. The present study examined how and when T0 was acquired by 108 Beijing Mandarin-speaking children (3-5 years) relative to 33 adult controls. Lexicalized (familiar) and non-lexicalized (unfamiliar) T0 words were elicited in different preceding tonal contexts. Unlike previous reports, the present study revealed that children as young as 3 years have already developed a phonological category for T0, exhibiting contextually conditioned tonal realizations of T0 for both familiar and unfamiliar items. However, mastery of adult-like pitch and duration implementation of T0 is a protracted process not completed until age 5. The implications for the acquisition of weak syllables more generally are discussed.

Key words: weak syllables; Mandarin Chinese; neutral tone.

Introduction

During early phonological acquisition, young children demonstrate challenges in acquiring weak syllables (Gerken, 1994). For example, it has been shown that English-learning children at age 5 still omit weak syllables that appear pretonically, before a stressed syllable, e.g., *banana* produced as 'nana' (e.g., Ingram, 1974; Haelsig & Madison, 1986; Kehoe, Stoel-Gammon & Buder, 1995; Demuth, 1996). Even when such syllables are produced, they can persist in having longer (unreduced) vowel durations (e.g., Yuen, Demuth & Johnson, 2011). Similar phenomena have also been observed in other languages, including Dutch (Fikkert, 1993), French (Demuth & Johnson, 2003), German and Spanish (Kehoe & Lleó, 2003; Lleó, 2006).

However, most of these studies have focused on Germanic and Romance languages, which use vowels and consonants to contrast word meanings, and involve some type of stress or phrasal lengthening that provides the context for syllable omission/reduction. As most of the languages around the world are tonal, using lexical tone in addition to consonants and vowels to contrast meanings (Yip, 2002), but often having a more limited role for 'stress', this raises the question of how 'weak' syllables are acquired in such languages.

In Mandarin Chinese, for instance, in addition to the four lexical tones which occur on stressed/full syllables, there is also a toneless category, i.e., neutral tone, which occurs only on weak (short) syllables. Thus, neutral tone syllables share the acoustic attribute of being a short syllable, similar to weak syllables in Germanic and Romance languages (Fry, 1955). In addition, as a toneless category, neutral tone exhibits contextually conditioned acoustic realizations after different lexical tones. However, little is known about whether children learning Mandarin show difficulty in acquiring the acoustic aspects of neutral tone. Therefore, the goal of the present study was to explore when neutral tone is acquired by Mandarin-speaking children, crucial for providing a comprehensive account of these children's phonological acquisition above the level of the segment.

Mandarin Chinese has four lexical tones that contrast in pitch contour, i.e., Level tone 1 (T1), Rising tone 2 (T2), Dipping tone 3 (T3) and Falling tone 4 (T4; see Figure 1). Word meaning varies as a function of lexical tone (e.g., /ma1/ 'mother', /ma2/ 'hemp', /ma3/ 'horse' and /ma4/ 'scold'). These four tones appear early in acquisition, i.e., during the single-word stage of development (Li & Thompson, 1977), although confusion between T2 (the rising tone) and T3 (the dipping tone) continues into the 2/3-word stage of development. By the age of 3, when children begin to combine words into longer sentences, all lexical tones are reportedly acquired (Li & Thompson, 1977).

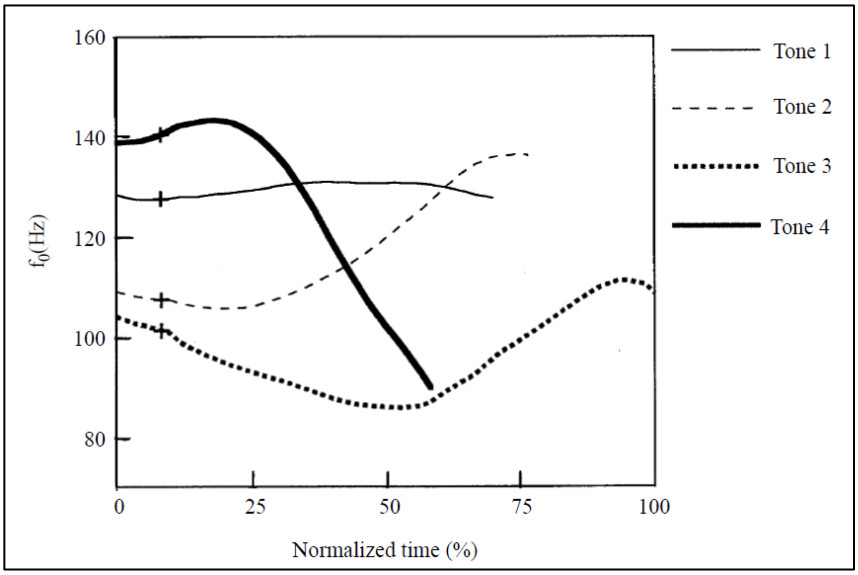


Figure 1. Mandarin Chinese Lexical tone pitch contours, from Xu (1997).

The toneless category, neutral tone, is also called the "fifth tone" or T0. Neutral tone is only carried by weak (short) syllables, appearing in the final position of a word. Neutral tone syllables can be classified into three semantic types: (morphological) suffix, reduplicative and lexeme types, as exemplified in Table 1 (Li & Thompson, 1977; Hua & Dodd, 2000). Of these various types, some neutral tone syllables belong to part of a disyllabic lexicalized/holistic word, which are not productive, such as the /**tsi0**/ in /t^{h}u4 **tsi0**/ *rabbit*. Other neutral tone syllables, such as the possessive particle /**tɤ0**/, can be combined with any noun, to form non-lexicalized/new neutral tone words (e.g., /tʂu1 **tɤ0**/ *pig's*; Table 1). The noun suffix /tsi0/ also has a full tone counterpart /tsi3/ 'child' bearing a tone 3; this is true for many neutral tone syllables. While these neutral tones were historically derived from their full tone counterparts, most of them are phonologically and semantically distinct from the full tone counterparts today (Shen, 1992).

Table 1. Neutral tone types and lexical tone counterparts

Type	**Example**	**Productivity**	**Lexical tone counterpart**
Suffix	Possessive particle /**tɤ0**/, e.g., /tʂu1 **tɤ0**/ 'pig's'	Productive	--
	Classifier particle /**kɤ0**/, e.g., /san1 **kɤ0**/ 'three items'	Productive	/kɤ4/
	Noun suffix /**tsi0**/, e.g., /t^{h}u4 **tsi0**/ 'rabbit'	Lexicalized	/tsi3/
Reduplicative	/ti4 **ti0**/ 'older brother'	Lexicalized	/ti4/
Lexeme	/t^{h}ou2 **fa0**/ 'hair'	Lexicalized	/fa1/

Unlike lexical tones, a feature of neutral tone is that it is phonologically 'under-specified', i.e. it does not have its own fixed tone. As a toneless category, its pitch implementation varies as a function of the preceding tone. Adults realise neutral tone as a fall in pitch after T1, T2 and T4, but as a rise or a level tone after T3[12] (see Figure 2; Tang, 2014). Neutral tone syllables are also reduced compared to lexical tone syllables, although the magnitude of shortening again varies as a function of the preceding tone. The overall mean duration of a neutral tone syllable is longer following a T3 syllable (about 70% of the preceding T3 syllable duration) than following T1/2/4 syllables (about 50% to 60% of the preceding T1/2/4 syllables; Cao, 1992; Tang, 2014). In other words, the preceding tonal context influences the tonal contour and duration of neutral tone. Learning to implement the pitch and durational features of this phonologically under-specified tonal category may therefore present a challenge for young children, as they must also learn to correctly modify its realization according to the tonal context.

[12] The mechanism that drives this contextual variation is debatable. See Yip (1989) for a tonal spreading (phonological) account and Chen and Xu (2006) for a mid-target (phonetic) account. Yet these accounts agreed on the phonetic observations that neutral tone varies its tonal contour as a function of preceding tonal contexts. In the present study, the focus was to examine when children can acquire neutral tone, despite its varied realizations.

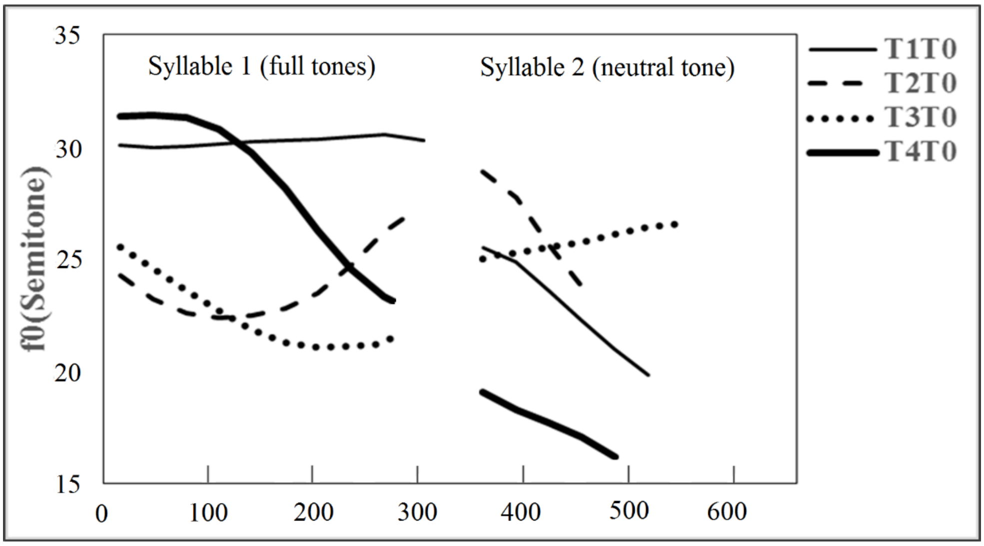


Figure 2. Pitch contour of Mandarin neutral tones (T0) in different tonal contexts (after T1-4), from Tang (2014). Please note that the duration of T0 is different when following different tones, i.e., about 70% of the preceding T3 syllable and about 50% to 60% of the preceding T1/2/4 syllables.

However, relative to lexical tone studies, neutral tone has received much less attention with respect to when it is acquired. According to the few references on this topic, neutral tone seems to be acquired much later than lexical tones, not being mastered until around the age of 4;6. For example, Hua and Dodd (2000) reported that no 3-year-olds correctly produced all neutral tone words, and only 36% of 4-year-olds could do so, suggesting that the acquisition of neutral tone is a protracted process for Mandarin-speaking children. Three types of errors were previously observed in children's neutral tone productions: (1) substituting the neutral tone with the lexical tone counterpart, (2) lengthening the neutral tone syllable, or (3) omitting the neutral tone syllable (for children under age 3) (Hua & Dodd, 2000). These results add to the expectation that neutral tone might be challenging to acquire.

However, in analysing children's neutral tone productions, previous studies were only based on subjective perceptual/auditory judgements, i.e., the authors transcribed children's productions and judged their neutral tone error patterns based on their own perceptual observations. So far there has been no *acoustic* investigation of children's neutral tone productions, it is therefore still unclear how and when children develop adult-like acoustic realizations of the neutral tone *category*. That is, do children show contextually conditioned pitch and duration variations for neutral tone, like adults? Can children reduce the neutral tone duration to the same degree as adults? These issues were addressed in the present study by conducting an acoustic analysis of children's neutral tone productions across different tonal contexts, comparing their productions to that of adult controls.

Moreover, previous studies only examined children's neutral tone productions in lexicalized items, i.e., noun suffixes, reduplicatives, etc., where the neutral tone syllable was part of a lexicalized word. However, it might be the case that a child who has successfully produced a lexicalized neutral tone word, such as "ma1 ma0" *mother*, may know nothing about the neutral tone category, but may be simply repeating the disyllabic word they know from their language input. Thus, previous investigations have mainly tapped into children's word knowledge (i.e. vocabulary) rather than examining their productive knowledge in generalising the neutral tone category to new words, as in a 'wug' task (e.g., Berko, 1958). In the present study, we therefore wanted to know when children develop a phonological representation for the neutral tone that can be productively generalized to form new words. For example, the neutral tone possessive particle /tɤ0/ can be productively composed with any noun to form a new possessive word.

To gain a better understanding of children's acquisition of the neutral tone category, the current study therefore examined the pitch and duration realization of 3-5-year-olds' neutral tone productions. This age range was selected since previous studies had reported that neutral tone syllable omission is not a problem for children above age 3 (Hua, 2002), and neutral tone is not fully acquired at 4;6 (Hua & Dodd, 2000). Two types of neutral tone words were adopted as stimuli, i.e., lexicalized/familiar and non-lexicalized/unfamiliar items. The non-lexicalized items were disyllabic words containing a monosyllabic noun and a monosyllabic possessive particle /tɤ0/, which does not have a lexical tone counterpart, e.g., /tʂu1 tɤ0/ 'pig's'. Exploring children's knowledge of these non-lexicalized items allow us to examine whether they have developed a robust neutral tone category generalizable to learning new words. Lexicalized neutral tone words (e.g., /ti4 **ti0**/ 'younger brother') were then included as known word control items. These words were all kinship terms which are familiar to young children and are reported to emerge early (Hua, 2002). Neutral tone syllables in these words also have a lexical tone counterpart that is identical to the first syllable of the disyllabic (reduplicated) word. The lexicalized neutral tones were therefore used to test whether children would substitute neutral tone with its lexical tone counterpart, as reported by Hua and Dodd (2000) and Hua (2002).

Based on previous studies (Li & Thompson, 1977; Hua & Dodd, 2000; Hua 2002), we assumed that the younger children would not have developed a robust neutral tone category, i.e., the contextually conditioned acoustic realization (pitch: falling after T1/2/4 and rising/level after T3; duration: shorter after T1/2/4 than T3) and reduced duration of neutral tone. Therefore, they would not be able to produce adult-like acoustic realizations of neutral tone for either the lexicalized or non-lexicalized items, at least not in the youngest group, i.e., 3-year-olds. We made the following predictions:

Hypothesis 1 (H1): (a) young children (i.e., 3-year-olds) would face challenges in producing contextually conditioned neutral tone *pitch* in an adult-like way for both lexicalized (reduplicative) and non-lexicalized (possessive) items; (b) the *pitch* of children's neutral tone productions would become more adult-like with age.

Hypothesis 2 (H2): (a) young children (i.e., 3-year-olds) would face challenges in producing contextually conditioned neutral tone *duration* in an adult-like way, producing longer durations than adults for both the lexicalized and non-lexicalized syllables, and (b) the duration of children's neutral tone productions would become more adult-like with age.

Method

Participants

One-hundred and eight children aged 3;1 to 6;2 were recruited and tested in the affiliated kindergarten of Beijing Language and Culture University. All children spoke Mandarin as their first language and were born and raised in Beijing. According to reports from the kindergarten, the recruited children did not have any speech, hearing, language or intellectual difficulty. These children were divided into three age-groups (3-, 4- and 5-year-olds) within a range of approximately one year (see Table 2). Since most 5-year-olds had graduated at the time of testing, there were fewer 5-year-olds than the other two age groups. Two young 6-year-old children (6;1 and 6;2) were also included into the group of 5-year-olds. In addition, 33 adult university students (mean: 20 yrs.; range: 19-25 yrs.) were recruited as controls. The adults were all native monolingual speakers of Mandarin and were born and raised in Beijing as well. The study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Panel.

Table 2. Number of participants in each age group

Age group	Male	Female	Total
3 yrs. (3;1-3;12, mean: 3;8, SD: 3 months)	13	31	44
4 yrs. (4;1-4;12, mean: 4;5, SD: 3 months)	20	22	42
5 yrs. (5;1-6;2, mean: 5;7, SD: 5 months)	9	13	22
Adults (19-25 yrs., mean: 20, SD: 2 years)	9	24	33

Stimuli

The investigation of neutral tone was part of a larger study investigating Mandarin-speaking children's acquisition of tones in context. To investigate the acquisition of neutral tone, eight disyllabic words were selected as stimuli. These consisted of four lexicalized kinship reduplicative words and four non-lexicalized possessive words. Within each disyllabic word, the first syllable was T1, T2, T3 or T4 and the second syllable was always T0 (see Table 3).

Table 3. Target words were of two types (reduplicative and possessive). Within each disyllabic word, the lexical tone of the first syllable was varied to be either T1, T2, T3 or T4 and the tone of the second syllable was always the neutral tone (T0).

Target word	Word type	Lexicalization	Tones	Meaning	Phonetic transcription
哥哥	Reduplicative	Lexicalized	T1+T0	*Older brother*	/kɤ1 **kɤ0**/
爷爷	Reduplicative	Lexicalized	T2+T0	*Grandpa*	/je2 **je0**/
姐姐	Reduplicative	Lexicalized	T3+T0	*Older sister*	/tɕie3 **tɕie0**/
弟弟	Reduplicative	Lexicalized	T4+T0	*Younger brother*	/ti4 **ti0**/
猪的	Possessive	Non-lexicalized	T1+T0	*Pig's*	/tʂu1 **tɤ0**/
牛的	Possessive	Non-lexicalized	T2+T0	*Cow's*	/niu2 **tɤ0**/
狗的	Possessive	Non-lexicalized	T3+T0	*Dog's*	/kou3 **tɤ0**/
鹿的	Possessive	Non-lexicalized	T4+T0	*Deer's*	/lu4 **tɤ0**/

Note that in each of the four disyllabic *lexicalized* reduplicative stimuli, both syllables shared the same segmental information, but differed in tone. Each of the four *non-lexicalized* disyllabic possessive stimuli consisted of a monosyllabic animal’s name plus the possessive particle /**tɤ0**/. All disyllabic reduplicative words and the first syllable of the possessive words fell within the top 20% of the most frequent disyllabic and monosyllabic words in the language input to children below 3 years, according to the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018) from the CHILDES database (MacWhinney, 2000).

Procedure

Children were tested in a quiet room at their kindergarten, and the adults were tested in a quiet room at their university. Only one participant was tested in each session. During test sessions participants wore a head-mounted cardioid-directional condenser microphone (AKG-C520) which was connected to a solid-state recorder (Marantz PMD661MKII).

The procedure used was an elicited production task. In eliciting the reduplicative kinship terms, four pictures were presented in a sequence to the participant on a computer screen. In each picture, a pair of family members depicted in line drawings were presented side-by-side, such as an older brother (illustrated by a tall boy) and a younger sister (illustrated by a shorter girl). To elicit the word /kɤ1 **kɤ0**/ 'older brother', for example, the Mandarin-speaking experimenter would point to the picture and ask the participant: "这个男生叫这个女孩妹妹，那么这个女孩叫这个男生什么?" 'This boy calls this girl 'younger sister', so what does the girl call the boy?' Once the participant produced the target word, e.g., /kɤ1 kɤ0/ 'older brother', the experimenter proceeded to the next picture to elicit the next target word. If the target word was not produced, the experimenter would rephrase and repeat the question. For example, if instead of saying the target word "older brother" a child said "boy", the experimenter would rephrase the question and ask them again, i.e., "他叫她妹妹，她叫他什么?" 'He calls her younger sister, so what does she call him?' The target words were never used when the experimenter rephrased the question. All participants succeeded in producing the target word after one or two attempts. The tone of the first syllable varied across items, providing the tonal context for variable acoustic realization of the neutral toned syllable as well.

To elicit the possessive items, two animals depicted in line drawings were simultaneously presented side-by-side, e.g., a pig (/tʂu1/) and a cow (/niu2/). To elicit the target neutral tone / tɤ0/ in a disyllabic word, such as /tʂu1 **tɤ0**/ 'pig's' and /niu2 **tɤ0**/ 'cow's', the experimenter first introduced the pig and the cow respectively: "这是一只猪, 这是一头牛" 'This is a pig; this is a cow', and then pressed a button to play a pre-programmed animation (e.g., tail spinning) on one of the animals, (e.g., a pig). During the animation, the experimenter asked: "这个尾巴是谁的?" 'Whose tail is it?' After the participant produced the target word (e.g., /tʂu1 tɤ0/ 'pig's'), the experimenter then pressed the button to trigger an animation of a spinning tail on the other animal, for example, the cow, and asked the same question to elicit another target item (e.g., /niu2 tɤ0/ 'cow's'). Note that, again, the first word/syllable (in this case, the animal) varied in tone, so that the acoustic realization of the neutral toned syllable 'tail' would be different in each newly formed disyllabic word. The same procedure and protocol were followed for the rest of the possessive stimuli. The order of the target words and their corresponding pictures were counterbalanced across participants to avoid any order effect. If participants did not produce the target words, the experimenter would rephrase and repeat the question. For example, instead of saying the target word in isolation "猪的" 'pig's', some children produced the target word in a clause, e.g., "**是**猪的" '**is** pig's'. In this case, the experiment would ask them to repeat without the verb "是" 'is'. All participants succeeded in producing the target words after one or two attempts. A total of eight test items were collected per participant, with four reduplicative words and four possessive stimuli.

Coding and measurements

The vowels were acoustically coded for both duration and pitch using Praat (Boersma & Weenink, 2016). To do this, the temporal landmarks (vowel onset and vowel offset) were identified for each syllable of the word based on the onset and offset of the second formant (F2) information in the spectrogram (see Figure 3). One of the reduplicative stimuli /je2 **je0**/ 'grandfather' contained the glide /j/, making it difficult to separate /j/ from /e/. Therefore, for this item, the glide was also included as part of the vowel, with the sonority trough in /j/ between the two vowels used to demarcate the first from the second syllable. This did not affect the overall durational measures because our analysis of neutral tone duration was normalized with reference to the first syllable. Since the neutral toned syllable of all remaining test words began with either a stop or affricate, such as /kɤ1 **kɤ0**/ 'older brother' or /tɕie3 **tɕie0**/ 'older sister', closure duration helped to demarcate the two syllables. This is illustrated in Figure 3. Duration and pitch contour (as measured by f0) were then extracted from each annotated syllable. F0 points were tracked within the annotated interval, using the default autocorrelation algorithm in Praat (Boersma, 1993); these were checked and manually revised to correct for any "doubling" or "halving" errors in pitch tracking. The revised pitch track was then interpolated and smoothed with a bandwidth of 20 Hz. Ten percent of the items were re-coded by a second trained native speaker of Mandarin. Interrater reliability on the duration of annotated vowels was good ($r = 0.92$).

Two types of measures were derived from the vowel portion of neutral tone syllables: F0 and normalized duration. The pitch contour of the neutral tone was based on 10 pitch points taken from each vowel. In order to minimize tonal coarticulation and pitch perturbation from neighbouring consonants, the initial and final 5% of the vowel proportion were excluded. In the analysis, pitch values were transformed to semitones from Hz values with 50 Hz as the reference to match human perception, using the following formula:

$$\text{Semitone} = 12 * \log_2 (\text{target Hz}/50).$$

The neutral tone duration was the temporal difference between the vowel onset and offset of the neutral tone; the lexical tone duration was the temporal difference between the vowel onset and offset of the lexical tone. The normalized vowel duration was a ratio between the neutral tone duration in millisecond (ms) and the lexical tone duration in ms, using the following formula:

$$\text{Normalized Neutral Tone Duration} = \frac{\text{Raw Neutral Tone Duration (ms)}}{\text{Raw Preceding Lexical Tone Duration (ms)}}$$

Thus, a ratio that was smaller than 1 indicated that the neutral tone duration was reduced relative to the lexical tone.

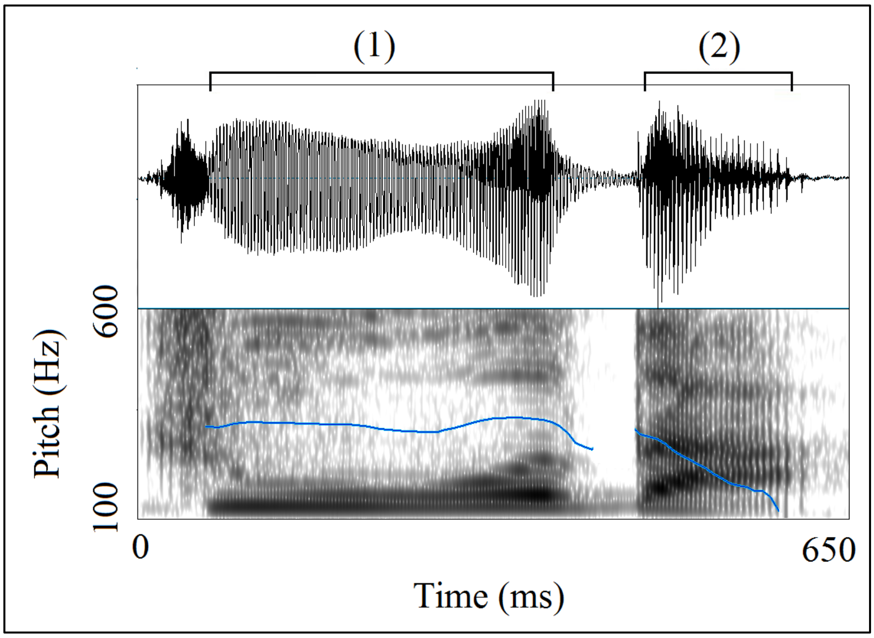


Figure 3. Waveform and spectrogram for the word /tʂu1 tɤ0/ 'pig's'. This token was produced by a 4-year-old boy. (1) and (2) illustrate the vowel portion of syllables /tʂu1/ and /tɤ0/ respectively.

Statistical analysis

A total of 1104 tokens were included in the analysis, with 842 tokens from children and 262 from adults. An additional 24 tokens produced by children were excluded from the analysis due to poor acoustic quality, including environmental noise, whispered speech or productions with prolonged creak.

The data were analysed using R (R Core Team, 2016). To quantify the pitch contour of children's and adults' neutral tone productions, a second-order orthogonal polynomial equation was used to model tone production, using the *poly* function of R. The second-order polynomials were adopted since the most complex pitch contour of tones in our data has only a convex or concave contour shape. The poly equation generated three parameters to characterise each pitch contour, i.e., (1) the intercept,

(2) the linear trend and (3) the quadratic trend. With reference to Mirman (2014), the intercept captures f0 contour's overall *height* (a large intercept indicates a high pitch height and vice versa). The linear trend models the *general direction of the contour* (a positive linear trend indicates a rising pitch and vice versa). The curvature approximates the *pitch contour* by estimating the shape deviated from a linear trend (a positive quadratic trend indicates a concave in the linear approximation of a pitch contour, whereas a negative quadratic trend indicates a convex in an f0 contour, and a small positive quadratic trend indicates a level f0 contour). These parameters were used to evaluate any group differences in f0 contour.

Linear mixed-effect models were built to compare the three parameters of tone production from children and adults, using LME4 package (Bates, Mächler, Bolker, & Walker, 2014). All random slopes were included in the model to make it maximally generalizable across the data (Barr, Levy, Scheepers, & Tily, 2013). There are different approaches to estimate the significance of fixed factors in linear-mixed effects model (Luke, 2017). In the present study, the significance of fixed effects was estimated using the *anova* function in the R package LMERTEST, which provides Satterthwaite's approximation to derive degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2015). This function reported omnibus effects for multi-level factors or interactions using *F*-tests, rather than comparisons with the baseline level using *t*/*z* tests. The main/interaction effects reported in the results were averaged across all levels of the other effects, and the effect of multi-level factors was an omnibus effect (see Peters, Hanssen, & Gussenhoven (2014) and Tang, Xu Rattanasone, Yuen & Demuth (2017b) for examples). Relative to other approaches adopted to generate *p* values for fixed effects, such as the likelihood ratio test (LRT) for model comparison, the Satterthwaite's method is a good alternative as it outperforms LRT especially in cases with unbalanced and/or small sample designs (Kuznetsova, Brockhoff, & Christensen, 2015). When a significant main effect of a multi-level factor or a significant interaction effect was observed, Bonferroni adjusted post-hoc comparisons were performed on the multi-level factor, as well as interactions, using LSMEANS package (Lenth, 2016).

Results

Pitch

The pitch contours of both the children's and adults' productions are illustrated in Figure 4. Visual inspection of Figure 4 shows that, for both lexicalized and non-lexicalized neutral tone words, all child groups and adults produced a falling pitch for neutral tones after T1/2/4 and a level/rising pitch after T3, consistent with the pitch variations of neutral tone reported in previous studies (Tang, 2014).

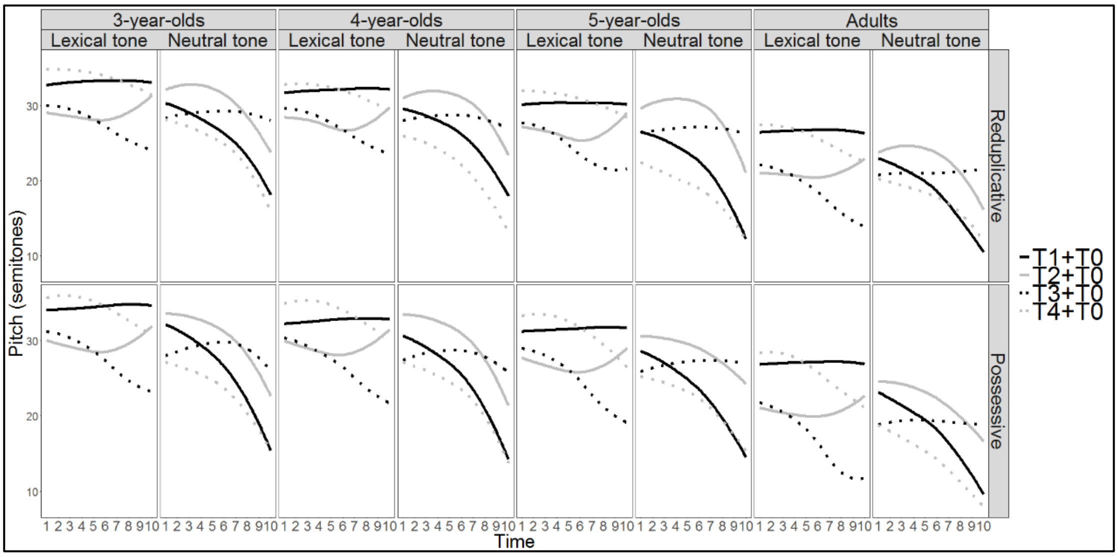


Figure 4. Pitch contours of lexical tone and neutral tone in two word types (reduplicatives on the top row, possessives on the bottom row) across tonal contexts (following T1/2/3/4 syllable) produced by children (3-, 4- and 5-year-olds, from left to right) and adults. The duration on the x-axis is normalized.

We predicted in H1a that children might face challenges in producing contextually conditioned neutral tone *pitch* in an adult-like way for both lexicalized (reduplicative) and non-lexicalized (possessive) items, and in H2b that the pitch of children's neutral tone productions would become more adult-like with age. To test these hypotheses, a linear mixed regression model was constructed using the three pitch parameters, i.e., pitch height (intercept), slope (linear trend) and curvature (quadratic trend), across 10 time steps to explore the group difference of pitch. Three fixed factors: "Group" (3-, 4-, 5-year-olds and adults), "Type" (reduplicative and possessive) and "Tonal Context" (T1, T2, T3 and T4), and a random factor: "Participant" were included in the model.

As children tend to have higher pitch than adults, the pitch height (intercept) would be predictably higher in children than adults. Our results confirmed this: the main effect of "Group" was significant on the pitch height ($F(3, 141) = 48.59$, $p < 0.001$; Table 4). Post-hoc analysis revealed that 3-year-olds exhibited the highest pitch height, followed by 4- and 5-year-olds, with adults showing the lowest pitch height (Appendix 1).

In terms of the *shape* of the pitch contours (pitch slope or curvature), an interaction of "Group × Tonal Context" was expected as we predicted that children would not produce adult-like pitch variation of neutral tone. In addition, developmental differences between the younger children and adults were anticipated, though perhaps not between the oldest children and adults in pairwise comparisons.

The results of the model are presented in Table 4, which shows that there was a significant three-way "Group × Type × Tonal Context" interaction on the **pitch slope,** and a significant two-way "Group × Type" interaction on the **pitch curvature**. Bonferroni adjusted post-hoc tests were then performed on the two interactions to further compare the group difference in terms of neutral tone pitch contours across conditions.

Table 4. Results of linear mixed regression model with second-order polynomials on pitch points of neutral tone across age groups (3-, 4- and 5-year olds and adults), types (reduplicatives and possessives) and tonal contexts (T1-4). Items in bold indicate significant findings.

Trends	**Factors**	df 1	**df 2**	**F**	***p***
Intercept	**Group**	**3**	**141**	**48.59**	**<0.001*****
	Type	1	141	1.31	0.254
	Tonal Context	**3**	**140**	**157.06**	**<0.001*****
	Group × Type	**3**	**141**	**4.19**	**<0.01****
	Group × Tonal Context	9	140	1.31	0.236
	Type × Tonal Context	**3**	**10283**	**9.54**	**<0.01****
	Group × Type × Tonal Context	**9**	**10279**	**18.76**	**<0.001*****
Linear trend	Group	3	141	2.32	0.078
	Type	**1**	**10214**	**73.71**	**<0.001*****
	Tonal Context	**3**	**10213**	**1258.09**	**<0.001*****
	Group × Type	**3**	**10214**	**8.36**	**<0.01****
	Group × Tonal Context	9	10213	1.31	0.236
	Type × Tonal Context	**3**	**10214**	**9.99**	**<0.01****
	Group × Type × Tonal Context	**9**	**10214**	**7.43**	**<0.001*****
Quadratic trend	**Group**	**3**	**147**	**4.86**	**<0.01****
	Type	1	10211	0.54	0.464

Tonal Context	**3**	**10211**	**53.77**	**<0.001*** **
Group × Type	**3**	**10211**	**4.43**	**<0.01** **
Group × Tonal Context	9	10210	1.56	0.121
Type × Tonal Context	**3**	**10211**	**11.7**	**<0.001*** **
Group × Type × Tonal Context	9	10211	1.38	0.192

Note. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Type * Context + (1 + Type + Context + Linear trend + Quadratic trend | Participant : Group)

Post-hoc comparisons of the **pitch slope** for the three-way interaction revealed three observations. Relative to adults, (a) 3-year-olds showed a more falling pitch contour for the reduplicative neutral tone after T4, and for the possessive neutral tone after T1; (b) 4-year-olds also showed a more falling pitch contour for the reduplicative neutral tone after T4, and for the possessive neutral tone after T2; (c) 5-year-olds did not differ from adults in the neutral tone pitch contour for any condition (also reflected by small effect sizes between 5-year-olds and adults, i.e., Cohen's d < 0.1; see Appendix 1). Consistent with these results, there were then also some pitch differences among the child groups: relative to 5-year-olds, 3-year-olds showed a more falling pitch contour for the possessive neutral tone after T1, and 4-year-olds showed a more falling pitch contour for the possessive neutral tone after T2. However, there were no significant differences between the 3- and 4-year-olds. These results indicate that all child groups showed contextually conditioned pitch variation of neutral tone syllables across tonal contexts, i.e., falling after T1/2/4 and rising/level after T3. While 3- and 4-year-olds exhibited a more falling pitch contour for neutral tones after T1/2/4, this did not distort their contextual pitch variation.

Post-hoc comparisons of the **pitch curvature** for the two-way interaction revealed two observations. Relative to adults, 3- and 4-year-olds produced a more curved pitch contour across all tonal contexts for the possessive neutral tone (see Figure 4). There were no other significant differences among groups (see Appendix 2). These results suggest that the 3- and 4-year-olds exhibited a more falling pitch contour for neutral tone in the possessive than adults. In contrast, the 5-year-olds did not differ from adults in the pitch curvature for either the possessive and reduplicative neutral tone, and this is also supported by small effect sizes (Cohen's d: -0.12 for reduplicatives and -0.04 for possessives).

To further explore the developmental trend in the pitch contour of children's neutral tone, a linear regression model was constructed on children's productions with age coded as a continuous factor (in months). Three fixed factors "Age" (from 38 to 74 months), "Type" (reduplicative and possessive) and "Tonal Context" (T1, T2, T3 and T4) and a random factor: "Participant" were included. The dependent variables were pitch slope and curvature since these two parameters reflected the *shape* of pitch contours. We expected a main effect of "Age" because we predicted that children's realization of the neutral tone contour would become more adult-like as they mature. The results are presented in Table 5, which shows that there was neither a main effect of "Age" nor an interaction between "Age" and other factors, and therefore post-hoc test was not conducted. Thus, overall, the direction and curvature of the pitch contour for neutral tone did not exhibit developmental differences between the ages of 3 and 5.

Table 5. Results of linear mixed regression model with second-order polynomials on the pitch points in children's neutral tone productions only, with children's age coded as continuous factor (in month). Three fixed factors were included: children's Age (from 38 to 74 months), type (reduplicatives and possessives) and tonal context (T1-4). Items in bold indicate significant findings.

Trends	**Factors**	**df 1**	**df 2**	**F**	***p***
Linear trend	Age	1	108	3.10	0.081
	Type	1	8211	12.76	**<0.001*** **
	Tonal Context	3	8210	12.07	**<0.001*** **
	Age × Type	1	8213	1.17	0.35
	Age × Tonal Context	3	8210	0.94	0.422
	Type × Tonal Context	3	8212	3.64	**<0.05***
	Age × Type × Tonal Context	3	8212	2.15	0.092
Quadratic trend	Age	1	112	1.84	0.178
	Type	**1**	**8206**	**5.02**	**<0.05***
	Tonal Context	3	8205	0.52	0.672
	Age × Type	1	8206	1.75	0.166
	Age × Tonal Context	3	8205	1.16	0.323
	Type × Tonal Context	3	8205	1.65	0.177
	Age × Type × Tonal Context	3	8205	1.80	0.146

Note. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Age * Type * Context + (1 + Linear trend + Quadratic trend | Participant)

In summary, counter to the prediction in H1a, all child groups generally showed adult-like pitch variations across tonal contexts, namely, a falling pitch for neutral tones after T1/2/4 and a rising or level pitch for neutral tones after T3, for both lexicalized and non-lexicalized neutral tone types (see Figure 4). Our results are also partly consistent with H1b, showing that, although there was no general developmental trend in the neutral tone pitch contour as a function of children's age, the younger children (3- and 4-year-olds) showed a more falling pitch component in a few tonal contexts compared to adults. By 5 years, however, children produced neutral tone pitch contours that were very similar to those of adults across all tonal contexts.

Duration

The normalized durations of children's and adults' neutral tone productions are presented in Figure 5. Visual inspection shows that both children and adults produced longer duration for neutral tones following T3 compared to T1/2/4 across word types. This is consistent with the durational variations of neutral tone reported in previous studies (Tang, 2014; Cao, 1992).

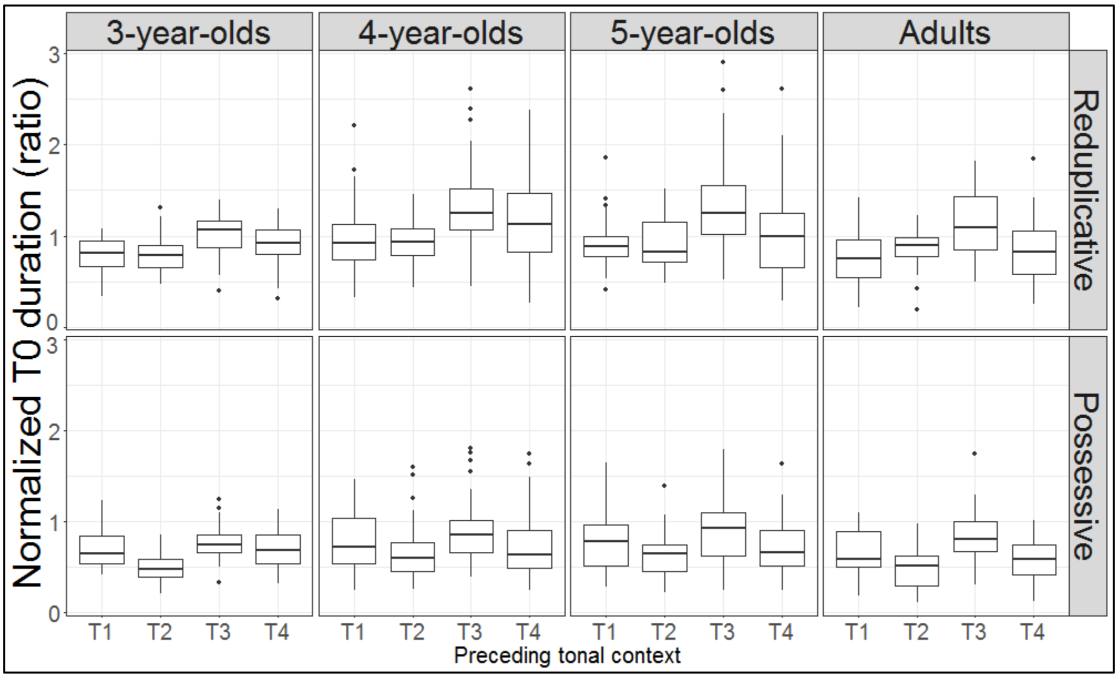


Figure 5. Normalized duration of neutral tone across word types (reduplicatives on the top row, possessives on the bottom row) and tonal contexts (following T1/2/3/4 syllable) produced by children (3-, 4- and 5-year-olds, from left to right) and adults.

We had predicted in H2a that 3-year-olds might face challenges in producing contextually conditioned neutral tone duration in an adult-like way, producing longer duration than adults for both lexicalized and non-lexicalized neutral tones. We also predicted in H2b that the duration of children's neutral tone productions would become more adult-like with age. To test these hypotheses, a linear mixed-effects model was built for the normalized duration of neutral tone syllables with three fixed factors: "Group" (3-, 4-, 5-year-olds, and adults), "Type" (the reduplicative and the possessive words), preceding "Tonal Context" (T1, T2, T3 and T4), and the random factor: "Participant". For H2a, the model anticipated a main effect of "Group" since we predicted that children would not reduce the neutral tone duration to the same degree as adults; the model also expected an interaction of "Group × Tonal Context" because children might not produce adult-like durational variation of neutral tone across all tonal contexts. For H2b, the model predicted significant pairwise comparisons between 3-year-olds and adults, but perhaps not between 5-year-olds and adults.

The results of the comparison are presented in Table 6, which shows as predicted a main effect of "Group", but also a significant three-way interaction of "Group × Type × Tonal Context". A Bonferroni post-hoc test was then conducted on the 3-way interaction to compare the durational difference of neutral tone between children and adults across word types and tonal contexts.

Table 6. Results of linear mixed-effect model of the normalized duration of neutral tone across age groups (3-, 4- and 5-year olds and adults), types (reduplicatives and possessives) and tonal contexts (T1-4). Items in bold indicate significant findings.

Factors	df 1	df 2	F	*P*
Group	**3**	**139**	**6.03**	**<0.01****
Type	**1**	**142**	**220.88**	**<0.001*****
Tonal Context	**3**	**168**	**44.38**	**<0.001*****
Group × Type	3	142	2.06	0.109
Group × Tonal Context	9	168	1.05	0.400
Type × Tonal Context	**3**	**549**	**15.22**	**<0.001*****
Group × Type × Tonal Context	**9**	**547**	**2.28**	**<0.05***

Note. R code for this model: Normalized duration ~ Group * Type * Context + (1 + Type + Context | Participant : Group)

Results of the post-hoc test showed that children generally reduced the neutral tone duration to a similar degree as adults across tonal contexts and types, except for the following conditions: relative to adults, (a) 3-year-olds produced longer durations for the reduplicative neutral tone after T3 and T4, and for the possessive neutral tone after T2; (b) 4-year-olds exhibited longer duration only for the reduplicative neutral tone after T3; (c) 5-year-olds did not differ from adults for the two neutral tone types (also reflected by small effect sizes: Cohen's d from -0.2 to 0.14; Appendix 3). Moreover, the results also showed that, among the child groups, there was no difference between either the 3- and 4-year-olds nor between the 4- and 5-year-olds.

Yet relative to 5-year-olds, 3-year-olds produced longer durations in the reduplicative words after T4, and in the possessive word after T2[13].

To further explore the developmental trend in the duration of neutral tone among children, a linear regression model was constructed on children's neutral tone productions. Children's age (in months) was coded as a continuous factor, with three fixed factors "Age" (from 38 to 74 months), "Type" (reduplicative and possessive) and "Tonal Context" (T1, T2, T3 and T4) and a random factor: "Participant". H2b would predict a main effect of "Age".

The results of the model are presented in Table 7, showing that there was a significant main effect of "Age" on the duration of neutral tone. No interaction between "Age" and other factors was found. To further explore the main effect of "Age", a Pearson correlation test was used to examine the relationship between children's age and the duration of their neutral tone. The results revealed a significant but weak negative correlation between these two parameters (r (840) = -0.134, $p < 0.001$), suggesting that (normalized) neutral tone duration becomes shorter, and more adult-like, as children mature.

[13] One of the anonymous reviewers pointed out that the procedure to elicit the possessives might induce the use of contrastive focus, because we presented two objects (i.e., a cow and a pig) and then played an animation (tail rotation) on each animal to elicit the target words 'cow's' and 'pig's' respectively. This might lead to the use of contrastive focus induced by the order effect. However, acoustic analysis revealed that the order in which the items were presented and produced yielded the same durational ratio for the same words. Thus, there is no order effect: $p = 0.09$, suggesting no contrastive focus.

Table 7. Results of linear mixed regression model on the duration in children's neutral tone productions only, with children's age coded as continuous factor (in month). Three fixed factors were included: children's Age (from 38 to 74 months), type (reduplicatives and possessives) and tonal context (T1-4). Items in bold indicate significant findings.

Factors	**df 1**	**df 2**	**F**	***P***
Age	**1**	**107**	**6.58**	**<0.05***
Type	**1**	**110**	**5.28**	**<0.05***
Tonal Context	3	133	2.02	0.114
Age × Type	1	110	0.00	0.988
Age × Tonal Context	3	133	1.01	0.389
Type × Tonal Context	3	423	2.52	0.057
Age × Type × Tonal Context	3	424	1.30	0.273

Note. R code for this model: Normalized duration ~ Group * Type * Context + (1 + Type + Context | Participant : Group)

In summary, the results from the analysis of neutral tone duration indicate that all children generally showed adult-like durational variation of neutral tone syllables across tonal contexts, with a shorter neutral tone duration after T1/2/4 than T3 (see Figure 5). In addition, children also generally reduced the neutral tone syllable duration to the same degree as adults, though in a few tonal contexts the 3- and 4-year-olds produced longer neutral tone durations than adults, e.g., the reduplicative neutral tone after T3/4 and the possessive neutral tone after T2. This suggests that these young children might be inconsistent in their realization of duration for neutral tone. By 5 years, however, children exhibited adult-like realizations of neutral tone duration in all tonal contexts. Thus, children's neutral tone productions generally become shorter as they get older.

Discussion

This study investigated the acoustic realization of neutral tone by 3-5-year-old Mandarin-learning children. The results showed that 3-year-olds have already developed the neutral tone category for both lexicalized/familiar and non-lexicalized/unfamiliar neutral tone items, as reflected by their contextually conditioned tonal realizations. However, adult-like acoustic implementation of neutral tone was more protracted, with differences in pitch and duration between child and adult productions disappearing by age 5.

These results thus provide partial support for previous studies (e.g., Li & Thompson, 1977; Hua & Dodd, 2000) that reported that children at 4;6 still did not acquire neutral tone, tending to replace the pitch of neutral tone with a full lexical tone and/or lengthened the duration of neutral tones, showing a limited understanding of the neutral tone category. However, our results showed that 3- and 4-year-olds have already developed the neutral tone category, albeit with occasional use of more falling pitch contours and longer durations.

A possible reason for the different results between the present and previous studies might be related to the task difference. The present study used a new word-formation task (for possessives) and a picture-naming task (for reduplicatives), whereas previous studies only employed a picture-naming task for known words (Li & Thompson, 1977; Hua & Dodd, 2000). The new word-forming task requires children to generate a new disyllabic word by combining a lexical tone with a neutral tone. This taps into children's productive knowledge of neutral tone as a phonological category. In contrast, the picture-naming task adopted in previous studies only taps children's word knowledge (i.e. vocabulary). Therefore, the word-formation task used in the present study is more challenging than the picture-naming task adopted in previous studies, for both the familiar words (kinship reduplicatives) and the unfamiliar (possessive) words. In addition, in the present study, the picture-naming task required children to use neutral tone in communicative situations, i.e., producing a kinship term to indicate the relationship between two relatives, i.e., grandma vs. grandpa. In previous studies, however, the picture-naming task only required children to name a neutral tone object, i.e., "xing1 xing0" *star*. Given the more complex tasks (both the word-formation task and the picture-naming task) used in the present study, one might have expected a higher proportion of errors. However, our results suggest that this was not the case. The different results between the present and previous studies must therefore be driven by other factors.

One possibility for the different results might be in the coding methods used. The present study used acoustic analysis to investigate the fine-grained pitch and durational realization of children's neutral tone productions, whereas previous studies used a subjective auditory transcription method, where the accuracy of children's neutral tone productions was determined by a single transcriber, with no reported intertranscriber reliability for tones (though Hua & Dodd (2000) and Hua (2002) report intertranscriber reliability for consonants and vowels). As children tend to speak more slowly than adults, neutral tone in children's productions will be longer than that in adult productions, and this could have biased the transcriber's judgement in previous studies, leading them to misinterpret neutral tone in slower speech as a full lexical tone. Indeed, our data suggest that the mean raw duration of neutral tone productions is 0.21s for 3- and 4-year-olds, 0.18s for 5-year-olds and 0.15s for adults. It has also been shown that phonetic expectation can bias perceptual transcription (Oller & Eilers, 1975). It is precisely for this reason that we used ratios in the current study (syllable 2/syllable 1) to compare child and adult productions. This showed that children's syllable ratios (comparing the duration of the first and second (T0) syllable) for disyllables, were adult-like, with only occasional lengthening of the neutral tone syllable by 3- and 4-year-olds. In future study it would be interesting to compare perceptual and acoustic analysis/coding of children's neutral tone production to determine the extent to which the two would yield similar results.

Another possibility that might explain the different results between the current study and previous studies might be related to the stimulus difference. The current study examined children's neutral tone productions in two different word types, i.e., reduplicatives and possessives; whereas Li and Thompson (1977) and Hua and Dodd (2000) also examined neutral tone in noun suffixes and lexemes. It might be the case that these different types of neutral tone pose different challenges for young children. This deserves further investigation in future studies.

However, despite the early emergence of neutral tone representations, our study found that adult-like acoustic implementation of neutral tone was not fully achieved until 5 years. Relative to adults, the 3- and 4-year-olds occasionally showed a more falling pitch for neutral tones. These might be related to the acoustic features of children's early language input, where the pitch contour of Mandarin lexical tone and neutral tone is exaggerated, leading to tone hyperarticulation, i.e., more falling tone and rising pitch contours (e.g., Tang et al., 2017a, b). Alternatively, these more falling pitch contours might reflect a tendency for young children to replace neutral tone with a lexical falling tone T4. This could be examined in future studies. Our results also found that 3- and 4-year-olds did not shorten the duration of their neutral tone productions to the same degree as adults, especially for the lexicalized reduplicatives (children lengthened reduplicatives in more tonal contexts than possessives). Perhaps this difference is due to the fact that the lexicalized words are learnt at a younger age, with a lengthened duration, and children at 3 or 4 years have not yet updated the acoustic realization of these early-learnt forms. The more adult-like implementation of neutral tone in non-lexicalized items, in contrast, indicates that children *have* developed a robust category for neutral tone and can generalise this productively to new words.

Our results therefore reveal a slightly different pattern of weak syllable acquisition from the acquisition of pretonic weak syllables in English. For instance, the predominant error pattern in English-learning children's weak syllable productions is syllable omission (Haelsig & Madison, 1986), and this phenomenon interacts with the stress pattern of words. For example, children were more likely to omit the weak syllable of a weak-strong word like *giraffe* than the weak syllable of a strong-weak word like *tiger* (Gerken, 1994; Demuth, 1996). In Mandarin, however, weak syllables are manifested in a toneless category with short duration and contextually conditioned tonal realization. The stress pattern of neutral tone words is always strong-weak, i.e., full tone + neutral tone, and therefore Mandarin-learning children do not typically omit neutral tone syllables. However, our results suggested that 3- and 4-year-olds still produced lengthened neutral tone productions. This is similar to findings in English where children sometimes produced weak syllables with longer (unreduced) vowel durations (e.g., Yuen, Demuth & Johnson, 2011). Taken together, these studies suggest that mastering adult-like acoustic realizations of weak syllables is a protracted process, sometimes deleting and sometimes lengthening, depending on the specific linguistic contexts of the language.

Finally, there are some limitations in the present study. First, as this study was part of a larger study on the acquisition of tones in context, only a few tokens of each neutral tone context were tested. In addition, the number of participants was unbalanced across groups (fewer 5-year-olds), and this might have resulted in insufficient power for the group comparison. Future study of neutral tone could include more items and a more balanced number of participants across age groups to confirm the reliability and generalizability of the current results. It would also be interesting, in future acoustic studies, to test children younger than 3 years, to investigate when and how the neutral tone category begins to be acquired.

Conclusion

The goal of this study was to conduct an acoustic analysis of how and when Mandarin-speaking children begin to produce adult-like pitch and durational cues for the short (weak) neutral tone. This is all the more challenging as both cues vary depending on the tone of the preceding syllable, leading to previous claims that neutral tone acquisition is a protracted process. However, our results show that children have extracted the phonological category of neutral tone from its varied surface forms by age 3, though they continue to refine their acoustic implementation of neutral tone in terms of pitch contour and duration, becoming more adult-like by the age of 5. This result is consistent with findings from other languages, suggesting that the mastery of adult-like weak syllable implementation is protracted. The acoustic analysis used here provides a framework for exploring these issues in a more nuanced manner, providing insight into the development of not only Mandarin tonal contrasts, but other (weak syllable) phonological processes as well.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R Package, version 1 (7), 1-23.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2-3), 150-177.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences* (Vol. 17, No. 1193, pp. 97-110).

Boersma, P., & Weenink, D. (2016). Praat: Doing Phonetics by Computer. [Computer program]. Version 6.0. 19.

Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics*, 20(3), 49-50.

Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese-speaking children's play narratives. *Narrative Inquiry*, 8 (1), 77-111. DOI: 10.1075/ni.8.1.05cha

Chen, Y., & Xu, Y. (2006). Production of Weak Elements in Speech–Evidence from F_0 Patterns of Neutral Tone in Standard Chinese. *Phonetica*, 63(1), 47-75.

Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: the Tong corpus. *Journal of Chinese Linguistics*, 46 (1), 69-92.

Demuth, K. 1996. The prosodic structure of early words. In J. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates. pp. 171-184.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 48(3-4), 211-241.

Fikkert, P. (1993). The acquisition of Dutch stress. In M. Verrips & F. Wijnen (eds), *Amsterdam Series in Child Language Development*, Vol. 1. Publisher: Amsterdam (Amsterdam): Universiteit van Amsterdam (University of Amsterdam), Instituut voor Algemene Taalwetenschap (Institute of General Linguistics). pp. 21-35.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4), 765-768.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, 21(3), 565-584. https://doi.org/10.1017/S0305000900009466

Haelsig, P. C., & Madison, C. L. (1986). A study of phonological processes exhibited by 3-, 4-, and 5-year-old children. *Language, Speech, and Hearing Services in Schools*, 17(2), 107-114.

Hua, Z. (2002). *Phonological Development in Specific Contexts: Studies of Chinese-Speaking Children*. Clevedon: Multilingual Matters Ltd.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3-42.

Ingram, D. (1974). Phonological rules in young children. *Journal of Child Language*, 1(1), 49-64. https://doi.org/10.1017/S0305000900000076

Kehoe, M., & Lleó, C. (2003). A phonological analysis of schwa in German first language acquisition. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 48(3-4), 289-327.

Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). *Acoustic correlates of stress in young children's speech*. *Journal of Speech, Language, and Hearing Research*, 38(2), 338-350.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. R Package, version 2(0).

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1-33. https://doi.org/10.18637/jss.v069.i01

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199. https://doi.org/10.1017/S0305000900001598.

Lleó, C. (2006). The acquisition of prosodic word structures in Spanish by monolingual and Spanish-German bilingual children. *Language and Speech*, 49(2), 205-229. https://doi.org/10.1177/00238309060490020401.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494-1502.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Chapman & Hall/CRC.

Oller, K., & Eilers, R. E. (1975). Phonetic expectation and transcription validity. *Phonetica*, 31(3-4), 288-304. https://doi.org/10.1159/000259675.

Peters, J., Hanssen, J., & Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics*, 46, 185-209.

R Core Team (2016). “R: A language and environment for statistical computing” [Computer program]. Version 3.3.1, https://www.R-project.org/ (Last viewed June 21, 2016).

Shen, X. S. (1992). Mandarin neutral tone revisited. *Acta Llinguistica Hafniensia*, 24(1), 131-152. https://doi.org/10.1080/03740463.1992.10412273.

Tang, P. (2014). A study of prosodic errors of Chinese neutral tone by advanced Japanese students (Chinese version). *TCSOL Studies*, 56(4), 39-47.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017a). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(2), 493-503.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017b). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(5), 2823-2835.

Vienna, Austria: R Foundation for Statistical Computing; 2014. R Foundation for Statistical Computing.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83.

Yip, M. (1989). Contour tones. *Phonology*, 6(1), 149-174.

Yip, M. (2002). *Tone*. Cambridge University Press.

Yuen, I., Demuth, K., & Johnson, M. (2011). Prosodic structure in child speech planning and production. *In Proceedings from the 17th International Congress of Phonetic Sciences*. pp. 2248-2251.

Appendix

Appendix 1. Bonferroni adjusted post-hoc test of the **linear trend (slope)** of the f0 contour of neutral tone syllables between children and adults across word types (reduplicatives and possessives) and tonal contexts (neutral tone syllables after T1, T2, T3 and T4 syllables). Items in bold indicate significant findings.

Group difference	Type	Tonal context	β	SE	df	*t*	*p*	Cohen's d
3 yrs. - Adults	Reduplicative	T1	0.89	1.06	651	0.83	1.000	0.05
		T2	-0.88	1.07	659	-0.82	1.000	-0.05
		T3	-0.75	1.06	651	-0.70	1.000	-0.04
		T4	**-3.71**	**1.07**	**659**	**-3.48**	**<0.01****	**-0.19**
	Possessive	**T1**	**-2.98**	**1.07**	**659**	**-2.79**	**<0.05***	**-0.15**
		T2	-2.50	1.07	659	-2.34	0.117	-0.13
		T3	-1.03	1.07	659	-0.96	1.000	-0.05
		T4	-0.45	1.07	659	-0.42	1.000	-0.02
4 yrs. - Adults	Reduplicative	T1	1.33	1.08	651	1.24	1.000	0.07
		T2	0.18	1.08	651	0.16	1.000	0.01
		T3	-1.55	1.08	651	-1.44	0.902	-0.08
		T4	**-4.65**	**1.08**	**669**	**-4.29**	**<0.001*****	**-0.23**
	Possessive	T1	-2.49	1.08	659	-2.31	0.128	-0.13
		T2	**-3.04**	**1.09**	**688**	**-2.78**	**<0.05***	**-0.15**

		T3	-1.19	1.08	659	-1.10	1.000	-0.06
		T4	1.33	1.08	651	1.24	0.990	-0.08
5 yrs. - Adults	Reduplicative	T1	-1.10	1.27	651	-0.86	1.000	-0.05
		T2	-0.31	1.27	651	-0.24	1.000	-0.01
		T3	-0.65	1.27	651	-0.51	1.000	-0.03
		T4	-2.04	1.30	700	-1.57	0.702	-0.08
	Possessive	T1	-0.65	1.27	651	-0.51	1.000	-0.03
		T2	1.69	1.28	674	1.31	1.000	0.07
		T3	1.29	1.27	651	1.01	1.000	0.06
		T4	1.19	1.31	729	0.91	1.000	0.05
3 yrs. - 4 yrs.	Reduplicative	T1	-0.44	1.00	651	-0.44	1.000	-0.02
		T2	-1.06	1.00	660	-1.06	1.000	-0.06
		T3	0.80	1.00	651	0.80	1.000	0.04
		T4	0.94	1.01	681	0.93	1.000	0.05
	Possessive	T1	-0.49	1.01	670	-0.49	1.000	-0.03
		T2	0.54	1.02	704	0.53	1.000	0.03
		T3	0.16	1.01	670	0.16	1.000	0.01
		T4	1.06	1.01	681	1.05	1.000	0.06
3 yrs.	Reduplicative	T1	1.99	1.21	651	1.64	0.604	0.09

-		T2	-0.57	1.21	657	-0.47	1.000	-0.03
5 yrs.		T3	-0.10	1.21	651	-0.08	1.000	0.00
		T4	-1.67	1.24	712	-1.35	1.000	-0.07
	Possessive	**T1**	**-4.19**	**1.22**	**683**	**-3.43**	**<0.01****	**-0.19**
		T2	-2.33	1.21	657	-1.92	0.330	-0.11
		T3	-2.32	1.21	657	-1.91	0.337	-0.11
		T4	-1.64	1.25	744	-1.31	1.000	-0.07
4 yrs. - 5 yrs.	Reduplicative	T1	2.43	1.22	651	2.00	0.279	0.11
		T2	0.49	1.22	651	0.40	1.000	0.02
		T3	-0.90	1.22	651	-0.74	1.000	-0.04
		T4	-2.61	1.25	719	-2.09	0.223	-0.11
	Possessive	T1	-1.84	1.22	657	-1.51	0.794	-0.08
		T2	**-4.73**	**1.24**	**706**	**-3.80**	**<0.001*****	**-0.20**
		T3	-2.48	1.22	657	-2.03	0.257	-0.11
		T4	-2.69	1.27	751	-2.13	0.201	-0.11

Appendix 2. Bonferroni adjusted post-hoc test of the **quadratic trend (curvature)** of the pitch contour of neutral tone between groups across word types (the reduplicative and the possessive). Note. Items in bold indicate significant findings.

Group difference	Type	β	SE	df	*t*	*p*	Cohen's d
3yrs. - Adults	Reduplicative	-0.76	0.54	329	-1.40	0.983	-0.11
	Possessive	**-1.88**	**0.54**	**331**	**-3.46**	**<0.01****	**-0.27**
4yrs. - Adults	Reduplicative	-0.89	0.55	330	-1.62	0.638	-0.13
	Possessive	**-1.94**	**0.55**	**336**	**-3.52**	**<0.01****	**-0.27**
5yrs. - Adults	Reduplicative	-0.98	0.65	333	-1.51	0.799	-0.12
	Possessive	-0.36	0.65	339	-0.55	1.000	-0.04
3yrs. - 4yrs.	Reduplicative	0.13	0.51	332	0.26	1.000	0.02
	Possessive	0.06	0.51	342	0.12	1.000	0.01
3yrs. - 5yrs.	Reduplicative	0.22	0.62	335	0.36	1.000	0.03
	Possessive	-1.52	0.62	344	-2.44	0.091	-0.19
4yrs. - 5yrs.	Reduplicative	0.09	0.62	335	0.15	1.000	0.01
	Possessive	-1.58	0.63	347	-2.51	0.075	-0.19

Appendix 3. Bonferroni adjusted post-hoc test of the **duration** of neutral tone syllables between children and adults across word types (the reduplicative and the possessive) and tonal contexts (neutral tone after T1, T2, T3 and T4). Items in bold indicate significant findings.

Group difference	**Type**	**Tonal context**	**β**	**SE**	**df**	***t***	***p***	**Cohen's d**
3 yrs. - Adults	Reduplicative	T1	0.18	0.07	200	2.51	0.078	0.25
		T2	0.13	0.06	223	2.06	0.243	0.20
		T3	**0.31**	**0.09**	**176**	**3.54**	**<0.01****	**0.38**
		T4	**0.28**	**0.09**	**174**	**3.18**	**<0.05***	**0.34**
	Possessive	T1	0.10	0.07	203	1.52	0.781	0.15
		T2	**0.20**	**0.06**	**251**	**3.39**	**<0.01****	**0.30**
		T3	0.12	0.08	186	1.56	0.723	0.16
		T4	0.02	0.07	183	0.21	1.000	0.02
4 yrs. - Adults	Reduplicative	T1	0.11	0.07	200	1.52	0.782	0.15
		T2	0.13	0.06	222	1.96	0.304	0.19
		T3	**0.33**	**0.09**	**176**	**3.69**	**<0.01****	**0.39**
		T4	0.10	0.09	176	1.09	1.000	0.12
	Possessive	T1	0.09	0.07	202	1.29	1.000	0.13
		T2	0.15	0.06	257	2.51	0.076	0.22
		T3	0.13	0.08	186	1.59	0.678	0.17
		T4	0.04	0.08	185	0.54	1.000	0.06

5 yrs. - Adults	Reduplicative	T1	0.00	0.09	200	0.04	1.000	0.00
		T2	0.06	0.08	222	0.74	1.000	0.07
		T3	0.14	0.10	176	1.31	1.000	0.14
		T4	0.07	0.11	180	0.66	1.000	-0.07
	Possessive	T1	-0.05	0.08	201	-0.63	1.000	-0.06
		T2	0.00	0.07	254	0.01	1.000	0.00
		T3	0.09	0.09	184	0.93	1.000	0.10
		T4	-0.18	0.09	194	-1.93	0.330	-0.20
3 yrs. - 4 yrs.	Reduplicative	T1	0.07	0.07	200	1.04	1.000	0.10
		T2	0.01	0.06	224	0.09	1.000	0.01
		T3	-0.02	0.08	176	-0.21	1.000	-0.02
		T4	0.18	0.08	178	2.20	0.176	0.23
	Possessive	T1	0.01	0.06	204	0.23	1.000	0.02
		T2	0.05	0.06	260	0.86	1.000	0.08
		T3	0.00	0.07	188	-0.05	1.000	-0.01
		T4	-0.03	0.07	187	-0.36	1.000	-0.04
3 yrs. - 5 yrs.	Reduplicative	T1	0.18	0.08	200	2.25	0.155	0.22
		T2	0.07	0.07	223	1.04	1.000	0.10
		T3	0.17	0.10	176	1.74	0.506	0.19

		T4	**0.35**	**0.10**	**182**	**3.45**	**<0.01****	**0.36**
	Possessive	T1	0.15	0.08	202	2.01	0.277	0.20
		T2	**0.20**	**0.07**	**256**	**2.97**	**<0.05***	**0.26**
		T3	0.04	0.09	185	0.40	1.000	0.04
		T4	0.19	0.09	197	2.20	0.172	0.22
4 yrs. - 5 yrs.	Reduplicative	T1	0.11	0.08	200	1.38	1.000	0.14
		T2	0.07	0.07	222	0.96	1.000	0.09
		T3	0.19	0.10	176	1.89	0.362	0.20
		T4	0.17	0.10	183	1.63	0.624	0.17
	Possessive	T1	0.14	0.08	202	1.80	0.438	0.18
		T2	0.15	0.07	261	2.21	0.167	0.19
		T3	0.04	0.09	185	0.44	1.000	0.05
		T4	0.22	0.09	198	2.47	0.086	0.25

Chapter Five: The Acquisition of Phonological Alternations: the Case of the Mandarin Tone Sandhi Process

This chapter is based on the following paper:

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., Demuth, K. (revision submitted). The acquisition of phonological alternations: the case of the Mandarin tone sandhi process. *Applied Psycholinguistics.*

The paper has been modified according to reviewers' comments in the first round, and is now undergoing the second round of review process.

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors and collaborators) when needed.

Abstract

Phonological processes can pose a learning challenge for children, where the surface form for an underlying contrast may vary as a function of the phonological environment. Mandarin tone sandhi is a complex phonological process which requires knowledge about both the tonal and prosodic context in which it applies. The present study explored the *productive knowledge* of tone sandhi processes by 108 3-5-year-old Mandarin-speaking children and 33 adults. Participants were asked to produce novel tone sandhi compounds in different tonal contexts and prosodic structures. Acoustic analysis showed that 3-year-olds have abstracted the tone sandhi process and can productively apply it to novel *disyllabic* words across tonal contexts. However, mastering the application of tone sandhi in response to the *trisyllabic* contexts appears to be a protracted process not fully completed by the age of 5. The results are discussed in terms of the factors that influence how tone sandhi processes, and phonological alternations more generally, are acquired.

Key words: phonological acquisition, phonological alternation, phonological representations, tone sandhi, Mandarin Chinese

Introduction

During the process of acquiring language, children need to develop an understanding of how words are phonologically represented, and then store and retrieve these words from their mental lexicon during both perception and production. This is complicated by the presence of phonological processes that occur when words or morphemes come together, which can lead to phonological alternations (e.g., Kazazis, 1969; Zamuner, Kerkhoff & Fikkert, 2006; Kerkhoff, 2007; Albright & Hayes, 2011; van de Vijver & Baer-Henney, 2013). Phonological processes such as *final devoicing* in Dutch, for instance, prohibits final voiced obstruents. Thus, children learning Dutch have to learn that the word-final /t/ in [bɛt] 'bed' is the result of a final devoicing process, where the underlying lexical representation is actually /bed/, surfacing as such intervocalically in the plural form [bɛd+en] 'beds'. However, it has been shown that Dutch-learning children have difficulty learning such alternations, acquiring this process with novel words only after 4 years of age (Zamuner, Kerkhoff, & Fikkert, 2006; Kerkhoff, 2007).

While there is a large body of literature investigating how children learn various types of phonological processes, these investigations have tended to focus on processes involving vowels and consonants, mostly in Germanic and Romance languages. However, 60%-70% of human languages also involve tonal contrasts (Yip, 2002). This raises questions about whether children learning tonal languages encounter problems in acquiring tonal alternations in the same way as reported for segmental alternations.

One of the most well studied tonal processes includes the tone sandhi phenomenon, leading to tonal changes in specific tonal and prosodic contexts. This is part of a larger set of language-specific phonological constraints that prohibit similar/identical phonological features from being adjacent to each other, known as the Obligatory Contour Principle (OCP). The OCP was originally observed in African tone languages (Goldsmith 1976; Odden, 1986), and then adopted in studies of other tonal languages (Yip, 1989). Mandarin Chinese has four lexical tones as well as various tone sandhi processes. Among those most well studied is the tone 3 sandhi process, typically modifying the realization of the third tone (T3) when followed by another T3 within the same prosodic domain (Yip, 2002, p. 180). Thus, tone sandhi is a complex phonological process conditioned by both the tonal context and the prosodic structure in which it appears (Shih, 1997). The aim of the present study was thus to explore when and how children develop *productive* knowledge of the tone sandhi process, with correct implementation across tonal contexts and prosodic structures. The findings could have implications not only for the learning of tonal alternations, but for learning phonological alternations more generally.

Lexical tones

Mandarin Chinese has four lexical tones contrasting in pitch contour, i.e., Level tone 1 (T1), Rising tone 2 (T2), Dipping tone 3 (T3) and Falling tone 4 (T4; see Figure 1). Word meaning differs when carrying different tones, e.g., /ma1/ *mother*, /ma2/ *hemp*, /ma3/ *horse* and /ma4/ *scold*. These four tones, with associated word meanings, appear early in acquisition, during the single-word stage of development, although confusion between T2 (the rising tone) and T3 (the dipping tone) continues into the 2/3-word stage of development (Li & Thompson, 1977). By the age of 3, when children begin to combine words into longer sentences, all lexical tones have generally been acquired (Li & Thompson, 1977).

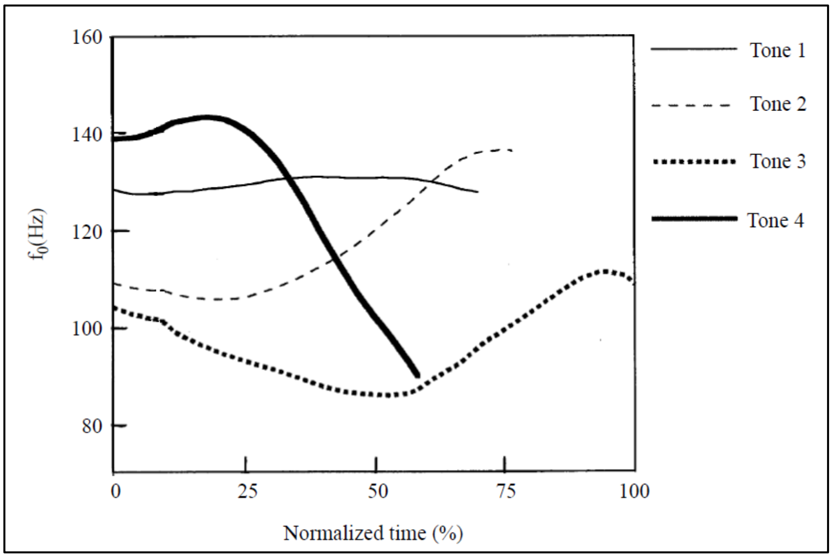


Figure 1. Mandarin Chinese Lexical tone pitch contours, from Xu (1997, p. 67).

Tone sandhi application as a function of tonal context in disyllabic words

Among the four lexical tones, T3 is unique in that it has highly context-conditioned phonetic realizations. T3 can only be realized in its citation/canonical form (a dipping tone) in isolation or in word final position. In the initial position of a **disyllabic word**, however, its realization is governed by the tone sandhi process, which alters the surface tone of T3 as a function of the following tonal context (Chen, 2000, p. 20). For example, in a T3T3 disyllabic word, the first T3 changes to T2, with a rising pitch; this phonological process is referred as the **full sandhi** process. In contrast, T3 is realized as a low-falling tone before T1/2/4, and this phonological process is referred as the **half sandhi** process. Both are exemplified in (1) and Figure 2. The tone sandhi process therefore complicates the phonological representation of T3, as it results in different surface tones in different tonal contexts, i.e., a dipping pitch in citation form, a rising pitch in the full sandhi context and a low-falling pitch in the half sandhi context.

(1) *Mandarin tone sandhi processes*

Full sandhi process: T3 → T2 (rising tone) / _ T3

Half sandhi process: T3 → low falling tone / _ T1/2/4

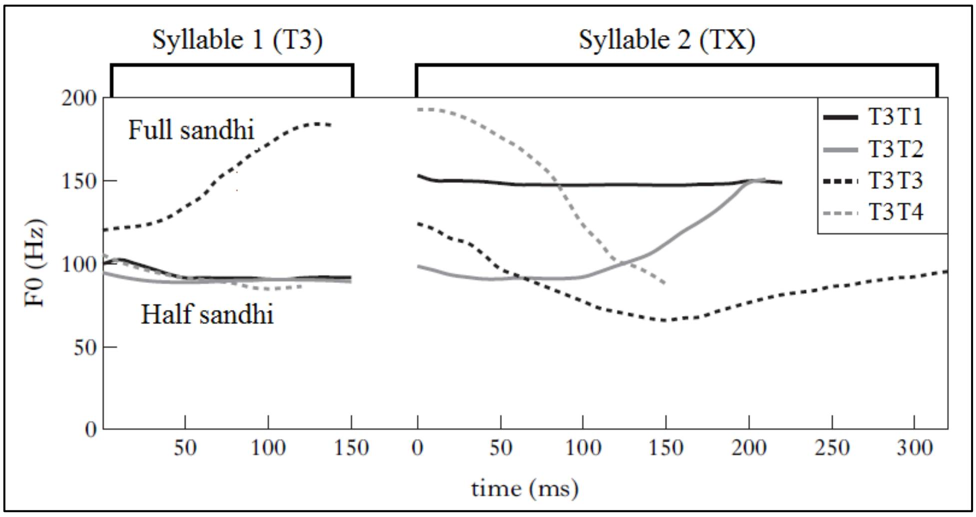
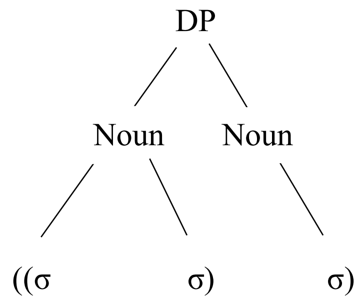


Figure 2. Mandarin Chinese tone sandhi pitch contours in disyllabic words (adapted from Zhang and Lai (2010, p. 163)).

Tone sandhi application as a function of prosodic structure in trisyllabic words

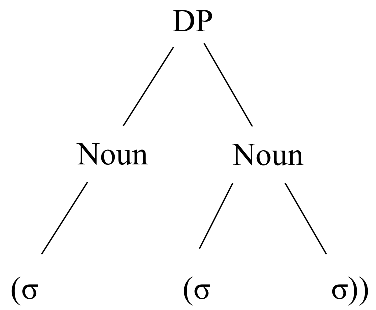
The application of tone sandhi is more complex for **trisyllabic** words. In a word containing **T3T3T3** as underlying tones, the full tone sandhi process may apply recursively, depending on the internal prosodic structure (Shih, 1997). Consider the two trisyllabic prosodic words: ((lao3 hu3) dan3) *tiger's gall* and (zhi3 (lao3 hu3)) *paper tiger*. The former involves a **left-branching** structure ((T3T3) T3) and the latter involves a **right-branching** structure (T3 (T3T3)), where 'left/right' refers to the position of the disyllabic prosodic unit (foot) where tone sandhi first applies within the larger trisyllabic prosodic word. The tone sandhi process thus applies first within the disyllabic unit [σσ], changing its tones from T3T3 to T2T3. With the innermost bracket erased, the tone sandhi process then applies again, within the higher prosodic domain of the entire trisyllabic prosodic word. If there are still two adjacent T3T3 syllables (as in the left-branching case in (2)), the tone sandhi process applies again. As a result of this recursive tone sandhi process, the left-branching structure will generate **T2T2**T3 as the surface output form (where the bold font indicates the result of tone sandhi application). In contrast, the right-branching structure in (3) surfaces with T3**T2**T3 output (Shih, 1997).

(2) A left-branching trisyllabic prosodic word ((T3T3) T3).



((老	虎)	胆)	Chinese characters: 老虎 *tiger*; 胆 *gall*
((lao3	hu3)	dan3)	Chinese pinyin
((T3	T3)	T3)	Underlying tone
((**T2**	T3)	T3)	Tone sandhi applies within the disyllabic word
(T2	**T2**	T3)	Boundary erased; tone sandhi applies again; surface tone

(3) A right-branching trisyllabic prosodic word (T3 (T3T3)).



(纸	(老	虎))	Chinese characters: 纸 *paper*; 老虎 *tiger*
(zhi3	(lao3	hu3))	Chinese pinyin
(T3	(T3	T3))	Underlying tone
(T3	(**T2**	T3))	Tone sandhi applies within the disyllabic word
(T3	T2	T3)	Boundary erased, no additional process; surface tone
(**T2**	**T2**	T3)	In fast speech, larger unit built, optional surface tone

However, in the case of the structure in (3), it has also been suggested that, especially in fast speech, the tone sandhi process applies recursively from left to right within the entire trisyllabic word, regardless of the internal prosodic structure. This then results in identical **T2T2**T3 surface forms for both the left- and right-branching prosodic words in (2) and (3) respectively (Chen, 2000, p. 379). Thus, for the right-branching structure there may be some variable surface realizations in the input children hear. However, there has yet to be any acoustic investigation of this phenomenon to verify a possible speaking rate explanation for such variable surface realizations.

Previous studies of tone sandhi acquisition

Studies have often reported the acquisition of phonological processes/alternations to be protracted (e.g., Kazazis, 1969; Zamuner, Kerkhoff & Fikkert, 2006; Kerkhoff, 2007; Albright & Hayes, 2011; van de & Baer-Henney, 2013), where learners need to compare morphologically related forms, choosing a basic or an underlying form/representation, and then abstracting a grammatical structure that can generate the various surface forms (Albright & Hayes, 2011). This could also be the case for Mandarin-speaking children when learning the tone sandhi process, as the various acoustic realizations across tonal contexts (i.e., full and half sandhi contexts) may complicate learning how and where it applies.

Although Mandarin tone sandhi processes have been well-studied in adults, they have received less attention with respect to when and how they are acquired by children. A few studies have examined children's production of **lexicalized disyllabic** tone sandhi words such as /ʂou3 tʂi3/ *finger*, revealing the early emergence of appropriate tone realization on these words before age 3 (e.g., Li & Thompson, 1977; Hua & Dodd, 2000; Xu Rattanasone, et al., 2018). Evidence from a perception study also suggests that children at age 3 are sensitive to tone sandhi mispronunciations on lexicalized tone sandhi words (Wewalaarachchi & Singh, 2016). However, "wug" tests with *novel* words (Berko, 1958) investigating productive mastery of phonological process often revealed errors even when children can correctly produce alternations on *real* words (Zamuner, Kerkhof & Fikkert, 2006; van de Vijver & Baer-Henney, 2013). This raises the question of whether 3-year-olds can generalize the tone sandhi process, productively applying it to novel items they have never heard before**.**

Note that application of tone sandhi in the T3T3T3 trisyllabic prosodic words requires children to build either a left-branching or a right-branching prosodic word structure. Evidence from non-tonal languages suggests the acquisition of adult-like prosodic knowledge is also a protracted process (e.g., Vogel & Raimy, 2001; Wells, Peppé & Goulandris, 2004; Yuen, et al., 2014). This suggests that Mandarin-learning children might also face problems in building different prosodic word structures to guide tone sandhi application. Moreover, from the perspective of the *variational learning model*, variability in the input may cause children to assume a different grammar from that of the adult, leading to later acquisition (Yang, 2002). Given that right-branching **T3T3T3** prosodic words can be realized with two possible surface forms (T3**T2**T3 ~ **T2T2**T3; Chen, 2000, p. 379), we might expect that adult-like tonal representations for the right-branching structure will take longer to acquire compared to the left-branching forms, where no such variability is found.

To our knowledge, there is only one study eliciting 3-5-year-olds' **trisyllabic** tone sandhi productions. In her thesis, Wang (2011; Chapter 6) examined children and adult controls' productions of **T3T3T3 noun + noun** compounds with two structures, i.e., ((T3T3) T3) vs. (T3 (T3T3)), such as ((shui3 guo3) niao3) *fruit-bird* and (shui3 (lao3 hu3)) *water-tiger*. Based on the author's perceptual transcriptions, it was observed that children at 3 years have adult-like tone sandhi productions for both structures, producing **T2T2**T3 for ((T3T3) T3) prosodic word structures, and T3**T2**T3 for (T3 (T3T3)) prosodic word structures. However, the prosodic structure of the stimuli used in her study was confounded with the word structure, where the child could easily parse a T3T3T3 as a familiar monosyllabic word plus familiar disyllabic word, leading to high accuracy in tone sandhi production. Thus, the interaction between children's tone sandhi application and prosodic structure has yet to be fully tested.

In summary, it is unclear when children have abstracted the tone sandhi process to be able to productively apply it to novel compounds in response to different tonal contexts and prosodic structures. This issue was addressed in the present study by conducting a series of "wug" tests. We first elicited **novel disyllabic** tone sandhi compound words across tonal contexts, testing children's productive knowledge of the tone sandhi process in these prosodically simple contexts. We then elicited **novel trisyllabic** T3T3T3 tone sandhi compound words in two different prosodic structures (left-branching vs, right-branching prosodic words), testing children's productive knowledge of tone sandhi application as a function of prosodic structure. In this case, all trisyllabic compounds were composed of 3 monosyllabic words to eliminate any potential bias of lexicalized/known disyllabic word structures.

We addressed two questions: (1) can 3-5-year-olds productively apply the tone sandhi process to **novel disyllabic compounds**? (2) If so, can they also extend this to **novel trisyllabic compounds**, showing sensitivity to prosodic context? Two main hypotheses were formulated:

Hypothesis 1 (H1) – **Disyllabic Words**:

As phonological processes can be challenging to acquire, we predicted that (a) the younger children (3-year-olds) might not *productively* use the tone sandhi process with *novel* disyllabic compounds, and that (b) children's ability to apply the tone sandhi process to novel disyllabic compounds would become acoustically more adult-like with age.

Hypothesis 2 (H2) – **Trisyllabic Words**:

As the trisyllabic T3T3T3 words involve *recursive* sandhi application in the left-branching structure, but *single* (or variably *recursive*) sandhi application in the right-branching structure, it might be more challenging to learn. We therefore predicted that (a) the younger children (3-year-olds) might not correctly apply the tone sandhi process to novel trisyllabic T3T3T3 compounds – especially for the more variably realized right-branching prosodic words, and that (b) children's ability to apply the tone sandhi process to novel trisyllabic compounds would become acoustically more adult-like with age.

Method

Participants

One-hundred and eight children aged 3, 4, and 5 years were recruited and tested in the affiliated kindergarten of Beijing Language and Culture University (see Table 1). All spoke Mandarin as their first language and were born and raised in Beijing. According to reports from the kindergarten, the recruited children did not have any speech, hearing, language or intellectual difficulty. There were fewer 5-year-olds than the other two age groups since many had already graduated from the kindergarten at the time of testing. In addition, 33 adult university students were recruited as controls. The adults were all native monolingual speakers of Mandarin and were born and raised in Beijing as well. The study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Panel.

Table 1. Number of participants in each age group

Age group	Male	Female	Total
3 yrs. (3;1-3;12, mean: 3;8, SD: 3 months)	13	31	44
4 yrs. (4;1-4;12, mean: 4;5, SD: 3 months)	20	22	42
5 yrs. (5;1-6;2, mean: 5;7, SD: 5 months)	9	13	22
Adults (19-25 yrs., mean: 20, SD: 2 years)	9	24	33

Stimuli

A total of 32 words were selected as stimuli, including eight monosyllabic TX (T1-4) words, 16 disyllabic T3TX novel compounds and eight novel trisyllabic T3T3TX compounds. All were picturable and associated with a line drawing. The eight monosyllabic words and 16 disyllabic compounds were used to address H1 and test children's productive use of the full and half tone sandhi process in various tonal contexts. The eight trisyllabic T3T3TX compounds, including six T3T3T1/2/4 controls and two T3T3T3 targets with left-branching and right-branching prosodic structures, were used to address H2 and investigate children's ability to apply the tone sandhi process to novel compounds with different prosodic structures.

Eight monosyllabic stimuli were used to compose the disyllabic compounds. The four first syllable words were all T3 animal names, and the second syllable words all referred to inanimate objects, with tones varying from T1 to T4 (see Table 2). All words fell within the top 50% of the most frequent monosyllabic words in language input to children below 3 years, according to the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018) from the CHILDES database (MacWhinney, 2000).

Table 2. Monosyllabic stimuli list

Group	**Target word**	**Tones**	**Meaning**	**Phonetic transcription**
a	马	T3	*horse*	/ma3/
	鸟	T3	*bird*	/niau3/
	狗	T3	*dog*	/kou3/
	鼠	T3	*mouse*	/ʂu3/
b	书	T1	*book*	/ʂu1/
	球	T2	*ball*	/tɕiu2/
	鼓	T3	*drum*	/ku3/
	画	T4	*picture*	/hua4/

Sixteen T3TX disyllabic compounds were then generated using these known monosyllabic words, resulting in four disyllabic full sandhi (T3T3) compounds and 12 disyllabic half sandhi (T3T1/2/4) compounds (see Table 3).

Table 3. Disyllabic stimuli list

Context	Compounds	Underlying Tones	Surface/Target Tones	Meaning	Phonetic transcription
Full sandhi	马鼓	T3T3	**T2**T3	*horse-drum*	/**ma3** ku3/
	鸟鼓	T3T3	**T2**T3	*bird-drum*	/**niau3** ku3/
	狗鼓	T3T3	**T2**T3	*dog-drum*	/**kou3** ku3/
	鼠鼓	T3T3	**T2**T3	*mouse-drum*	/**ʂu3** ku3/
Half sandhi	马书	T3T1	**Low-falling**T1	*horse-book*	/**ma3** ʂu1/
	马球	T3T2	**Low -falling**T2	*horse-ball*	/**ma3** tɕiu2/
	马画	T3T4	**Low -falling**T4	*horse-painting*	/**ma3** hua4/
	鸟书	T3T1	**Low -falling**T1	*bird-book*	/**niau3** ʂu1/
	鸟球	T3T2	**Low -falling**T2	*bird-ball*	/**niau3** tɕiu2/

鸟画	T3T4	**Low -falling**T4	*bird-painting*	/**niau3** hua4/
狗书	T3T1	**Low -falling**T1	*dog-book*	/**kou3** ʂu1/
狗球	T3T2	**Low -falling**T2	*dog-ball*	/**kou3** tɕiu2/
狗画	T3T4	**Low -falling**T4	*dog-painting*	/**kou3** hua4/
鼠书	T3T1	**Low -falling**T1	*mouse-book*	/**ʂu3** ʂu1/
鼠球	T3T2	**Low -falling**T2	*mouse-ball*	/**ʂu3** tɕiu2/
鼠画	T3T4	**Low -falling**T4	*mouse-painting*	/**ʂu3** hua4/

Eight trisyllabic compounds were also constructed, four with left-branching structure ((T3T3) TX) and four with right-branching structure (T3 (T3TX)) (see Table 4). In the *left-branching* structures, the same initial two monosyllabic word sequence (/tsi3 ma3/ *purple-horse*) was combined with each of the monosyllabic words (TX) from the monosyllabic word set, e.g. ((zi3 ma3) gu3) *purple-horse drum*) (Table 2b). Similarly, in the *right-branching* structures, the leftmost monosyllabic word was kept constant as /tsi3/ *purple* to combine with a disyllabic unit composed of two monosyllabic words in a T3TX sequence, e.g. ((zi3 (ma3 gu3)) *purple horse-drum*[14]. The left-branching structure contained **Noun + Noun** items ("purple-horse" + "drum"), whereas the right-branching structure contained **Adjective + Noun** items ("purple" + "horse-drum"). In total, there were six T3T3T1/2/4 compounds and two T3T3T3 compounds. The six T3T3T1/2/4 compounds were used as *controls* since the tone sandhi process is only applied on the first T3 syllable of these trisyllabic novel compounds. The two T3T3T3 novel compounds were thus the *target items* used to test participants' sensitivity to applying tone sandhi as a function of different prosodic structures. Following Shih (1997), the surface tones of the six control items and the two target items are exemplified in Table 4.

[14] One of the anonymous reviewers points out that there is only one test item in the critical initial position, i.e., purple. However, it is difficult to find multiple T3 test items for this position that 3-year-olds are familiar with. Adjectives more frequently appear in this position compared to nouns, especially for trisyllabic words. 'Purple' is a T3 adjective this is highly frequent in children's language input, based on the Mandarin corpora (the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018)) from the CHILDES database. Other highly frequent T3 adjectives such as "xiao3" *small* would induce a relative contrast effect, requiring 2 items to be presented on the screen, complicating the display/procedure.

Table 4. Trisyllabic stimuli list.

Groups	Compounds	Underlying tones/structures	Surface/Target tones	Meaning	Phonetic transcription
T3T3T1/2/4 (Controls)	((紫马) 书)	((T3T3) T1)	**T2**T3T1	*((purple horse) book)*	/tsi3 ma3 ʂu1/
	((紫马) 球)	((T3T3) T2)	**T2**T3T2	*((purple horse) ball)*	/tsi3 ma3 tɕiu2/
	((紫马) 画)	((T3T3) T4)	**T2**T3T4	*((purple horse) painting)*	/tsi3 ma3 hua4/
	(紫 (马书))	(T3 (T3T1))	**T2**T3T1	*(purple (horse book))*	/tsi3 ma3 ʂu1/
	(紫 (马球))	(T3 (T3T2))	**T2**T3T2	*(purple (horse ball))*	/tsi3 ma3 tɕiu2/
	(紫 (马画))	(T3 (T3T4))	**T2**T3T4	*(purple (horse painting))*	/tsi3 ma3 hua4/
T3T3T3 (Targets)	((紫马) 鼓)	((T3T3) T3)	**T2T2**T3	*((purple horse) drum)*	/tsi3 ma3 ku3/
	(紫 (马鼓))	(T3 (T3T3))	T3**T2**T3~**T2T2**T3	*(purple (horse drum))*	/tsi3 ma3 ku3/

All trisyllabic novel compounds were picturable, with prosodic structure constructed via eliciting different novel compounds. For example, the *left-branching* target item ((zi3 ma3) gu3) *purple-horse drum* was depicted by a drum with a purple horse on it, while the *right-branching* target item (zi3 (ma3 gu3)) *purple horse-drum* was depicted by a horse-drum (a drum with a horse on it) which was then coloured purple, resulting in the final representations depicted in Figure 3.

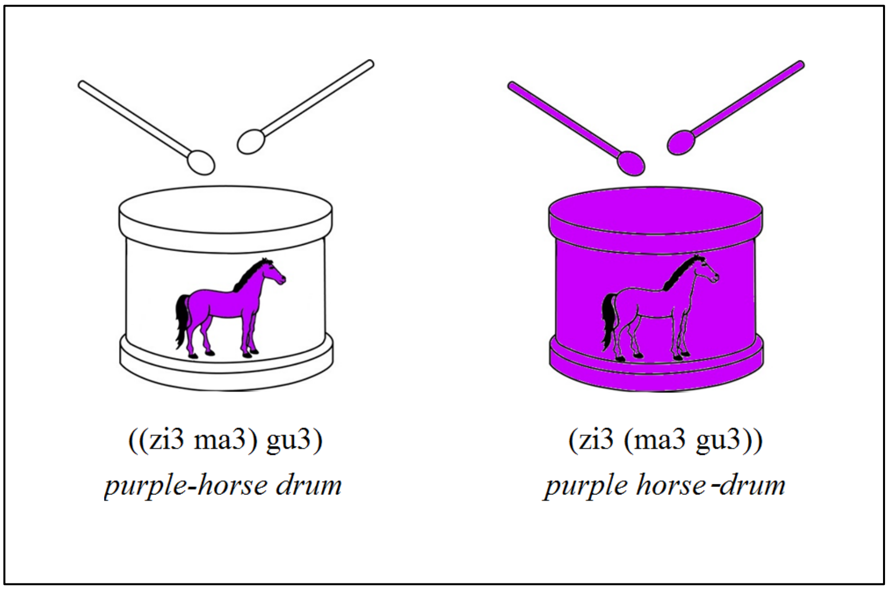


Figure 3. Two pictures used to elicit the *left-branching* trisyllabic item ((zi3 ma3) gu3) *purple-horse drum* and the *right-branching* trisyllabic item (zi3 (ma3 gu3)) *purple horse-drum*.

Procedure

Children were tested in a quiet room at the kindergarten; adults were tested in a quiet room at the university. During the test session, each participant wore a head-mounted cardioid-directional condenser microphone (AKG-C520) which was connected to a solid-state recorder (Marantz PMD661MKII) in order to capture their word productions.

The experiment had three parts, eliciting monosyllabic, disyllabic and trisyllabic items respectively, each with an initial familiarization/practice phase. In Part 1, monosyllabic words were elicited in a picture-naming task. Prior to testing, there was a familiarization phase during which the Mandarin-speaking experimenter produced the names of all monosyllabic words in Table 2, including the two practice items, i.e., /mau1/ *cat* and /jaŋ2/ *sheep*. The experimenter first presented a picture of an item on the computer screen and asked the participant to name the picture. If the participant failed to name it or produced it with a different word, the experimenter would correct it and ask the child to name it again. Once the participant was able to correctly produce the name of all the pictured words, the experimenter proceeded to the test phase. In the test phase, the experimenter showed pictures of all the objects one by one and asked the participant to name the pictures using the words produced during the familiarization phase. The order of items was randomized across participants. These data provided both a tonal and lexicalized item control for assessing participants' knowledge of tone sandhi processes with novel word combinations in Parts 2 and 3.

In Part 2, the sixteen disyllabic compounds were elicited in a novel word-compounding game. Four novel practice trials (with feedback, i.e., the correct target word) were first used to familiarise participants with the procedure for forming novel compounds: /mau1 ʂu1/ *cat-book*, /mau1 tɕʰiu2/ *cat-ball*, /jaŋ2 ku3/ *sheep-drum* and /jaŋ2 hua4/ *sheep-painting*. This was identical to the procedure then used for the test trials where, for example, to elicit the disyllabic compound /ma3 ku3/ *horse-drum*, a picture of a horse and a picture of a drum were displayed on the left and the right side of a robot-like cartoon Figure respectively (see step 1 in Figure 4). The experimenter then pressed a button to play a pre-programmed animation where the two pictures disappeared behind the robot and it jiggled to output a novel object: a horse-drum (see step 2 in Figure 4). The experimenter then asked the participant to produce the name of the resulting item. The order of the test trials was randomized across participants.

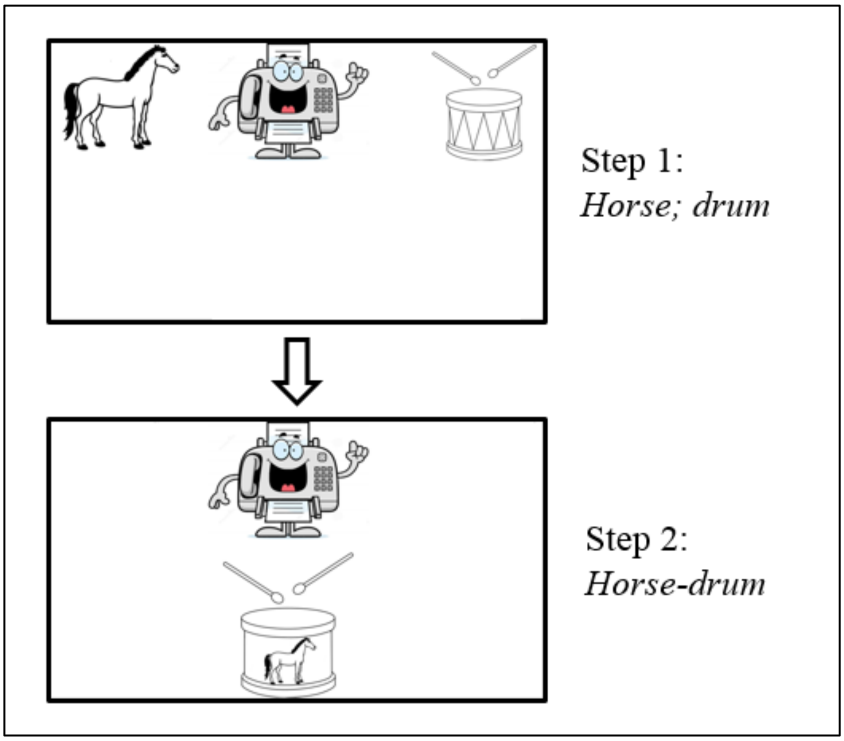


Figure 4. To elicit the compound "horse-drum", a horse and a drum were first presented on either side of the robot. The horse and the drum were then combined into a new item "horse-drum" by the robot.

In Part 3, the eight trisyllabic compounds were elicited using a similar procedure. Compounds with left-branching and right-branching structures were elicited in two separate blocks to minimize the processing load for young children. Both blocks began with a non-sandhi practice trial, (e.g., ((hong2 mao1) shu1) *red-cat book* or (hong2 (mao1 shu1)) *red cat-book*), followed by four test trials. During practice, feedback was provided when the child used a non-target word in forming the novel compounds. For instance, instead of using the target word 'hong2' (red), some children produced ‘hong2 se4’ (red-coloured) for “(hong2 (mao1 shu1))” *red cat-book*, or “(hong2 se4 de0 (mao1 shu1))” *red-colour cat-book*. In this case, the experimenter would provide the target word ‘hong2’ and prompt the child to use it to produce the novel compound again. This was done to ensure that the child understood the task as no feedback was provided for the test items.

Thus, to elicit the *left-branching* target item ((zi3 ma3) gu3) *purple-horse drum*, for instance, a purple-horse and a drum appeared on opposite sides of the robot (step 1, Figure 5), and the experimenter asked the participant to name both items. Next, the experimenter pressed a button to play the same animation of the robot (jiggling) to combine the purple-horse and the drum into a new object, i.e., a purple-horse drum (step 2, Figure 5). The experimenter then asked the participant to name the novel compound.

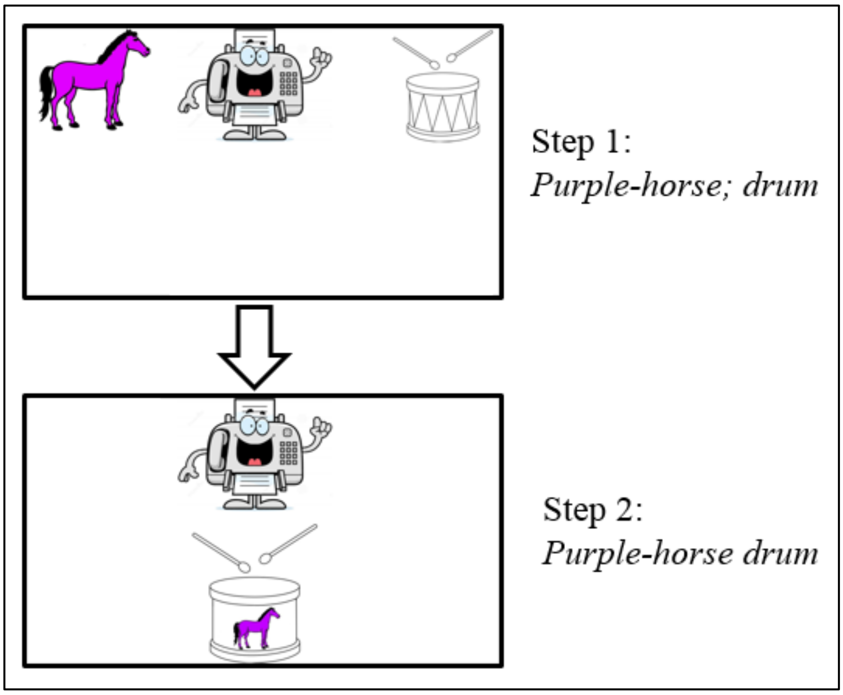


Figure 5. To elicit the compound “purple-horse drum”, a purple horse and a drum were firstly presented on opposite sides of the robot. The purple-horse and the drum were then combined into a new item “purple-horse drum” by the robot.

A similar procedure was used to elicit the *right-branching* item (zi3 (ma3 gu3)) *purple horse-drum*. First a “horse-drum” was generated, using identical steps to those for the left-branching items (steps 1 and 2, Figure 6). The participant was asked to produce its name, as it also served as the input for the trisyllabic word. Step 3 was then added to render the novel object purple (step 3, Figure 6) so as to elicit the novel compound “purple horse-drum”. The order of the two blocks and the order of testing trials within each block were randomized across participants. The order of the two blocks and the order of testing trials within each block were also randomized across participants.

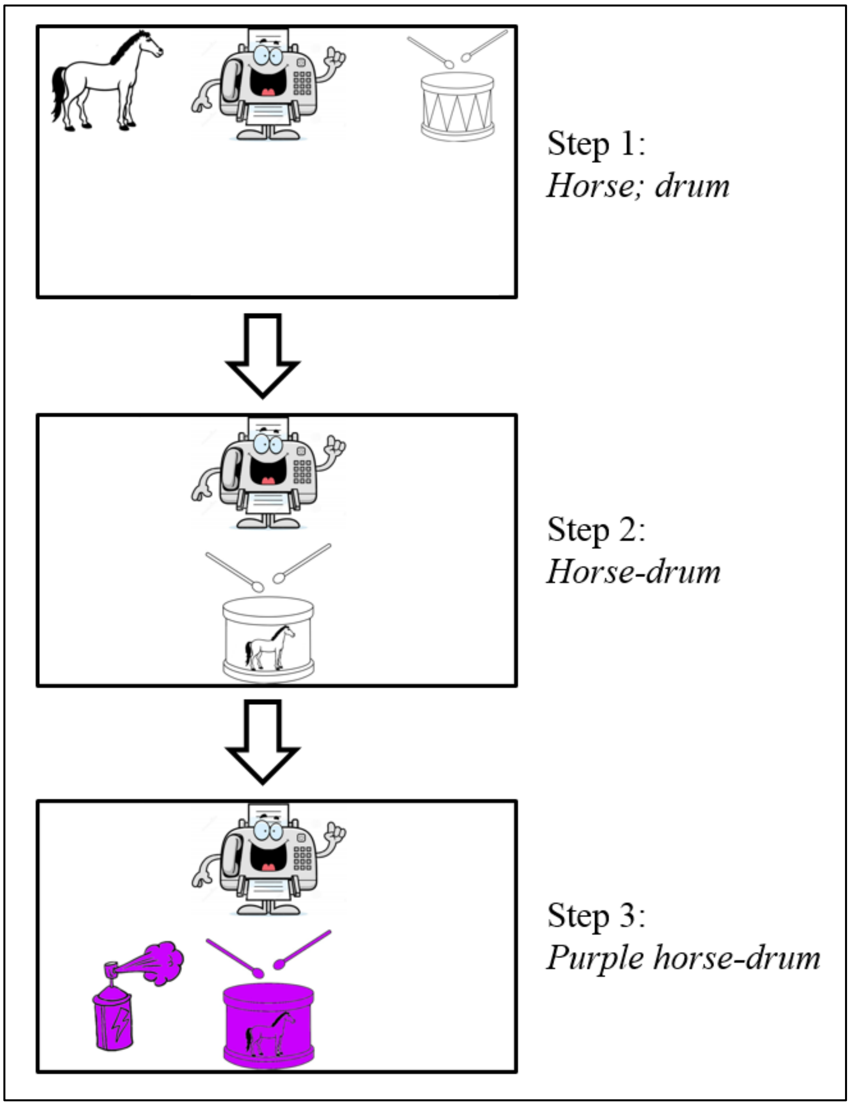


Figure 6. To elicit the compound “purple horse-drum”, a horse and a drum were firstly presented on opposite sides of the robot. The horse and the drum were then combined into a new item ‘horse-drum’. A spray-paint appeared and coloured the horse-drum purple to create a new object ‘purple horse-drum’.

Coding and measurements

Participants’ productions were acoustically coded in Praat (Boersma & Weenink, 2016). The vowel of each syllable in the compound was identified and annotated in terms of the onset and offset of the second formant (F2). Pitch tracks of annotated vowels were checked and manually revised to correct for any ‘doubling’ or ‘halving’ errors. The revised pitch track was then interpolated and smoothed with a bandwidth of 20 Hz. The f0 values were extracted at 10 equidistant points within each annotated vowel using the default autocorrelation algorithm in Praat (Boersma, 1993). The extracted f0 values were then converted from Hz to semitones with a reference of 50 Hz). Ten percent of the items were re-coded by a second trained native speaker of Mandarin. Inter-rater reliability on the extracted pitch points was good ($r = 0.94$). To minimize possible inter-speaker variation in pitch, the f0 values of each token were normalized against the mean pitch across all tokens for each individual speaker.

Statistical analysis

A total of 4444 tokens were included in the data analysis (1377 tokens from 3-year-olds, 1313 tokens from 4-year-olds, 698 tokens from 5-year-olds and 1056 tokens from adults). An additional 68 tokens (2% exclusion rate: 31 tokens from 3-year-olds, 31 tokens from 4-year-olds and 6 tokens from 5-year-olds) were excluded from the analysis due to poor acoustic quality, including environmental noise, whispered speech or productions with prolonged creak.

The data were analysed using R (R Core Team, 2016). To quantify the pitch contour of child and adult productions across conditions, a second-order orthogonal polynomial equation was fitted for each tone production, using the *poly* function of R. The second-order polynomials were adopted since the most complex pitch contour of tones in our data had only a convex or concave contour shape. The polynomial function generated three parameters for each pitch contour, i.e., (1) the intercept, (2) the linear trend and (3) the quadratic trend. According to Mirman (2016), the three parameters capture the f0 contour overall *height* (as reflected in the intercept: The higher the intercept, the higher the pitch), *direction* (slope as reflected in the linear trend: a positive linear trend indicates a rising pitch and vice versa; a larger absolute value of the linear trend represents a steeper pitch and vice versa) and *curvature* (as reflected in the quadratic trend: a positive quadratic trend indicates a concave f0 contour, a negative quadratic trend indicates a convex contour; and a larger quadratic trend indicates a more curved f0 contour and vice versa). These parameters were used to evaluate any group differences in the overall f0 contour.

Linear mixed regression models were built to compare the tone productions between children and adults, using the LME4 package (Bates, Mächler, Bolker, & Walker, 2014). All random slopes were included in the model to make it maximally generalizable across the data (Barr, Levy, Scheepers, & Tily, 2013). The *anova* function, which provides Satterthwaite's approximation to degrees of freedom for estimating *p-values* using F-statistics in R package LMERTEST (Kuznetsova, Brockhoff, & Christensen, 2015), was used to test for statistical significance. The main and interaction effects reported in the results were averaged across all levels of the other effects (see Luke, 2017 for a detailed explanation and Peters, Hanssen, & Gussenhoven, 2014 for a practical example). When a significant main effect of a multi-level factor or a significant interaction effect was observed, Tukey-HSD post-hoc comparisons were performed on the multi-level factor, as well as interactions, using LSMEANS package (Lenth, 2016). The parameter estimates are provided in the results of post-hoc comparisons rather than in the results of the main model.

Results

Tone sandhi in disyllabic words

In this section we address H1: (a) young children (i.e., 3-year-olds) might not productively apply the tone sandhi process to novel disyllabic compounds; (b) children's ability to apply the tone sandhi process to novel disyllabic compounds will become more acoustically adult-like with age. We first examined children's lexical/citation tone productions to make sure that they had correct tonal representations for all four lexical tones, and then examined their disyllabic tone sandhi productions.

Lexical tones

Figure 7 illustrates the normalized lexical tone pitch contours for the three child groups compared to adults. All child groups showed very similar lexical tone pitch patterns, and were similar to the adults, with a level pitch for T1, a rising pitch for T2, a dipping pitch for T3 and a falling pitch for T4 (Figure 7). A linear mixed regression model was built using the 3 pitch curve parameters (intercept, slope and curvature) across 10 time points to explore any group differences in tone contour. Two fixed factors "Group" (3-, 4-, 5-year-olds and adults) and "Tone" (T1, T2, T3 and T4) and the random factor "Participant" were entered into the model. Since we expected all children to have acquired all four lexical tone categories, in line with previous studies (e.g., Li & Thompson, 1977; Hua & Dodd, 2000), we predicted that the model would predict no effect of "Group" on the pitch shape, i.e., the pitch slope (linear trend) and the pitch curvature (quadratic trend).

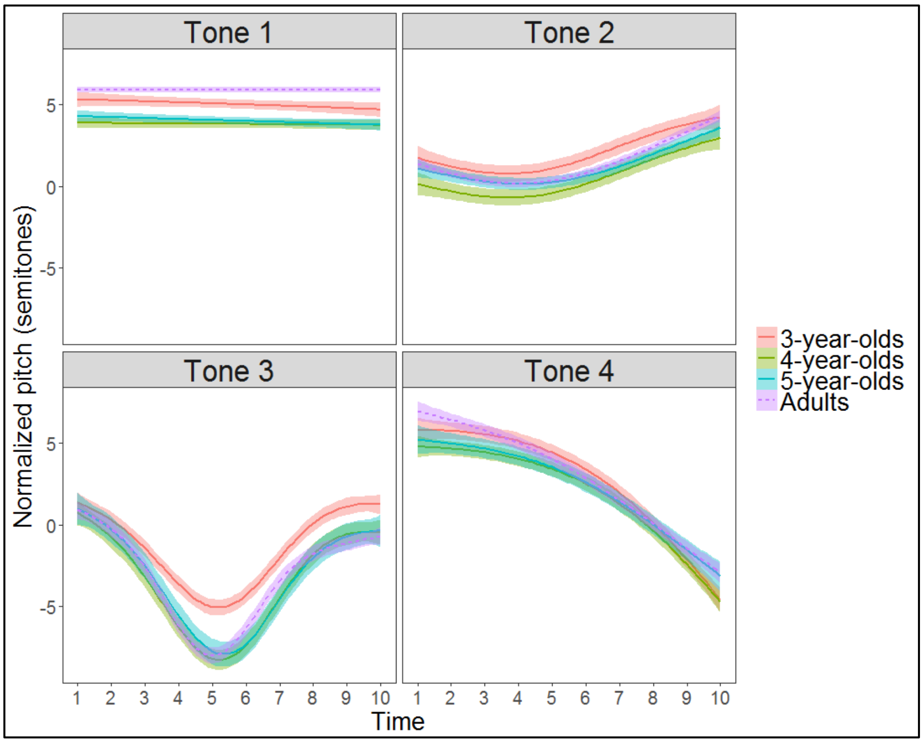


Figure 7. Pitch contour of lexical tone productions from children (3-, 4- and 5-year-olds) and adults.

The results are presented in Appendix 1. As expected, neither the main effect of "Group" nor the interaction between "Group" and "Tone" was significant on the slope or the curvature, indicating that the pitch contours of children's lexical tone productions were not significantly different from those of adults[15]. This corroborates previous findings showing that, by 3 years, children have established appropriate representations for Mandarin lexical tones (e.g., Li & Thompson, 1977; Hua & Dodd, 2000).

[15] Note that the effect of "Group" on the pitch intercept reflects the predictable group difference on overall pitch height, i.e., younger children typically have higher overall pitch than older children and adults, and was therefore not further explored in the current or following analysis.

Disyllabic tone sandhi compounds

We now turn to the results of the disyllabic novel compounds. Figure 8 shows 3-, 4- and 5-year-olds' and adults' tone sandhi productions across the two types of tonal contexts: one triggering full sandhi (T3T3), and the others triggering half sandhi (T3T1/2/4). Visual inspection of Figure 8 suggests that children from all three age groups have correctly applied the tone sandhi process to novel disyllabic compounds across both types of tonal context. That is, they changed the tone of the first syllable (T3) to a rising tone (T2) in the full sandhi context, and changed it to a low falling tone in the three half sandhi contexts.

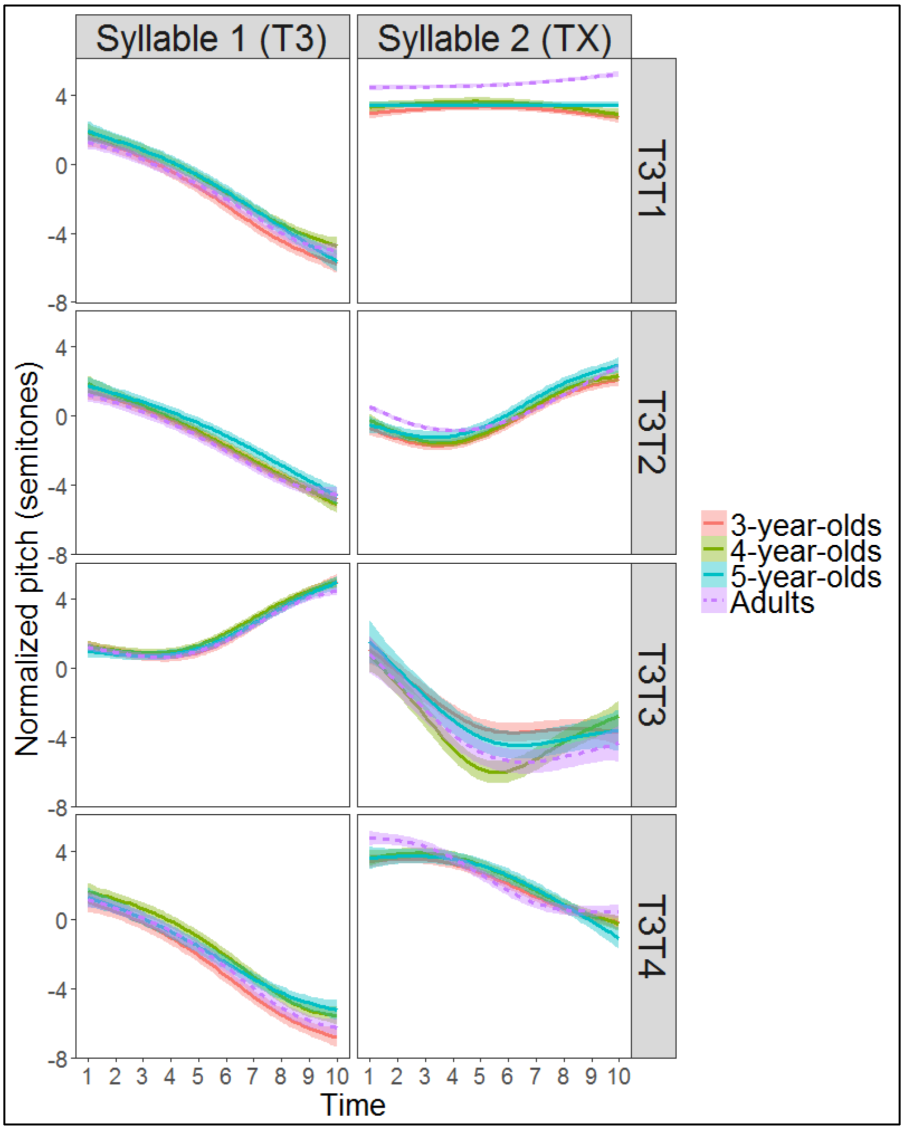


Figure 8. Pitch contours of T3TX productions from children (3-, 4- and 5-year-olds) and adults, where T3T3 items result in full sandhi compounds and T3T1/2/4 items result in half sandhi compounds.

We then conducted a linear mixed regression model to look for any group differences in sandhi production. The comparison focused on the first T3 syllable of the disyllabic compound, since this is where the tone sandhi change takes place. Two fixed factors "Group" (3-, 4-, 5-year-olds and adults) and "Type" (full and half sandhi), with the random factor "Participant" were included in the model. According to H1 (a), if young children did not productively apply the tone sandhi process, the model would predict a main effect of "Group" and an interaction between "Group" and "Type" on the pitch shape (slope and curvature). According to H1 (b), if children's ability to apply the tone sandhi process becomes more adult-like as they get older, a post-hoc test on the main effect of "Group" would reveal a smaller pitch difference between older children and adults compared to that between young children and adults.

The results are presented in Appendix 2. Neither the main effect of "Group" nor the interaction between "Group" and "Type" was significant for pitch shape (slope and curvature), suggesting that children's sandhi productions were not significantly different from adults for either full or half sandhi syllables. Moreover, the significant main effect of "Type" suggests that, for both child and adult groups, the pitch contour was different between full and half sandhi syllables. A post-hoc test of "Type" showed that, relative to half sandhi syllables, the pitch contour of full sandhi syllables was more rising and curved (slope: $\beta = 11.6$, SE = 0.16, t (22084) = 71.38, $p < 0.001$; curvature: $\beta = 2.25$, SE = 0.16, t (22082) = 13.81, $p < 0.001$), consistent with the acoustic feature of a rising T2. Thus, children from all age groups were able to productively apply tone sandhi processes to novel disyllabic compounds in both full and half sandhi tonal contexts.

Tone sandhi in trisyllabic words

In this section we address H2: (a) 3-year-olds might not apply the full sandhi process to novel trisyllabic (T3T3T3) compounds, and (b) children's tone sandhi application to novel trisyllabic compounds would become acoustically more adult-like with age. We first examined children's and adults' productions of trisyllabic control items (underlying T3T3T1/2/4), where the tone sandhi process does not interact with the prosodic structure. We then examined their tone sandhi application on the target items (underlying T3T3T3), to test children's ability to build different prosodic structures to guide recursive tone sandhi application.

Control items: T3T3T1/2/4 compounds

Figure 9 shows children's and adults' productions of trisyllabic tone sandhi control words, i.e., ((T3T3) T1/2/4) vs. (T3 (T3T1/2/4)), where the tone sandhi process only applies to the first syllable of the T3T3 sequence. The surface tones should thus be **T2**T3T1/2/4 for both structures. Visual inspection of Figure 9 showed that all groups produced the correct surface tones across structures.

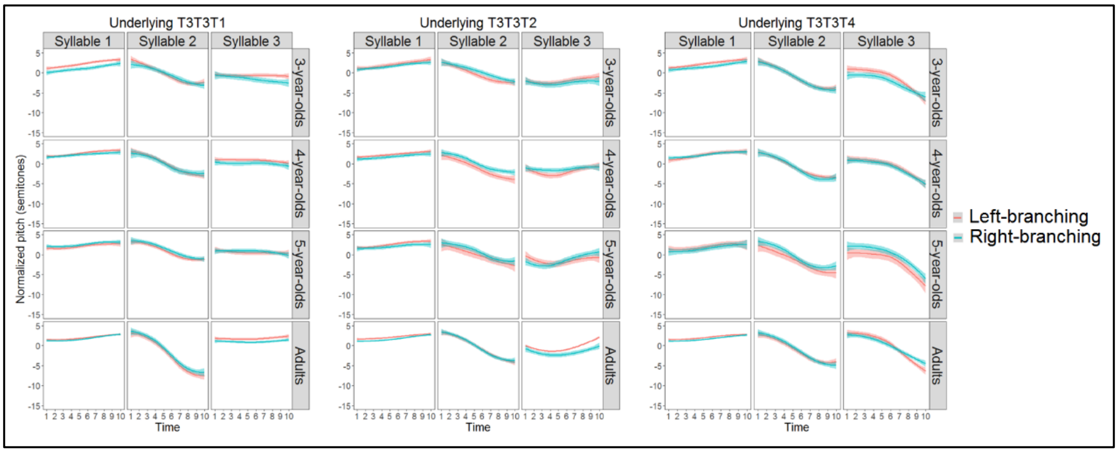


Figure 9. Pitch contours of child (3-, 4- and 5-year-olds) and adult productions of trisyllabic tone sandhi controls: ((T3T3) T1/2/4) and (T3 (T3T1/2/4)).

To examine any group differences in surface tone of the T3T3T1/2/4 control items, a linear mixed regression model was constructed using the three pitch contour parameters (intercept, slope and curvature) of the first T3 syllable within the trisyllabic compound, since this is where the tone sandhi process takes place. Two fixed factors "Group" (3-, 4-, 5-year-olds and adults) and "Structure" (left-branching and right-branching) and a random factor "Participant" were entered into the model. We expected that there would be no main effect of "Structure" on the pitch shape (i.e., pitch slope and pitch curvature) as the surface tone does not differ between structures. The results are presented in Appendix 3. Consistent with our expectation, there was no main effect of "Structure" or "Group" nor any interaction between "Structure" and "Group" on the pitch shape, indicating that the prosodic structure did not interact with their tone sandhi application on these control items, and that children's acoustic realizations were not significantly different from those of adults.

Target items: T3T3T3 compounds

For the target trisyllabic tone sandhi items, the prosodic structure was expected to guide sandhi application to result in different surface tone realizations for the two different prosodic structures. Recall that, in the *left-branching structure* ((T3T3) T3), tone sandhi applies to the *leftmost* disyllabic unit first, and then again at the level of the trisyllabic prosodic word, resulting in a surface **T2T2**T3 tonal realization. In the *right-branching structure* (T3 (T3T3)), tone sandhi applies to the *rightmost* disyllabic unit first, and this is then incorporated into the trisyllabic prosodic word, resulting in a T3**T2**T3 surface tone realization. However, this right-branching structure has also been reported to 'optionally' surface as **T2T2**T3, thought to occur in fast speech (Chen, 2000, p. 379).

Since the right-branching structure can exhibit two possible surface realizations of syllable 1, i.e., T3**T2**T3 and **T2T2**T3, it was necessary to check whether this is the case. Standard deviations (SD) of the pitch slope for syllable 1 from both the left- and right-branching structures were compared (see Table 5 for descriptive data). Adults' SD for the right-branching structure was twice that of the left-branching structure (0.2 vs. 0.48), whereas children's SD did not show this pattern. As suggested by Chen (2000), adults produced more variable surface outputs for the right-branching structure. Therefore, it is necessary to check whether the large variability fell into the two possible **T2T2**T3 or T3**T2**T3 forms as suggested in Chen (2000). Surface forms for the trisyllabic items were therefore first coded perceptually by a trained native Mandarin-speaker (the first author). Acoustic comparisons were made between adults and children to explore any group differences in the phonetic implementation of identical surface tones.

Table 5. Descriptive data of the F0 slope (mean, SD)) of Syllable 1 in children's and adults' T3T3T3 productions.

Structure	**Group**	**Slope (mean)**	**Slope (SD)**
((T3T3) T3)	3-year-olds	0.37	0.28
	4-year-olds	0.34	0.24
	5-year-olds	0.33	0.26
	Adults	0.21	**0.2**
(T3 (T3T3))	3-year-olds	0.41	0.36
	4-year-olds	0.29	0.25
	5-year-olds	0.32	0.31
	Adults	-0.12	**0.48**

Figure 10 shows the proportion of children and adults who produced the two surface tone patterns for 2 prosodic structures. Ten percent of the tokens were re-assessed by another trained native speaker, with an interrater reliability of 0.96.

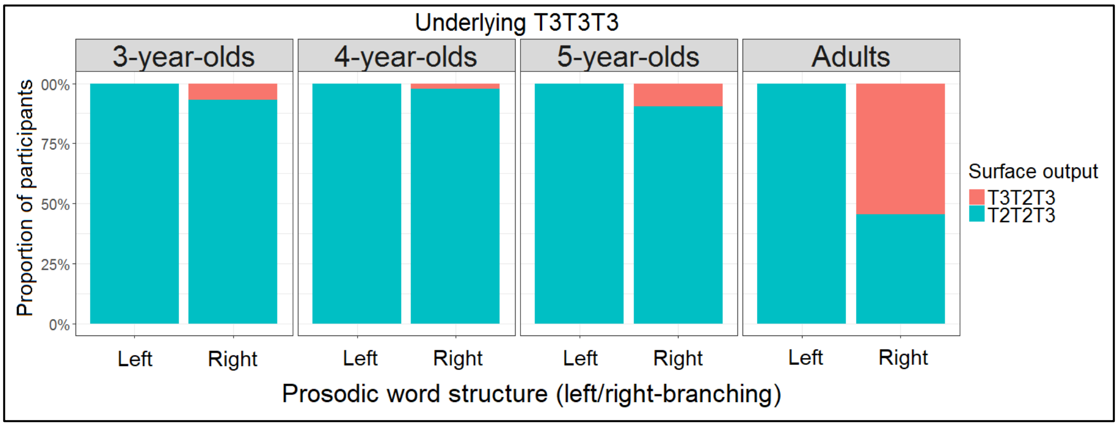


Figure 10. Proportion of participants who produced surface tones **T2T2**T3 vs. T3**T2**T3 for ((zi3 ma3) gu3) *purple-horse drum* and (zi3 (ma3 gu3)) *purple horse-drum*, based on perceptual coding.

All children and adults produced **T2T2**T3 as the surface tone for the *left-branching item* ((T3T3) T3). However, the children differed from adults in the choice of surface tone for the *right-branching item* (T3 (T3T3)), with most of the children producing **T2T2**T3. In contrast, 55% of adults (18 out of 33 participants) used the expected T3**T2**T3 surface form, whereas the other 45% (15 out of 33 participants) used the **T2T2**T3 surface form (recall that each participant produced only item per prosodic structure). To investigate the proportion of surface tone patterns for each prosodic structure, Chi-square tests were then conducted. The results showed that, (1) for the child groups, counts of **T2T2**T3 surface productions did not differ between the two prosodic structures (3-year-olds: $\chi^2(1) = 2.828$, $p = 0.093$; 4-year-olds: χ^2 $(1) = 0.329$, $p = 0.566$; 5-year-olds: χ^2 $(1) = 2.86$, $p = 0.091$); (2) for adults, however, there was a significant difference between the two conditions (χ^2 $(1) = 31.871$, $p < 0.001$). Thus, children used **T2T2**T3 as the surface tone for *both* trisyllabic prosodic word structures, while adults used **T2T2**T3 for the *left-branching structure*, but both T3**T2**T3 and **T2T2**T3 for the *right-branching structure*. In other words, children produced the same surface tones (**T2T2**T3) as adults in the **left-branching** structure, but differed from adults in sandhi application in the **right-branching** structure, with all but a few children (four 3-year-olds, one 4-year-old and two 5-year-olds) producing **T2T2**T3 in contrast to adults' variable production of both T3**T2**T3 and **T2T2**T3.

Figure 11 shows the overall child and adult acoustic realizations of pitch for the T3T3T3 items as a function of the two different prosodic structures. Note the two different realizations for the right-branching forms for the adults: both T3**T2**T3 and **T2T2**T3. We then performed two acoustic analyses to examine (1) whether child trisyllabic tone sandhi productions were adult-like and (2) whether the adult **T2T2**T3 output for the right-branching structure was a result of faster speaking rate, as proposed by Chen (2000).

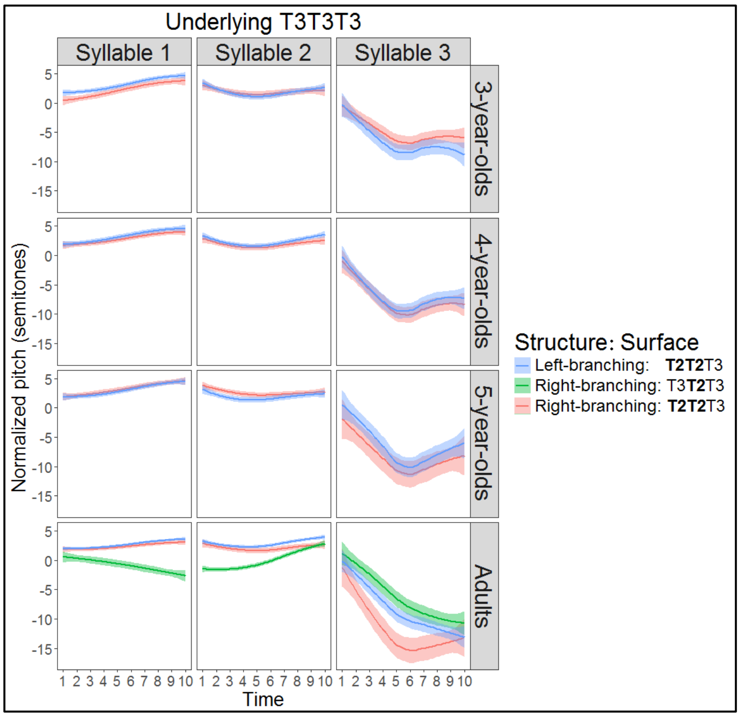


Figure 11. Pitch contours of child (3-, 4- and 5-year-olds) and adult productions of target trisyllabic tone sandhi items ((zi3 ma3) gu3) *purple-horse drum* and (zi3 (ma3 gu3)) *purple horse-drum*.

To answer the first question, we acoustically compared children's productions with those from adults. Since most of the children only produced **T2T2**T3 as the surface tone for both structures, group comparisons were **only** made with perceived **T2T2**T3 surface productions from children and adults. A linear mixed regression model was built on the first two syllables (where the tone sandhi process occurs), with two fixed factors "Group" (3-, 4-, 5-year-olds and adults) and "Structure" (left- and right-branching) and the random factor "Participant". The results showed that, for the first two syllables, there was no main effect of "Group" on the pitch slope (syllable 1: $F(3, 140) = 2.4$, $p = 0.07$; syllable 2: $F(3, 140) = 2.24$, $p = 0.09$), indicating that children's **T2T2**T3 surface tones were not different from those of adults.

It has been proposed that faster speaking rate may optionally change the surface tone of the right-branching structure from T3**T2**T3 to **T2T2**T3 in adult speech, leading to variation in tone sandhi application for this structure (Chen, 2000, p. 379). To examine whether the adult **T2T2**T3 productions for the *right-branching* structure were a result of fast speaking rate, we compared the normalized syllable duration (as a proxy of speaking rate) of adult *right-branching* surface outputs (T3**T2**T3 vs. **T2T2**T3). Syllable duration was normalized to control for any individual differences in speaking rate. The mean syllable durations of all three syllables (i.e. duration of vowel portion of /ts**i**3 m**a**3 k**u**3/) were measured and normalized against the mean duration across all tokens produced by each individual speaker. A linear mixed-effects model was constructed on the normalized syllable duration for all right-branching adult prosodic words, with a fixed factor "Surface-realization" (T3**T2**T3 vs. **T2T2**T3) and the random factor "Participant". If there was a speaking rate difference, we predicted a main effect of "Surface-realization" where the normalized syllable duration would differ between the two output forms. However, the result did not support this prediction: the main effect of Surface-realization was not significant: $F_{(1, 31)} = 1.03$, $p = 0.31$; the average normalized syllable duration for the T3**T2**T3 productions was 0.84 (SD = 0.23, range: 0.35 to 1.4), and for the **T2T2**T3 productions was 0.87 (SD = 0.31, range: 0.26 to 1.4). Thus, the normalized syllable duration did not differ between the T3**T2**T3 and **T2T2**T3 forms for the right-branching prosodic words, suggesting that speaking rate was not the source of these variable adult surface forms.

Discussion

Phonological processes can pose a challenge for learners, leading to late acquisition (Kazazis, 1969; Zamuner, Kerkhoff & Fikkert, 2006; Kerkhoff, 2007). In this study, we explored the acquisition of tone sandhi processes in Mandarin Chinese by 3 to 5-year-old children. We asked two questions. The first concerned children's ability to productively apply the tone sandhi process to novel disyllabic compounds in appropriate tonal contexts. The second concerned children's ability to construct different prosodic structures to guide single or recursive sandhi rule application in trisyllabic novel compounds. We addressed these two questions by conducting acoustic analysis to examine the fine-grained pitch contours of children's tone sandhi productions across tonal contexts and prosodic structures. To our knowledge, very few studies have carried out similar analyses of children's tone sandhi productions, either with novel compounds, or using acoustic methods.

Regarding the productive knowledge of the tone sandhi process, we hypothesized in H1 that 3-year-olds would not be able to productively apply the tone sandhi process to novel **disyllabic** compounds, but that this would become more adult-like with age. However, our results provide acoustic evidence that children from all age groups (3, 4 and 5 years) were able to productively apply the (full and half) tone sandhi process to novel disyllabic compounds, changing underlying T3 to T2 before another T3, and to a low-falling tone before T1/2/4. This result therefore extends our understanding of children's tone sandhi knowledge from the early emergence of tone sandhi productions in lexicalized items (by 3 years) (Li & Thompson, 1977; Hua & Dodd, 2000) to the productive application of tone sandhi processes in novel items. These findings are also consistent with the reported acquisition of certain grammatical tone sandhi processes in Bantu languages, also acquired by the age of 3 (Demuth, 1993).

The early emergence of tone sandhi in Mandarin lexicalized items could be related to the good realization of tone sandhi in the early language input these children hear. For example, Tang and colleagues (2017) found that, despite the slow speaking rate in Mandarin infant-directed speech (IDS), mothers correctly use tone sandhi in their speech to their 12-month-olds, providing Mandarin-learning infants with a good language environment in which to learn these tone sandhi processes.

We then explored children's ability to apply the tone sandhi process to novel trisyllabic words, where learners need to use different prosodic structures to guide the application of tone sandhi. Recall that the expected surface tone patterns for the left-branching and right-branching structures were **T2T2**T3 and T3**T2**T3 ~ **T2T2**T3, respectively (Chen, 2010).

Results from the perceptual and acoustic analyses found that both adults and children produced the expected **T2T2**T3 surface tones for the **left-branching** structure, and did not differ from one another in the phonetic implementation of these forms. This suggests that, by 3 years, children are already able to recursive apply the tone sandhi process to novel left-branching trisyllabic items. This extends previous findings from Wang (2011) that 3-year-olds can produce **T2T2**T3 as the surface form for left-branching compounds when these contain a familiar disyllabic lexicalized item (e.g. ((shui3 guo3) niao3) *fruit-bird*). This early ability to recursively apply the tone sandhi process in the left-branching structures again probably benefits from the language input, where Mandarin-speaking mothers consistently produce **T2T2**T3 for left-branching structures in IDS (Tang et al., 2017).

However, adults and children differed in their surface outputs for the **right-branching** trisyllabic words, where half of the adults produced the expected T3**T2**T3 surface form, but the other half, and most of the children, produced the unexpected recursive **T2T2**T3 surface form. This is *not* consistent with previous findings that 3-year-olds generally produce the expected T3**T2**T3 surface pattern for right-branching items such as (shui3 (lao3 hu3)) *water-tiger* (Wang, 2011), where the '(lao3 hu3)' *tiger* is a lexicalized disyllabic word. This raises many questions as to why the adults showed variability in their realization of the right-branching trisyllabic compounds in the present study[16], and why children generally used this unexpected form.

Adults: speaking rate

One possible explanation for adults' surface variation in the right-branching trisyllabic compounds was that the **T2T2**T3 surface pattern results from a fast speaking rate, as suggested by (Chen (2010, p. 379). To test this possibility, we compared the (normalized) durational difference between adults' expected T3**T2**T3

[16] Note that the surface variation in adults' right-branching productions was also unlikely to be driven by the experimental order in which they were presented, as the presentation order for the left-branching and right-branching forms was counterbalanced across participants, and there was no correlation between order of presentation and single vs. recursive tone sandhi application.

outputs and the alternative **T2T2**T3 outputs in the right-branching productions. There was no durational difference between the two, indicating that adults' surface variation was not driven by speaking rate, at least in the present study. Moreover, Tang et al. (2017) observed no surface tonal variation for right-branching trisyllabic words across different registers, even though the speaking rate differed as a function of IDS (hyperarticulated with a slow speaking rate), adult-directed speech (ADS, normal speaking rate) and Lombard speech (used in noise, also with a slow speaking rate); all (female) speakers consistently produced the expected T3**T2**T3 for the trisyllabic right-branching word (xiao3 (ma3 yi3)) *small ant* across all registers (including at least 10 tokens for each register, for a total of 442 tokens produced across registers). That is, there was no variation in these participants' production of tone sandhi in this right-branching context. This suggests again that speaking rate may not be involved in adults' variable realization of tone sandhi on right-branching compounds, and this variation may be driven by other factors. Note, however, that in the Tang et al. (2017) study, like the Wang (2011) study, these trisyllabic compounds were composed of a monosyllable followed by a lexicalized disyllabic word to form an Adjective + Noun sequence.

We therefore considered two other factors that might help explain the adult variation in surface tones for the right-branching structure found in both Chen's and our results: (1) the frequency distribution of the left-branching vs. right-branching structures in Mandarin; and (2) the lexical representation of the novel right-branching forms used in the present study.

Frequency distribution of left-branching vs. right-branching compounds in Mandarin

One factor possibly accounting for adults' surface variation in the right-branching structure might be related to the frequency distribution between the left-branching and right-branching structures in Mandarin Chinese. Duanmu (2012) conducted a corpus study on the word length in Mandarin Chinese, using the Lancaster Corpus of Mandarin Chinese (one million words of written Mandarin Chinese; McEnery & Xiao, 2004), observing that 95% of trisyllabic noun compounds had left-branching structure (disyllable + monosyllable) with only 5% having right-branching structure (monosyllable + disyllable). Therefore, relative to the predominant left-branching structure, perhaps the **prosodic representation** of the less frequent right-branching structure is weak, and this weakly represented structure might undergo prosodic reorganization under certain conditions, especially when realized in a novel word. Thus, perhaps adults' optional **T2T2**T3 surface pattern for the right-branching trisyllabic items is simply a default to the more frequent tonal pattern found on trisyllabic words, irrespective of the internal prosodic structure.

Weak lexical representation of novel trisyllabic compounds

Apart from the asymmetric frequency distribution between the two structures, the second factor leading to some adults' variable tonal realization for the right-branching structure might be related to the **weak lexical representation** of the novel compounds used in the present study. Recall that both trisyllabic words were first composed of a *novel disyllabic compound* (composed of two monosyllabic words) which was then composed with another monosyllabic word to form a *novel trisyllabic compound*, i.e., left-branching: ((**zi3 ma3**) gu3*) purple-horse drum*, and right-branching: (zi3 (**ma3 gu3**)) *purple horse-drum*. Perhaps the novel disyllabic component in each condition had only a weak lexical representation. This could have led some adults to reorganize the prosodic structure and produce the more frequent surface tonal pattern, i.e., **T2T2**T3.

Furthermore, the first syllable 'purple' (adjective) was identical in both the left- and right-branching structures. In the left-branching structure, the adjective was first combined with a noun to form a disyllabic noun, then incorporated as part of a trisyllabic compound with the structure of ((Adj + N) N). In the right-branching structure, the adjective was combined with a novel disyllabic noun as part of a noun phrase, with the structure of (Adj (N + N)). It has been suggested, in Chinese, the Adj + N sequence exhibit a strong monomorphemic status (Xu, 2018), which is in accord with the left-branching structure, i.e., ((Adj + N) N). This monomorphemic status might potentially bias some adults to reorganize the prosodic structure of the right-branching items toward the left-branching structure and produce the unexpected T2T2T3 surface pattern, especially when the disyllabic component does not have a robust lexical representation. In contrast, for the right-branching words such as (shui3 (lao3 hu3)) *water-tiger* used in Wang (2011), which consisted of two lexicalized items constituting a N+N compound, adults uniformly produced the expected T3**T2**T3 surface pattern. This suggests that the weak lexical representation of the novel disyllabic component in the right-branching structure used in the present study, along with the low overall frequency of right-branching structures in Mandarin and the monomorphemic status of the Adj + N sequence, may all have led some adults to reorganize these items into left-branching trisyllabic words.

Children's T2T2T3 productions for the right-branching structure

We turn now to the children's results and discuss why they used only the unexpected **T2T2**T3 surface pattern for the right-branching item[17]. As mentioned above, Tang et al. (2017) showed that, in children's early language input, Mandarin-speaking mothers consistently produced **T2T2**T3 vs. T3**T2**T3 for left-branching vs. right-branching items, suggesting that children abundant evidence for the canonical surface realizations for these respective trisyllabic words in the input they hear. Children's **T2T2**T3 productions then were not driven by variable tonal realizations of right-branching forms in their environment. Thus, it appears that, as for adults, children's **T2T2**T3 productions may be related to (1) the higher frequency of left-branching forms in the overall input they hear, and (2) the weak lexical representation of the novel disyllabic component and the monomorphemic bias of the Adj + N unit.

To examine the first issue, we calculated the frequency distribution of left-branching vs. right-branching trisyllabic words using data from caregiver child-directed speech from the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018) (cf. CHILDES database (MacWhinney, 2000)). A total of 2237 trisyllabic words were found, of which 67% had left-branching structure and 33% had right-branching structure. Again, these trisyllabic words all consisted of a monosyllabic and a disyllabic (lexicalized) word. Thus, it is possible that this predominant prosodic

[17] One anonymous review pointed out that, given that only one item (i.e., 'purple') was used in the critical initial position, it was unclear whether children's different performance from adults was driven by their knowledge of this particular word, or was due to other factors. However, as noted above, the word 'zi3' *purple* is a high frequency word in the language input to 3-year-olds. Moreover, children were able to correctly apply tone sandhi on 'purple' in the disyllabic unit 'purple horse', where they had to change 'purple' from T3 to T2, demonstrating that they understood that 'purple' is an underlying T3 word which should undergo tone sandhi before another T3. Therefore, children's different performance from adults on the trisyllabic words must be driven by factors other than their knowledge of this particular word.

structure in children's language input biased children's parsing of the novel trisyllabic word, leading to the **T2T2**T3 surface output.

However, the structural frequency distribution itself cannot fully account for children's **T2T2**T3 productions for the right-branching item, since the results from Wang (2011) suggest that children at 3 years consistently produced T3**T2**T3 as the surface output for right-branching compounds such as (shui3 (lao3 hu3)) *water-tiger*. Therefore, we proposed that, similar to adults, the weak lexical representation of the novel disyllabic component in our trisyllabic compounds and the monomorphemic status of the Adj + N sequence might help account for children's **T2T2**T3 realizations. The present study used three independent monosyllabic words to generate trisyllabic novel compounds (e.g., (zi3 (ma3 gu3)) *purple horse-drum*) in which the disyllabic component "ma3 gu3" *horse-drum* does not coincide with a lexicalized word, and thus does not have a strong lexical representation. Moreover, as mentioned earlier, the monomorphemic status of Adj + N sequence may have biased children towards the structure of 'purple-horse'. Without the boost from a lexical representation, this bias might have made the right-branching structure more vulnerable to structural reorganisation. In contrast, the right-branching items such as (shui3 (lao3 hu3)) *water-tiger* used in Wang (2011) consisted of two items with robust lexical representations, where children then produced the expected T3**T2**T3 surface pattern.

Our results therefore indicate that Mandarin Chinese-speaking children (and some adults) rely on lexical representations, the morphological structure and structural frequency, to guide their application of the tone sandhi rule. When the lexical representation of the word is robustly represented in the lexicon, and aligns with the prosodic structure, 3-year-olds apply the tone sandhi process with the expected surface pattern (see Wang, 2011). However, when the lexical representation of a word is weak, and differs from the most common morphological and prosodic structure, it may be susceptible to prosodic reorganization. This suggests that lexical representations may facilitate children's learning of both prosodic structure and phonological processes. Given that there was only one T3T3T3 test item for each structure in the current study, it would be useful in future studies to include more test items to probe these issues further, exploring within and between speaker variable application of tone sandhi as a function of lexical, morphological and prosodic structure, for both children and adults.

Conclusion

This study examined the acquisition of a complex phonological process in Mandarin Chinese, the tone sandhi process. The results showed that 3-year-olds were able to acquire the tone sandhi process and productively apply it to novel disyllabic compounds in different tonal contexts. However, building different prosodic structures to guide tone sandhi application in novel trisyllabic compounds appears to be a protracted process not yet completed by age 5, where the robustness of lexical representations and structural frequency in the input appear to influence children's application of the tone sandhi rule. This raises many questions about when and how these tonal alternations, and their variable realizations, are learned, and the factors that influence the establishment of adult-like tonal representations more generally.

References

Albright, A., & Hayes, B. (2011). Learning and learnability in phonology. *The Handbook of Phonological Theory (Second Edition)*, 661-690.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3), 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R Package, version 1 (7), 1-23.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2-3), 150-177.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences* (Vol. 17, No. 1193), pp. 97-110.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. [Computer program]. Version 6.0. 19.

Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese-speaking children's play narratives. *Narrative Inquiry*, 8 (1), 77-111.

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects* (Vol. 92). Cambridge University Press.

Deng, X. & Yip, V. (2018). A multimedia corpus of child Mandarin: The Tong corpus. *Journal of Chinese Linguistics*, 46 (1): 69-92.

Demuth, K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language*, 20(2), 275-301.

Duanmu, S. (2012). Word-length preferences in Chinese: a corpus study. *Journal of East Asian Linguistics*, 21(1), 89-114.

Goldsmith, J. A. (1976). *Autosegmental Phonology* (Vol. 159). Bloomington: Indiana University Linguistics Club.

Hendricks, A. E., Miller, K., & Jackson, C. N. (2018). Regularizing UnpredicTable 5ariation: Evidence from a Natural Language Setting. *Language Learning and Development*, 14(1), 42-60.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, 27(1), 3-42.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredicTable 5ariation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151-195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30-66.

Kazazis, K. (1969). Possible evidence for (near-) underlying forms in the speech of a child. In *Proceedings of the Chicago Linguistic Society* (Vol. 5, pp. 382-388). Chicago: Chicago Linguistic Society.

Kerkhoff, A. O. (2007). *Acquisition of Morpho-phonology: The Dutch Voicing Alternation* (Doctoral dissertation, LOT).

Kisseberth, C. W. (1970). On the functional unity of phonological rules. *Linguistic Inquiry*, 1(3), 291-306.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. R Package, version 2(0).

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1-33.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behaviour Research Methods*, 49(4), 1494-1502.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

McEnery, A., & Xiao, Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. *Religion*, 17, 3-4.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Chapman & Hall/CRC.

Odden, D. (1986). On the Role of the Obligatory Contour Principle in Phonological Theory. *Language*, 353-383.

Peters, J., Hanssen, J., & Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics*, 46, 185-209.

R Core Team (2016). “R: A language and environment for statistical computing” [Computer program]. Version 3.3.1, https://www.R-project.org/ (Last viewed June 21, 2016).

Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, 20, 81-124.

Sommerstein, A. H. (1974). On phonotactically motivated rules. *Journal of Linguistics*, 10(1), 71-94.

Song, J. Y., Shattuck-Hufnagel, S., & Demuth, K. (2015). Development of phonetic variants (allophones) in 2-year-olds learning American English: A study of alveolar stop/t, d/codas. *Journal of Phonetics*, 52, 152-169.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(5), 2823-2835.

Van de Vijver, R. & Baer-Henney, D. (2013). On the development of the productivity of plural suffixes in German. In *The 37th Annual Boston University Conference on Language Development* (pp. 444-455).

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language*, 29(2), 225-250.

Wang, C. Y. (2011). *Children's Acquisition of Tone 3 Sandhi in Mandarin*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, 7, 1-14.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83.

Xu, Z. (2018). The word status of Chinese adjective-noun combinations. *Linguistics*, 56(1), 207-256.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' Acoustic Realization of Mandarin Tone Sandhi and Lexical Tones in Context are Not Yet Fully Adult-like. *Frontiers in Psychology*, 9, 817.

Yang, C. D. (2002). *Knowledge and Learning in Natural Language*. Oxford University Press on Demand.

Yuen, I., Xu Rattanasone, N., McDonald, G., Holt, R.., & Demuth, K. (2014). Temporal planning in the production of Australian English compounds. In Hay. J. & Parnell, E. (Eds.), *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, pp.142-145.

Yip, M. (1989). Contour tones. *Phonology*, 6(1), 149-174.

Yip, M. (2002). *Tone*. Cambridge University Press.

Zamuner, T. S., Kerkhoff, A. & Fikkert, J. P. M. (2006). Acquisition of voicing neutralization and alternations in Dutch. Paper presented at the *Boston University Conference on Language Development*, Boston.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153-201.

Appendix

Appendix 1. Results of linear mixed regression model with second-order polynomials on pitch points of lexical tones across age groups and tones. Items in bold indicate significant findings: $p < 0.05$*, $p < 0.01$** and $p < 0.001$***.

Trends	Factors	df 1	df 2	F	*p*
Intercept (pitch height)	**Group**	**3**	**140**	**46.44**	**< .001*** **
	Tone	**3**	**10479**	**1826**	**< .001*** **
	Group × Tone	**9**	**10480**	**12.08**	**< .001*** **
Linear trend (pitch slope)	Group	3	210	0.11	.96
	Tone	**3**	**10502**	**293.85**	**< .001*** **
	Group × Tone	9	10505	1.64	.10
Quadratic trend (pitch curvature)	Group	3	170	1.24	.30
	Tone	**3**	**10492**	**417.61**	**< .001*** **
	Group × Tone	9	10494	1.64	.10

Note. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Tone + (1 + Group + Tone | Participant : Item)

Appendix 2. Results of linear mixed regression model with second-order polynomials on pitch points of tone sandhi syllables (the first syllable: T3) in disyllabic T3TX words across age groups and types. Items in bold indicate significant findings: $p < 0.05$*, $p < 0.01$** and $p < 0.001$***.

Trends	**Factors**	**df 1**	**df 2**	**F**	***p***
Intercept (pitch height)	Group	3	142	1.2	.30
	Type	**1**	**22084**	**5871.8**	**< .001*** **
	Group × Type	**3**	**22083**	**6.4**	**< .001*** **
Linear trend (pitch slope)	Group	3	146	0.1	.98
	Type	**1**	**22084**	**5095**	**< .001*** **
	Group × Type	3	22084	2.1	.09
Quadratic trend (pitch curvature)	Group	3	249	0.6	.59
	Type	**1**	**22082**	**190.6**	**< .001*** **
	Group × Type	3	22082	0.6	.63

Note. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Type + (1 + Group + Type | Participant : Item)

Appendix 3. Results of the linear mixed regression model with second-order polynomials on pitch points of tone sandhi syllable (the first syllable) in trsyllabic T3T3T1/2/4 words across age groups and structures. Items in bold indicate significant findings: p < 0.05*, p < 0.01** and p < 0.001***.

Trends	**Factors**	**df 1**	**df 2**	**F**	***p***
Intercept (pitch height)	Group	**3**	**137**	**84.61**	**< .001*** **
	Structure	**1**	**7600**	**92.81**	**< .001*** **
	Group × Structure	**3**	**7601**	**9.22**	**< .001*** **
Linear trend (pitch slope)	Group	3	135	1.7	0.17
	Structure	1	7639	0.97	.32
	Group × Structure	3	7641	2.32	.07
Quadratic trend (pitch curvature)	Group	3	284	1.37	0.25
	Structure	1	7593	0.11	.74
	Group × Structure	3	7592	0.44	.73

Note. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Structure + (1 + Group + Structure | Participant : Item).

Chapter Six: The Acquisition of Mandarin Tonal Processes by Children with Cochlear Implants

This chapter is based on the following paper:

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., Demuth, K. (revision submitted). The Acquisition of Mandarin Tonal Processes by Children with Cochlear Implants. *Journal of Speech, Language, and Hearing Research*

The paper has been modified according to reviewers' comments in the first round, and is now undergoing the second round of review process.

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors and collaborators) when needed.

Abstract

Purpose: Children with cochlear implants (CIs) face challenges in acquiring tonal languages, as CIs do not efficiently code pitch information. Mandarin is a tonal language with lexical tones and tonal processes such as neutral tone and tone sandhi, exhibiting contextually conditioned tonal realizations. Previous studies suggest that early implantation and long CI experience facilitate the acquisition of lexical tones by children with CIs. However, there is a lack of acoustic evidence on children's tonal productions demonstrating that this is the case, and it is unclear whether and how children with CIs are able to acquire contextual tones. This study therefore examined the acoustic realization of both lexical tones and contextual tones as produced by children fitted with CIs, exploring the potential effects of age at implantation and length of CI experience on their acquisition of the Mandarin tonal system.

Method: Seventy-two Mandarin-learning pre-schoolers with CIs, varying in age at implantation (13-42 months) and length of CI experience (2-49 months), and 44 normal-hearing (NH) 3-year-old controls were recruited. Tonal productions were elicited from both groups using picture-naming tasks and acoustically compared.

Results: Only the early implanted group (i.e., implanted before age 2) produced normal-like lexical tones, and generally produced contextual tones approximating those of the NH children. The other children, including those with longer CI experience, did not have typical tonal productions; their pitch patterns for lexical tones tended to be flatter and contextual tone productions were unchanged across tonal contexts.

Conclusion: Children with CIs face challenges in acquiring Mandarin tones, but early implantation may help them to develop normal-like lexical tone categories, which further facilitates their implementation of contextual tones.

Introduction

Advances in cochlear implant (CI) technology have made oral language acquisition an obtainable goal for many children with profound hearing loss. Studies have demonstrated that CIs can assist many of these children to achieve a satisfactory level of speech production and perception (e.g., Ching et al., 2018). However, it can be a challenge for children with CIs who are learning a tonal language to build a typical tonal system, since pitch information, which is critical for contrasting different tones and words, is not transmitted effectively via current CI speech processing programs, as a consequence of a restricted number of channels used to deliver a wide range of speech frequencies (Vandali & van Hoesel, 2012).

Mandarin is a tone language with four lexical tones, primarily contrasting in the following pitch contours (Yip, 2002): T1 (Level), T2 (Rising), T3 (Dipping) and T4 (Falling) (see Figure 1). These four tones are used to differentiate word meanings, e.g., /ma1/ *mother*, /ma2/ *hemp*, /ma3/ *horse* and /ma4/ *scold*. It has been suggested that, relative to segments (consonants and vowels), lexical tones are acquired much earlier by normal-hearing (NH) children, before age 3 (Li & Thompson, 1977; Hua & Dodd, 2000). A recent acoustic investigation (Tang et al., submitted) has also revealed that NH 3-year-olds were already able to produce adult-like pitch contours for all lexical tones, showing their mastery of lexical tone categories/representations.

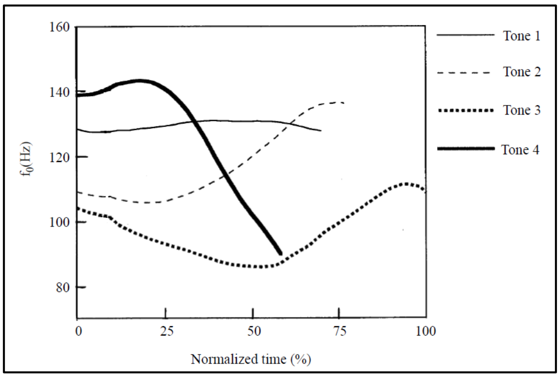


Figure 1. Mandarin Chinese lexical tone pitch contours, from Xu (1997, p. 67).

Studies of tonal acquisition have generally found that, relative to NH children, tone production accuracy is lower for children with CIs, but significant individual variability exists (see Tan, Dowell & Vogel, 2016 for a review). Age at implantation and length of CI experience have been found to be significant predictors in explaining some of this variability in overall tone production accuracy (e.g., Han et al. 2007). For instance, based on transcriber judgement, Han et al. (2007) reported that the tone production accuracy ranged from 17.4% to 77.9% (mean: 48.4%; SD: 18.9%) for Mandarin-learning children with CIs, while it ranged from 69.4% to 96.9% (mean: 78.0%; SD: 7.3%) for NH children. Despite this, Han et al. (2007) also reported significant correlations between the accuracy of tone production and both age at implantation (r = -0.57) and CI experience (r = 0.56), suggesting that early implantation and longer CI experience were critical for these children to master lexical tones.

However, most previous studies of tonal development in children with CIs have been limited in using subjective transcriber judgements only in their evaluation, rather than acoustic analysis. Relative to transcriber judgement, acoustic analysis reveals the fine-grained acoustic information of speech productions, allowing us to better understand the detailed surface pattern of these children's tonal realizations. Moreover, by acoustically comparing the tonal productions from children with different implantation ages or CI experience and those from NH children, acoustic analysis allows for further exploration of whether early implantation and long CI experience help children with CIs to acquire typical tonal productions, which is essential in acquiring typical speech communication ability of Mandarin Chinese. Although Xu et al. (2004) examined the pitch realization of lexical tones produced by children with CIs, and reported that their lexical tone pitch contours tended to be flatter, the authors did not explore the effects of implantation age or length of CI experience on the children's tonal productions, nor did they compare these children's tonal productions with those from NH children. It is therefore unclear if children with earlier age at implantation or longer CI experience are able to produce normal-like lexical tone productions that are acoustically comparable to those produced by NH children. The first aim of the present study was therefore to address this issue by conducting fine-grained acoustic analysis of lexical tone productions of prelingually deaf children with different ages at implantation (from 13 to 74 months) and different years of CI experience (from 2 months to 49 months), and to compare this with that of NH controls.

In everyday conversation, people communicate with each other using connected speech, characterized by various phonological processes. In connected speech, tones also undergo modifications across tonal contexts, resulting in tonal processes. Two well-studied tonal processes are neutral tone and tone sandhi phenomenon, exhibiting contextually conditioned tonal realizations. To acquire the Mandarin tonal system, children with CIs also have to learn how to produce neutral tone and tone sandhi as a function of tonal context. However, to date, studies of tone acquisition by children using CIs have focused exclusively on lexical tones. It is thus not clear how children with CIs acquire contextual tones such as neutral tone and tone sandhi.

Neutral tone (T0) is a “toneless” category without a fixed tonal representation, only appearing after a full lexical tone (Yip, 2002). Neutral tone exhibits pitch variation determined by the preceding tonal context, i.e., a falling pitch when following T1/2/4 syllables and a rising/level pitch when following T3 syllables (Tang, 2014; figure 2). Neutral tone is also a short/reduced (weak) syllable (around 40% to 60% of the duration of a full tone syllable). Neutral tone can be found in the second syllable of some reduplicative words, including many family relation terms, e.g., “ma1 **ma0**” *mother*, some function words/morphemes, e.g., the possessive particle “de0” in “niu2 **de0**” *cow’s*, and some other types (Hua & Dodd, 2000). Many neutral tone syllables are lexicalized and familiar to children, such as the reduplicatives; however, other neutral tone syllables are productive, e.g., the possessive particle /tɤ0/, which can be combined with any noun to form a non-lexicalized/new possessive word that might be unfamiliar to children, e.g., “niu2 **de0**” *cow’s*.

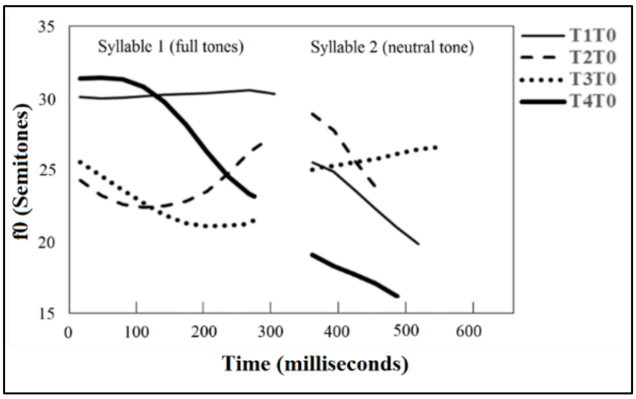


Figure 2. Pitch contours and durations of Mandarin neutral tones (T0) in different tonal contexts (after T1-4), from Tang (2014).

In Mandarin, tone sandhi is a phonological process modifying the surface realization of T3 in connected speech (Chen, 2000). In disyllables with underlying **T3**TX (T1-4) tones, the tone sandhi process changes the initial **T3** to a rising T2 before another T3 (i.e., **T3**T3 → **T2**T3; the "full sandhi" process), and to a low-falling tone before T1/2/4 (i.e., **T3**T1/2/4 → **low-falling** T1/2/4; the "half sandhi" process). For larger prosodic units, such as trisyllables with underlying tones **T3T3**TX, the tone sandhi process applies twice for triple-T3 sequences, e.g., T3T3T3 → **T2**T3T3 → **T2T2**T3 (Shih, 1997), while it only applies once for double-T3 contexts, e.g., **T3**T3T1/2/4 → **T2**T3T1/2/4.

Previous studies have found that, by age 3, NH children have already acquired both neutral tone and tone sandhi processes, showing adult-like pitch realizations across tonal contexts (Tang, et al., 2018; Tang, et al., submitted). Tang et al. (2018), for instance, conducted an acoustic analysis of NH 3-5-year-olds' neutral tone productions, finding that, by age 3, children were already able to correctly produce neutral tone pitch variations across tonal contexts for both familiar (reduplicative) and unfamiliar (possessive) items, generally with reduced/short duration. Regarding the acquisition of tone sandhi, Li and Thompson (1977) and Xu Rattanasone et al. (2018) found that NH 3-year-olds were able to produce familiar tone sandhi words, i.e., underlying T3T3 words such as "xiao3 niao3" *birdie*, with correct surface sandhi tones **T2**T3. However, given that these familiar tone sandhi words are only realized as T2T3 in children's language input, it was unclear whether these children understood the tone sandhi **process** or simply repeated what they heard in their language input. To address this issue, Tang et al. (submitted) used a novel word formation task, testing NH 3-5-year-olds' productions of novel disyllabic and trisyllabic tone sandhi compounds such as "ma3 gu3" *horse-drum* and "zi3 ma3 gu3" *purple-horse drum*.

The results showed that, by age 3, NH children were also able to productively apply the tone sandhi process with adult-like pitch implementation for both novel disyllabic and trisyllabic compounds. Overall, these results suggested that, although neutral tone and tone sandhi exhibit tonal variation across contexts, 3-year-old NH children have developed the neutral tone category and abstracted the tone sandhi process, as reflected by their contextually conditioned tonal realization of novel items.

For children using CIs, acquiring neutral tone and tone sandhi processes may be more challenging. Given previous findings that children with CIs generally face challenges in acquiring the simpler lexical tones (e.g., Tan, Dowell & Vogel, 2016; Chen & Wong, 2017), they might not be able to correctly perceive or produce the more complex pitch variations required for the realization of contextual tones. However, due to the lack of investigation examining their use of contextual tones, it was still unclear how they produce neutral tone and tone sandhi syllables across contexts, i.e., will they produce an unchanged pitch contour for these syllables across contexts, or produce contextually conditioned but incorrect pitch patterns? Given that most words in Chinese are disyllablic or polysyllabic (Duanmu, 2012), with a great number of contextual tones, exploring the surface pattern of contextual tone productions by children with CIs' would deepen our understanding of their connected speech processes. Moreover, since it has been suggested that earlier age at implantation and longer CI experience benefit lexical tone acquisition (Tan, Dowell & Vogel, 2016), it is possible that children who are implanted early, and/or have longer exposure with CIs, might have learned these tonal processes, and may be able to produce them accurately. Therefore, exploring the contextual tone productions from children with different implantation ages and CI experience would help us to better understand the effect these have on the acquisition of tones in connected speech.

Therefore, the second aim of the present study was to explore the acquisition of contextual tones by children with CIs. We asked whether these children were able to develop the neutral tone category and master the tone sandhi process as reflected by the contextually conditioned tonal realization of neutral tone and tone sandhi syllables.

This study thus consisted of three experiments involving Mandarin-learning children using CIs: examination of the acoustic realization of lexical tones (Experiment 1); examination of acquisition of the neutral tone category (Experiment 2); and examination of acquisition of tone sandhi processes (Experiment 3). To establish a baseline for comparison, we included NH Mandarin-learning 3-year-olds as controls, since previous studies showed that these children have already acquired both lexical and contextual tones (Tang et al., 2018; Tang et al., submitted). Generally, we asked whether children with early implantation or longer CI experience would show an advantage in tonal acquisition, with lexical tone productions acoustically comparable to those of NH children and correct neutral tone and tone sandhi productions across tonal contexts. Based on previous findings (e.g., Han et al., 2007; Tan, Dowell & Vogel, 2016; Chen & Wong, 2017), we predicted that the children with CIs would face challenges in producing typical tonal realizations, but perhaps this would be better for those with earlier age at implantation or longer CI experience.

General methods

The same participants, coding and measurements and statistical analyses were used for all three studies. These are briefly outlined below.

Participants

The same cohort of participants participated in all three experiments, including 72 3-7-year-old (mean age: 4;8; SD: 11 months) prelingually deafened Mandarin-speaking children with severe to profound hearing loss (CI group; Supplementary material S1), and 44 3-year-old NH controls. Children with CIs were recruited from speech rehabilitation centres in Beijing, China. These children were implanted at between 13 to 74 months of age, and their CI experience ranged from 2 to 49 months (see figure 3). Sixty-five were unilaterally implanted, and were also fitted with a hearing aid (HA) for the contralateral ear, wearing both CI and HA devices during the experiment. The other seven children were bilaterally implanted with CIs, excluded from the main analysis, but see the Supplementary material (S2) for an additional analysis on this group. Twelve children had a record of HA experience prior CI surgery (see additional analysis of the effect of HA experience prior to CI on their tonal development: Supplementary material S3). According to reports from the institutions, none of these children had intellectual difficulties. The NH controls (referred as: NH = 3; mean age: 3;8; SD: 3 months) were recruited from the affiliated kindergarten of Beijing Language and Culture University. The study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Panel.

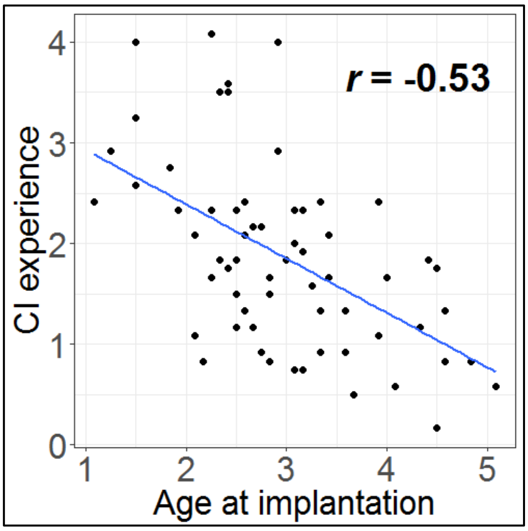


Figure 3. Participants' age at implantation and CI experience.

Each experiment consists of two types of analysis: *group comparison* and *regression analysis*. In the first analysis (group comparison), we acoustically compared the tonal productions between NH children and children with CIs across ages at implantation and lengths of hearing experience. The effects of age at implantation and length of CI experience were examined separately, since these two factors were significantly correlated (Pearson's r = -0.53, $p < 0.001$), i.e., children with earlier implantation also typically had longer CI experience, and children implanted later typically had shorter CI experience (see figure 3). Therefore, children with CIs were grouped according to age at implantation and length of CI experience respectively. Based on the age at implantation, they were grouped into four groups: 1-2 years (CI_Imp 1-2: 7 children; mean age: 4;7; SD: 10 months); 2-3 years (CI_Imp 2-3: 28 children; mean age: 4;9; SD: 11 months); 3-4 years (CI_Imp 3-4: 19 children; mean age: 5;1; SD: 7 months); and 4-5 years (CI_Imp 4-5: 11 children; mean age: 5;8; SD: 7 months). Based on the length of CI experience, they were grouped into four groups: 0-1 year (CI_Exp<1, 12 children; mean age: 4;6; SD: 8 months); 1-2 years

(CI_Exp 1-2, 27 children; mean age: 4;10; SD: 10 months); 2-3 years (CI_Exp 2-3, 19 children; mean age: 4;11; SD: 9 months); and 3-4 years (CI_Exp 3-4, 7 children; mean age: 5;12; SD: 7 months). In the second analysis (regression analysis), we performed linear-regression models on **CI groups'** productions only, with different predictors (e.g., age at implantation, CI experience, chronological age, residual hearing), to provide a supplementary analysis exploring the effect of these predictors on these children's tonal productions.

Coding and measurements

Participants' productions of lexical tones, neutral tone and tone sandhi were acoustically coded in Praat (Boersma & Weenink, 2016). Ten percent of the items were re-coded by a second trained native speaker of Mandarin (inter-rater reliability: 91%). Pitch information was extracted for lexical tone, neutral tone and tone sandhi tokens, and durational information was also extracted for neutral tone tokens.

Following standard procedures (e.g., Tang, et al., 2017a, 2017b), ten f0 points (in semitones with the reference of 50 Hz) were derived from the vowel portion of each syllable. To minimize possible inter-speaker variation in pitch, the f0 values of each token were normalized against the mean pitch across all tokens for each individual speaker.

For neutral tone, which is also characterized by short duration, normalized duration was also measured for each neutral tone syllable, calculated as the ratio between the neutral tone duration in millisecond (ms) and the preceding lexical tone duration in ms, using the following formula:

$$\text{Normalized Neutral Tone Duration} = \frac{\text{Raw Neutral Tone Duration (ms)}}{\text{Raw Preceding Lexical Tone Duration (ms)}}$$

Thus, a ratio smaller than 1 indicates that the neutral tone duration is shorter relative to the lexical tone.

Statistical analysis

The data were analysed using R (R Core Team, 2016). To quantify the lexical tone pitch contours, a second-order orthogonal polynomial equation was fitted for each token using the *poly* function of R. The second-order polynomials were adopted since the most complex pitch contour of tones in our data had only a convex or concave contour shape. The polynomial function generated two parameters for each pitch contour, i.e., (1) the linear trend and (2) the quadratic trend. According to Mirman (2016), the two parameters capture the f0 contour's *slope* (as reflected in the linear trend: a positive linear trend indicates a rising pitch and vice versa; a larger absolute value of the linear trend represents a steeper pitch and vice versa) and *curvature* (as reflected in the quadratic trend: a positive quadratic trend indicates a concave f0 contour, a negative quadratic trend indicates a convex contour; and a larger quadratic trend indicates a more curved f0 contour and vice versa).

Linear mixed regression models were built to compare the pitch contours and duration (of neutral tone) across tones and groups, using the LME4 package (Bates, Mächler, Bolker, & Walker, 2014). An advantage of mixed-effects modelling is that it is possible to fit models to large, unbalanced data, such as the unequally distributed children with different ages at implantation or CI experience in the current study (Baayen, Davidson & Bates, 2008). All random slopes were included in the model to make it maximally generalizable across the data (Barr, Levy, Scheepers, & Tily, 2013). The *anova* function, which provides Satterthwaite's approximation to degrees of freedom for estimating *p-values* in R package LMERTEST (Kuznetsova, Brockhoff, & Christensen, 2015), was used to test for statistical significance. The main and interaction effects reported in the results were averaged across all levels of the other effects (see Luke, 2017 for a detailed explanation and Peters, Hanssen, & Gussenhoven, 2014 for a practical example). When a significant main effect of a multi-level factor or a significant interaction effect was observed, Tukey-HSD post-hoc comparisons were performed on the multi-level factor, as well as interactions, using LSMEANS package (Lenth, 2016). Due to the page limitation of the paper, all tables for the results of statistical analysis are presented in the online supplementary materials.

Experiment 1: lexical tones

Stimuli

Four picturable monosyllabic words were selected as stimuli, each one including one of the four different lexical tones, i.e., "shu1" *book*, "qiu2" *ball*, "gu3" *drum* and "hua4" *painting*. All words were object names and fell within the top 50% of the most frequent monosyllabic words in the language input to NH children below the age of 3 according to the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018) from the CHILDES database (MacWhinney, 2000).

Procedure

Each participant was tested individually in a quiet space at their institutions. A picture-naming task was used in the elicitation task. All productions were recorded via an AGK C520 head-mounted microphone, placed around 3 cm in front of the participant's mouth and connected to a Marantz PMD661MKII recorder, with a sampling rate of 44.1 kHz.

The elicitation task was started with two practice trials, followed by four test trials with a randomized order across participants. The two practice trials included two monosyllabic items, i.e., “mao1” *cat* and “yang2” *sheep*, used to familiarize the participants with the task. In each trial, a Mandarin-speaking experimenter presented a picture on the computer screen and asked the participant to name it. If the participant failed to name it or produced it with a different name, i.e., produced “hua4 hua4” *drawing a painting* instead of the target word “hua4” *painting*, the experimenter would correct the child and ask him/her to name it again. The experiment proceeded once the participant correctly produced each target word. Each participant produced one token for each item. A total of 448 tokens were collected, in which 44 tokens were excluded from further analysis due to the poor acoustic quality. This resulted in 404 tokens included in the analysis (NH children: 168 tokens; CI children: 236 tokens).

Results

Group comparison

Age at implantation

Figure 4 illustrates the normalized lexical tone pitch contours for the NH 3-year-olds and the children with CIs. It shows that the NH children correctly produced all lexical tones (level (T1), rising (T2), dipping (T3), falling (T4)). For CI groups, however, it seems that the earlier implanted group (CI_Imp 1-2) produced comparable lexical tone pitch contours as NH children, but with larger variability, especially for T3 and T4, while other groups generally produced flatter pitch contours for all tones.

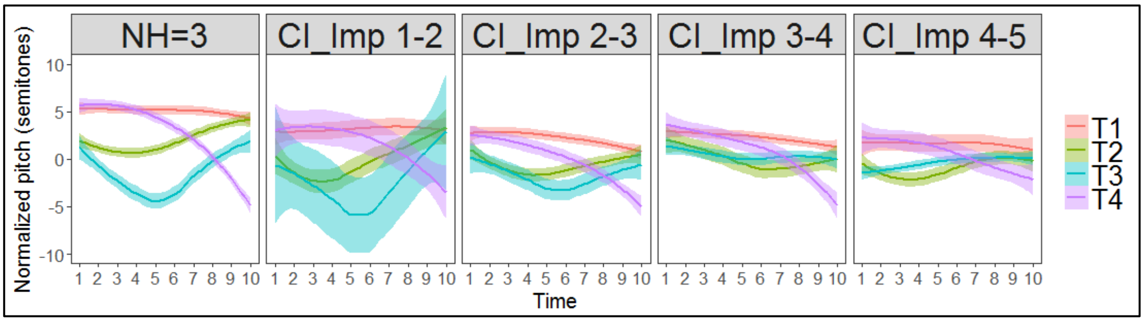


Figure 4. Pitch contours for lexical tone productions from NH 3-year-olds and CI children with different ages at implantation (1-2 years: CI_Imp 1-2, 2-3 years: CI_Imp 2-3, 3-4 years: CI_Imp 3-4, and after 4 years: CI_Imp 4-5).

To explore the effect of age at implantation on lexical tone productions, a linear mixed-effects model was performed on the pitch contour (slope and curvature) for each token, with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Tone" (T1, T2, T3 and T4), a covariate "Chronological age" (age in months; to control for the different chronological ages across participants) and a random factor "Participant". The results showed significant interactions of "Group × Tone" on pitch slope ($F (12, 5890) = 3.64$, $p < 0.001$) and curvature ($F (12, 5893) = 3.91$, $p < 0.001$; Supplementary materials S4). Tukey-HSD post-hoc tests showed that the CI_Imp 1-2 group produced pitch contours comparable to those of the NH children for all lexical tones, while other CI groups differed from the NH children in T2, T3 and T4 productions, with flatter pitch contours (Supplementary material S5).

Moreover, Figure 4 shows that NH children and CI_Imp 1-2 exhibited more stretched pitch contours (larger pitch range) than other CI groups. We thus performed a one-way ANOVA on pitch range (measured as the difference between the maximal and minimal values for each individual speaker) across groups, with an independent variable "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and a covariate "Chronological age". The results showed a significant main effect of "Group" ($F(4, 141) = 3.22$, $p < 0.05$) on pitch range (Supplementary material S6). The Tukey-HSD post-hoc test showed that CI_Imp 1-2 group showed similar pitch range as NH children, while other groups exhibited smaller pitch range than NH children (Supplementary material S7).

CI experience

Figure 5 illustrates the normalized lexical tone pitch contours from NH children (3-year-olds) and CI groups with different CI experience. It seems that children with CIs generally produced much flatter and compressed pitch contours, even those with longer CI experience.

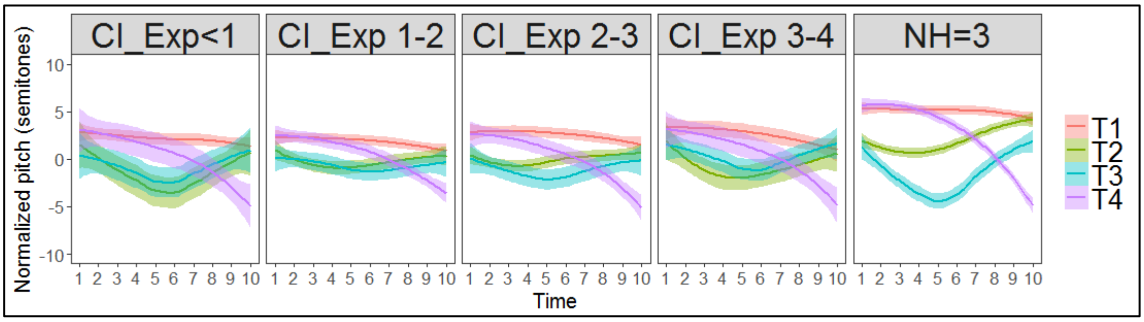


Figure 5. Pitch contours for lexical tone productions for NH 3-year-olds and CI children with different lengths of CI experience (less than 1 year: CI_Exp<1, 1-2 years: CI_Exp 1-2, 2-3 years: CI_Exp 2-3, and 3-4 years: CI_Exp 3-4).

To explore the effect of CI experience on the pitch realization of children's lexical tone productions, a linear mixed-effects model was performed on the pitch contour (slope and curvature) for each token, with two fixed factors "Group" (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3) and "Tone" (T1, T2, T3 and T4), a covariate "Chronological age" and a random factor "Participant". The results showed significant interactions of "Group × Tone" on pitch slope ($F (12, 5897) = 2.20$, $p < 0.05$) and curvature ($F (12, 5900) = 2.02$, $p < 0.05$; Supplementary material S8). Tukey-HSD post-hoc tests revealed that, relative to the NH children, all CI groups generally produced flatter pitch contours for T2, T3 and T4 (Supplementary material S9).

We also performed a one-way ANOVA to compare the pitch range across groups, with an independent variable "Group" (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3) and a covariate "Chronological age". The results showed a main effect of "Group" ($F (4, 141) = 2.93$, $p < 0.05$; Supplementary material S10). The Tukey-HSD post-hoc test revealed that, relative to NH children, the pitch range of all CI groups was smaller (Supplementary material S11).

Regression analysis

A linear regression model was performed on the pitch slope of CI groups' lexical tone productions with five predictors: "lexical tone category (T1-4)", "age at implantation", "CI experience", "chronological age" and "residual hearing". The results showed that, "age at implantation", "CI experience" and "chronological age" were all significant predictors in predicting the pitch slope of their lexical tone productions. Among these predictors, "age at implantation" exhibited larger effects on predicting the pitch of lexical tones than "CI experience" and "chronological age", as reflected by the larger standardized coefficients (β: age at implantation = 0.022, CI experience = 0.009, chronological age = -0.007; see Supplementary material S12 for detailed results).

Summary

The results of Experiment 1 showed that only children with early age at implantation (CI_Imp 1-2) produced lexical tones with pitch contours that were comparable to those of the NH children. The other CI groups generally produced flatter pitch patterns and smaller pitch ranges than the NH children, including the longer CI experience group. These results suggest that children with CIs may face challenges in producing contextual tones, since they have not yet developed the lexical tone categories upon which these depend.

Experiment 2: neutral tone

Experiment 2 examined children's neutral tone productions across tonal contexts in two word types: familiar **reduplicatives** and unfamiliar **possessives**. These two types were selected because previous studies found that NH 3-year-olds were already able to correctly produce the contextually conditioned tonal realization of neutral tone syllables in both familiar and less familiar types, showing productive application of this tonal process, as well as shortening of the short T0 syllable (Tang et al., 2018). We expected that children with CIs would face challenges in producing neutral tone productions, i.e., with contextual pitch variations and short duration, especially for the unfamiliar **possessives**. But perhaps earlier age at implantation or longer CI experience would help these children in the neutral tone acquisition.

Stimuli

Eight disyllabic neutral tone words were used as stimuli, including four familiar items (reduplicatives) and four unfamiliar items (possessives), creating four different tonal contexts for each (TXT0; Table 1). The familiar items were all (kinship) reduplicatives, such as "ge1 **ge0**" *older brother*; the unfamiliar items were all possessives composed of a monosyllabic animal's name plus the possessive particle "de0" (e.g., "niu2 **de0**" *cow's*). These two types of neutral tone words were used to test if children with CIs had developed the neutral tone category and could generalize it to unfamiliar words.

Table 1. Stimuli used in Experiment 2: disyllabic neutral tone words, including four familiar reduplicatives and four unfamiliar possessives.

Type	**Words**	**Meaning**	**Phonetic transcription**
Reduplicative (familiar items)	ge1 **ge0**	*Older brother*	/kɤ1 **kɤ0**/
	ye2 **ye0**	*Grandpa*	/je2 **je0**/
	jie3 **jie0**	*Older sister*	/tɕie3 **tɕie0**/
	di4 **di0**	*Younger brother*	/ti4 **ti0**/
Possessive (unfamiliar items)	zhu1 **de0**	*Pig's*	/tʂu1 **tɤ0**/
	niu2 **de0**	*Cow's*	/niu2 **tɤ0**/
	gou3 **de0**	*Dog's*	/kou3 **tɤ0**/
	lu4 **de0**	*Deer's*	/lu4 **tɤ0**/

Procedure

Two blocks were used to elicit familiar reduplicatives and unfamiliar possessives respectively. In the **reduplicative** neutral tone elicitation task, four pictures were presented in a sequence on a computer screen. In each picture, a pair of family members depicted in line drawings were presented side-by-side, such as an older brother (illustrated by a tall boy) and a younger sister (illustrated by a shorter girl). To elicit the word “ge1 **ge0**” *older brother*, for example, the experimenter said to the participant ‘the boy calls the girl younger sister, so what does the girl call the boy?' Once the participant produced the target word, the experimenter proceeded to the next picture to elicit the next target word.

In the **possessive** neutral tone elicitation task, two animals depicted in line drawings were simultaneously presented side-by-side, e.g., a *pig* ("zhu1") and a *cow* ("niu2"). To elicit the target neutral tone "de0" in a disyllabic word, such as "zhu1 **de0**" *pig's* and "niu2 **de0**" *cow's*, the experimenter firstly introduced the *pig* and the *cow* respectively and then pressed a button to play a pre-programmed animation (e.g., tail spinning) on one of the animals, (e.g., a *pig*). During the animation, the experimenter asked the participant whose tail it was. After the participant produced the target word "zhu1 **de0**" *pig's*, the experimenter then pressed the button to trigger an animation of a spinning tail on the other animal, for example, the *cow*, and asked the same question to elicit the target item "niu2 **de0**" *cow's*. The same procedure was used to elicit the other possessives. The order of the test trials was randomized across participants and each participant produced one token per item. A total of 842 tokens were collected, with 102 tokens excluded from further analysis due to the poor acoustic quality. This resulted in 740 tokens included in the analysis (NH children: 346 tokens; CI children: 394 tokens).

Results

Group comparison

Age at implantation

Figure 6 illustrates the normalized neutral tone pitch contours for the NH 3-year-olds and the CI groups with different ages at implantation. The NH children produced the pitch variations of neutral tone with adult-like falling pitch after T1/2/4 and a rising/level pitch after T3. The early implanted CI group (CI_Imp 1-2) also had target-like pitch variation for neutral tone syllables, but only for the lexicalized reduplicatives, not the unfamiliar possessives. The other CI groups, however, generally produced a falling pitch for neutral tones irrespective of tonal context.

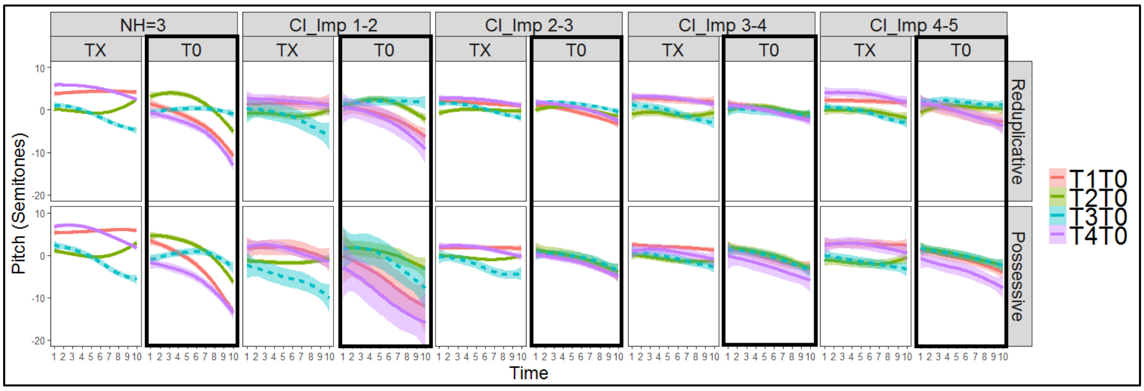


Figure 6. Pitch contours of neutral tone (T0) productions across tonal contexts (TX: T1-4) from NH 3-year-olds and CI children with different ages of implantation (1-2 years: CI_Imp 1-2, 2-3 years: CI_Imp 2-3, 3-4 years: CI_Imp 3-4, and after 4 years: CI_Imp 4-5). Black boxes indicate neutral tone syllables.

A linear mixed-effects model was performed on the pitch contour (slope and curvature) of each neutral tone syllable, exploring the effect of early implantation on children's neutral tone productions, with three fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5), "Context" (after T1/2/4 and after T3) and "Type" (reduplicatives and possessives), a covariate "Chronological age" and two random factors "Participant" and "Item" (8 neutral tone words). The results show that there was a significant three-way interaction of "Group × Context × Type" on the pitch slope (F (4, 11977) = 3.09, $p < 0.01$; Supplementary material S13), indicating that there were pitch slope differences across groups, contexts and types. Two Tukey-HSD post-hoc tests were performed on this interaction. The first test compared the pitch difference between two contexts (after T1/2/4 vs. after T3) across groups and types, exploring whether children produced pitch variation of neutral tone syllables across contexts. The results showed that: (1) NH children correctly produced the pitch variation of neutral tones across contexts; (2) CI_Imp 1-2 correctly produced the pitch variation of neutral tones for the familiar reduplicatives, but generally produced a falling pitch for the unfamiliar possessives across tonal contexts; (3) the other CI groups generally produced a falling pitch for neutral tone syllables irrespective of the tonal context (supplementary material S14). The second test compared the pitch slope difference of neutral tone between NH and CI groups, to explore whether children with CIs produced comparable pitch of neutral tone as NH children. The results showed that, consistent with the first analysis, (1) CI_Imp 1-2 produced similar pitch of neutral tone syllables as NH children, except for the (non-lexicalized) possessives after T3, where NH children produced a rising pitch but CI_Imp produced a falling pitch; (2) other CI groups generally produced flatter pitch contours of neutral tone than NH children (Supplementary material S15).

We then examined the duration of children's neutral tone productions. A linear mixed-effects model was performed on the normalized duration of children's neutral tone productions with three fixed factor "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5), "Context" (after T1/2/4 and after T3) and "Type" (reduplicatives and possessives), a covariate "Chronological age" and two random factors "Participant" and "Item". The results showed an interaction effect of "Group × Type": $F(4, 1046) = 8.864$, $p < 0.001$ (Supplementary material S16). Tukey-HSD post-hoc tests showed that the CI_Imp 1-2 group reduced neutral tone duration to the same degree as NH children, but only for (lexicalized) reduplicatives, not the (non-lexicalized) possessives, while other CI groups all produced longer neutral tone durations (Supplementary material S17).

CI experience

Figure 7 illustrates the time-normalized neutral tone pitch contours for the NH children and CI groups with different years of CI experience. Again, the children with CIs generally produced a falling pitch contour for neutral tone syllables across tonal contexts, including the longer CI experience group.

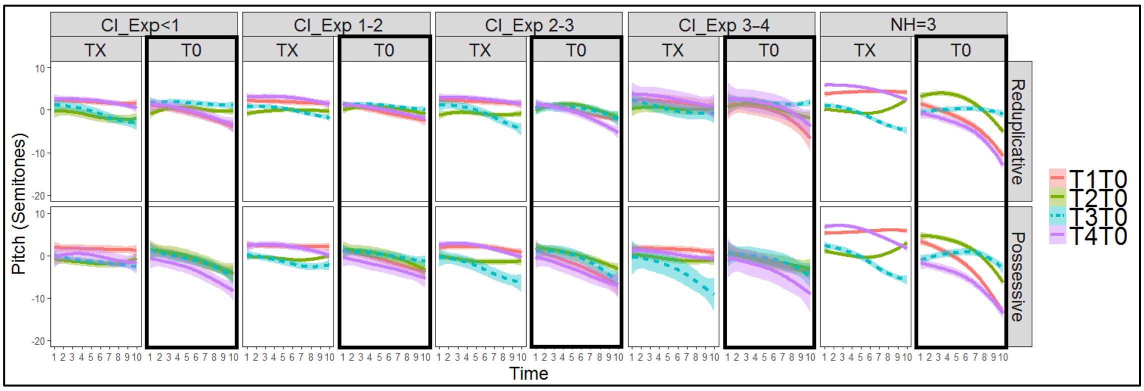


Figure 7. Pitch contours of neutral tone (T0) productions across tonal contexts (TX: T1-4) from NH 3-year-olds and CI children with different lengths of CI experience (less than 1 year: CI_Exp<1, 1-2 years: CI_Exp 1-2, 2-3 years: CI_Exp 2-3, and 3-4 years: CI_Exp 3-4). Black boxes indicate neutral tone syllables.

To investigate the effect of CI experience on the pitch realization of children's neutral tone productions, a linear mixed-effects model was performed on the pitch contour (slope and curvature) for each neutral tone syllable, with three fixed factors "Group" (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3), "Context" (after T1/2/4 and after T3) and "Type" (reduplicatives and possessives), a covariate "Chronological age" and two random factors "Participant" and "Item" (8 neutral tone words). The results show that there were significant interactions between “Group × Context” for pitch slope (F (4, 11981) = 9.485, $p < 0.01$; Supplementary material S18), indicating pitch slope differences across groups and contexts. Two Tukey-HSD post-hoc tests were then performed to explore the pitch slope difference of neutral tone (1) between contexts and (2) between groups. The first analysis showed that only NH children produced the correct pitch variation for neutral tone across contexts, while all CI groups produced an unchanged falling pitch contour for neutral tone for both contexts (Supplementary material S19). The second analysis showed that the CI groups produced falling pitch contours for neutral tone syllables similar to those of the NH children after T1/2/4, while they differed from NH children in the neutral tone productions after T3, with NH children producing a rising pitch and the CI groups producing a falling pitch (Supplementary material S20). This indicates that the CI groups did not produce appropriate pitch variation for neutral tone syllables across all contexts.

We then examined the duration of children’s neutral tone productions, exploring whether longer CI experience would help children with CIs to reduce their neutral tone duration to the same degree as NH children. A linear mixed-effects model was performed on the normalized duration of children’s neutral tone productions with three fixed factor “Group” (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3), “Context” (after T1/2/4 and after T3) and “Type” (reduplicatives and possessives), a covariate “Chronological age” and two random factors “Participant” and “Item” (8 neutral tone words). The results showed that there was a main effect of “Group”: F (4, 180) = 4.539, $p < 0.01$ (Supplementary material S21). A Tukey-HSD post-hoc test on this main effect revealed that, relative to the NH children, all CI groups produced longer neutral tone durations, and there was no significant durational differences between any two CI groups (Supplementary material S22).

Regression analysis

Linear regression models were performed on the pitch slope and duration of CI groups' neutral tone productions, with six predictors: "context (after T1/2/4 and after T3)", "type (reduplicative and possessive)", "age at implantation", "CI experience", "chronological age" and "residual hearing". The results showed, for the *pitch* of neutral tone, only "context", "type" and "age at implantation" were significant predictors; for *duration*, "type", "age at implantation", "CI experience" and "chronological age" were all significant predictors, with "age at implantation" exhibiting larger effects than "CI experience" and "chronological age", as reflected by the larger standardized coefficients (β: age at implantation = -0.228, CI experience = 0.138, chronological age = 0.123; see Supplementary materials S23 and S24 for more detail). These results suggested that the age at implantation was significant in predicting both the pitch and duration of neutral tone productions from children with CIs, with larger effects than CI experience and chronological age.

Summary

The results of Experiment 2 showed that children using CIs (including the longest CI experience group) generally faced challenges in producing the pitch variation and short duration of neutral tones, even for familiar items (reduplicatives), generally producing unchanged falling pitch and lengthened durations across tonal contexts. The early age at implantation group (CI_Imp 1-2) was able to produce correct pitch for neutral tone, but only for the familiar, lexicalized reduplicatives, rather than the unfamiliar, novel possessives.

In the next sections (experiment 3a and 3b), we examined children's productions of tone sandhi items. Experiment 3a used disyllabic compounds as stimuli, exploring children's knowledge of productive tone sandhi processes; experiment 3b used trisyllabic compounds as stimuli, exploring children's ability to recursively applying the tone sandhi process.

Experiment 3a: tone sandhi in disyllabic compounds

Experiment 3a examined children's tone sandhi productions. We asked whether children using CIs were able to acquire the tone sandhi process, productively applying it to novel compounds, i.e., producing **T2** + T3 for full sandhi (**T3** + T3) compounds, and **low falling** + T3 for half sandhi (**T3** + T1/2/4) compounds. We hypothesized that this would be challenging, but perhaps would be possible for those with earlier age at implantation or longer CI experience.

Stimuli

Sixteen novel *disyllabic* tone sandhi (T3+TX) compounds were used as stimuli. For each compound, the first syllable was always a monosyllabic animal name with T3 (e.g., "ma3" *hors*e), and the second syllable was always a monosyllabic object name with TX (the same as used to elicit the lexical tones; e.g., "gu3" *drum*); this resulted in 16 novel compounds (four T3 syllables × four TX syllables). These 16 disyllabic T3+TX words thus included four full sandhi (T3+T3) items and 12 half sandhi (T3+T1/2/4) items (see Table 2).

Table 2. Stimuli used in Experiment 3a: novel disyllabic tone sandhi compounds, including four full sandhi (T3+T3) compounds and 12 half sandhi (T3+T1/2/4) compounds.

Type	Words	Underlying tones	Surface tones	Meaning	Phonetic transcription
Full sandhi (T3+T3)	ma3 gu3	T3 + T3	**T2** + T3	*Horse-drum*	/ma3 ku3/
	niao3 gu3	T3 + T3	**T2** + T3	*Bird-drum*	/niau3 ku3/
	gou3 gu3	T3 + T3	**T2** + T3	*Dog-drum*	/kou3 ku3/
	shu3 gu3	T3 + T3	**T2** + T3	*Mouse-drum*	/ʂu3 ku3/
Half sandhi (T3+T1/2/4)	ma3 shu1	T3 + T1	**Low falling** + T1	*Horse-book*	/ma3 ʂu1/
	ma3 qiu2	T3 + T2	**Low falling** + T2	*Horse-ball*	/ma3 tɕiu2/
	ma3 hua4	T3 + T4	**Low falling** + T4	*Horse-painting*	/ma3 hua4/
	niao3 shu1	T3 + T1	**Low falling** + T1	*Bird-book*	/niau3 ʂu1/
	niao3 qiu2	T3 + T2	**Low falling** + T2	*Bird-ball*	/niau3 tɕiu2/
	niao3 hua4	T3 + T4	**Low falling** + T4	*Bird-painting*	/niau3 hua4/
	gou3 shu1	T3 + T1	**Low falling** + T1	*Dog-book*	/kou3 ʂu1/
	gou3 qiu2	T3 + T2	**Low falling** + T2	*Dog-ball*	/kou3 tɕiu2/

gou3 hua4	T3 + T4	**Low falling** + T4	*Dog-painting*	/kou3 hua4/
shu3 shu1	T3 + T1	**Low falling** + T1	*Mouse-book*	/ʂu3 ʂu1/
shu3 qiu2	T3 + T2	**Low falling** + T2	*Mouse-ball*	/ʂu3 tɕiu2/
shu3 hua4	T3 + T4	**Low falling** + T4	*Mouse-painting*	/ʂu3 hua4/

Procedure

Four practice trials were first used to familiarise participants with the procedure for forming novel compounds: “mao1 shu1” *cat-book*, “mao1 qiu2” *cat-ball*, “yang2 gu3” *sheep-drum* and “yang2 hua4” *sheep-painting*. This was identical to the procedure then used for the test trials. To elicit the disyllabic compound “ma3 gu3” *horse-drum*, for instance, a picture of a horse and a picture of a drum were displayed on the left and the right side of a robot-like cartoon figure respectively (see step 1 in figure 8). The experimenter then pressed a button to play a pre-programmed animation where the two pictures disappeared behind the robot and it jiggled to output a novel object: a horse-drum (see step 2 in figure 8). The experimenter then asked the participant to produce the name of the resulting item. The order of the test trials was randomized across participants.

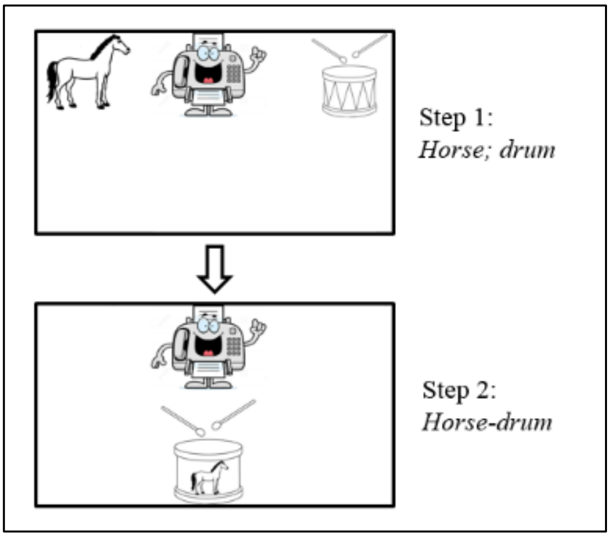


Figure 8. To elicit the compound "horse-drum", a horse and a drum were first presented on either side of the robot. The horse and the drum were then combined into a new item "horse-drum" by the robot.

Each participant produced one token per item (16 test items per participant in average). A total of 1600 tokens were collected, in which 119 tokens were excluded from the analysis due to the poor acoustic quality. This resulted in 1481 tokens included in the analysis (NH children: 698 tokens; CI children: 783 tokens).

Results

Group comparison

Age at implantation

Figure 9 illustrates the normalized tone sandhi pitch contours from the NH 3-year-olds and the CI groups with different ages at implantation. Both the NH children and the early implantation group (CI_Imp 1-2) applied the tone sandhi process in both the full sandhi and half sandhi contexts, producing a **rising T2** for in full sandhi syllables (**T3**+T3) and a **low-falling** pitch in half sandhi context (**T3**+T1/2/4), but it appeals that CI_Imp 1-2's sandhi productions were flatter than those of NH children. The other CI groups, in contrast, generally produced an unchanged (falling/level) pitch for the tone sandhi syllables across contexts.

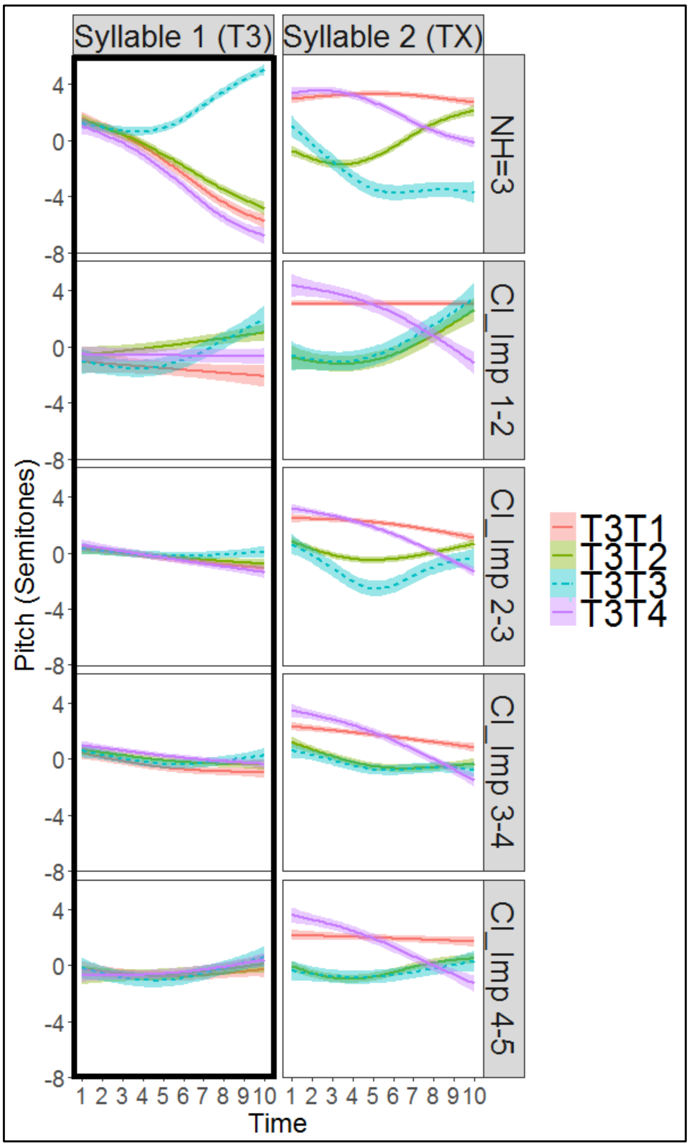


Figure 9. Pitch contours for the disyllabic tone sandhi productions for NH 3-year-olds and CI groups with different ages at implantation. For each word, the first syllable was the tone sandhi syllable that underwent the tone change.

To examine the pitch realization of the tone sandhi syllables from children with different ages at implantation, a linear mixed-effects model was performed on the pitch contour (slope and curvature) of the tone sandhi syllable (Syllable 1) in each word, with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Context" (full sandhi and half sandhi), a covariate "Chronological age" and two random factors "Participant" and "Item" (16 tone sandhi words). The results showed that there were significant interactions of "Group × Context" on the pitch slope ($F(4, 14715) = 4.53$, $p < 0.01$; Supplementary material S25), indicating that there were pitch slope differences across groups and contexts. Two Tukey-HSD post-hoc tests were then performed to compare the pitch difference of the tone sandhi syllables between contexts and between groups. The results showed that the NH children and the early implantation group (CI_Imp 1-2) correctly applied the tone sandhi process in both full and half sandhi contexts, while the other CI groups produced an unchanged pitch contour (i.e., a falling/level pitch) of tone sandhi syllables across contexts (Supplementary material S26). Moreover, relative to NH children, pitch contours of both full and half sandhi syllables produced by CI groups were flatter, including the early implantation group (Supplementary material S27).

CI experience

Figure 10 illustrates the normalized tone sandhi pitch contours for both the NH and CI children, showing that, even for the longer CI experience group (CI_Exp 3-4), children with CIs generally produced an unchanged (falling/level) pitch in both tone sandhi contexts.

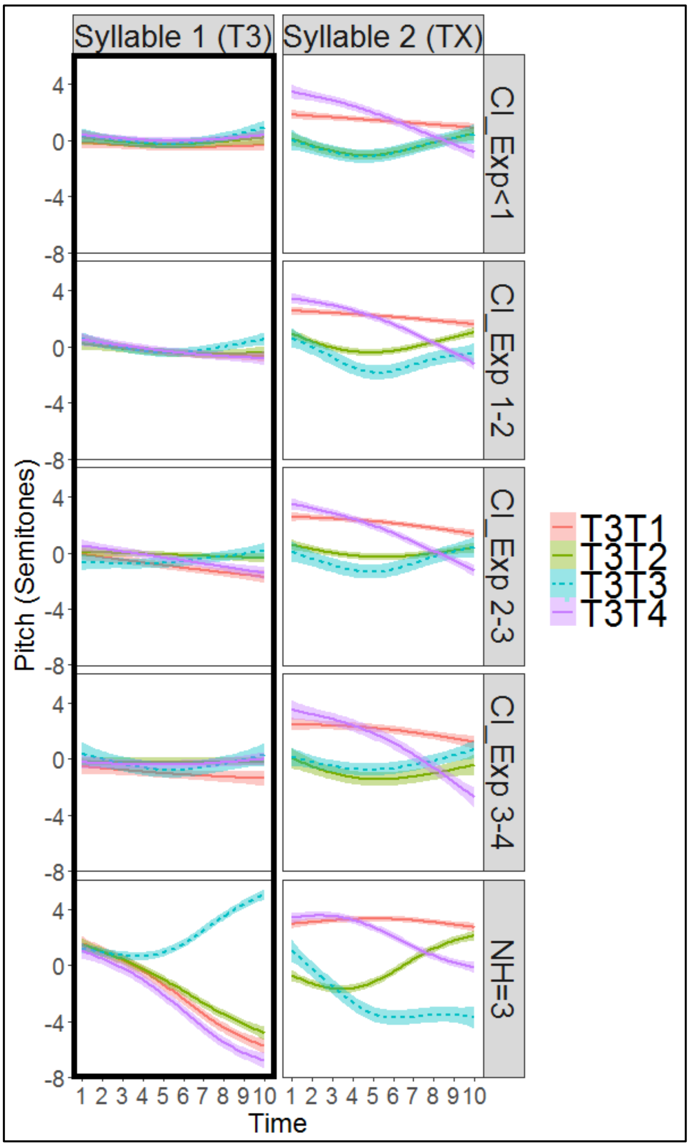


Figure 10. Pitch contours of disyllabic tone sandhi productions for NH 3-year-olds and CI groups with different years of CI experience. For each word, the first syllable was the tone sandhi syllable that underwent the tone change.

To examine the pitch realization of the tone sandhi syllables for the children with different years of CI experience, a linear mixed-effects model was performed on the pitch contour (slope and curvature) of the tone sandhi syllable (Syllable 1) in each word, with two fixed factors "Group" (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp

3-4, NH=3) and "Context" (full sandhi and half sandhi), a covariate "Chronological age" and the two random factors "Participant" and "Item" (16 tone sandhi words). The results show that there were significant interactions of "Group × Context" for pitch slope ($F (4, 14716) = 5.21$, $p < 0.001$; Supplementary material S28). Two Tukey-HSD post-hoc tests then compared the pitch difference for the tone sandhi syllables (1) between contexts and (2) between groups. The results showed that only the NH children applied the tone sandhi process in response to tonal context, while all CI groups produced an unchanged pitch contour (i.e., a falling/level pitch) for both contexts (Supplementary material S29). Moreover, none of the CI groups produced pitch contours on the tone sandhi syllables that approximated those of the NH children (Supplementary material S30).

Regression analysis

A linear regression model was performed on the pitch slope of CI groups' production of tone sandhi syllables (Syllable 1 in each word), with five predictors: "context (full sandhi and half sandhi)", "age at implantation", "CI experience", "chronological age" and "residual hearing". The results showed that age at implantation, CI experience, chronological age and context were significant predictors in predicting these children's tone sandhi productions. Among these predictors, "age at implantation" exhibited larger effects than "CI experience" and "chronological age", as reflected by the larger standardized coefficients (β: age at implantation = 0.123, CI experience = 0.113, chronological age = -0.065; see Supplementary material S31 for more detail).

Summary

The results of Experiment 3a showed that only the early implantation group (CI_Imp 1-2) showed evidence of producing the tone sandhi process, with a rising vs. low falling pitch for the full vs. half sandhi contexts, respectively; however, their (full and half) sandhi productions were flatter relative to those of the NH children. The other CI groups, including the long CI experience group, generally produced an unchanged pitch for the tone sandhi syllables irrespective of the tonal context, indicating that they had not acquired the tone sandhi process. These results implied that children with CIs might face greater challenges in recursively applying tone sandhi processes on larger, trisyllabic units.

Experiment 3b: tone sandhi in trisyllabic compounds

Experiment 3b examined children's tone sandhi productions in larger prosodic units, i.e., trisyllabic T3T3TX compounds. We asked whether children with CIs were able to correctly apply the tone sandhi process on trisyllabic compounds, i.e., applying tone sandhi once for double-T3 sequences (T3T3T1/2/4) and twice for the triple-T3 sequence (T3T3T3), resulting surface tones to be **T2**T3T1/2/4 and **T2T2**T3, respectively. Based on the results of Experiment 3a, we hypothesized that only the early implantation group (CI_Imp 1-2) might be able to correctly apply the tone sandhi process on trisyllabic compounds with double- and triple-T3 surface tones, as only they had partially acquired the tone sandhi process with the simpler, disyllabic words.

Stimuli

Four trisyllabic T3T3TX compounds were used as stimuli (Table 3). The first two syllables were always "zi3 ma3" *purple-horse*, while the last syllables varied from T1 to T4, i.e., "shu1" *book*, "qiu2" *ball*, "gu3" *drum* and "hua4" *painting*.

Table 3. Stimuli use in Experiment 3b: four novel trisyllabic tone sandhi compounds with underlying tones T3T3TX.

Words	Underlying tones	Surface tones	Meaning	Phonetic transcription
zi3 ma3 shu1	T3T3T1	**T2**T3T1	*Purple-horse book*	/tsi3 ma3 ʂu1/
zi3 ma3 qiu2	T3T3T2	**T2**T3T2	*Purple-horse ball*	/tsi3 ma3 tɕiu2/
zi3 ma3 gu3	T3T3T3	**T2T2**T3	*Purple horse-drum*	/tsi3 ma3 ku3/
zi3 ma3 hua4	T3T3T4	**T2**T3T4	*Purple-horse painting*	/tsi3 ma3 hua4/

Procedure

The task began with a non-sandhi practice trial, e.g., "hong2 mao1 shu1" *red-cat book*, followed by four test trials with a randomized order across participants. Again, the participants were trained on how to compose novel trisyllabic compounds in the practice phase (with feedback provided) before moving on to the test phase.

To elicit the target item "zi3 ma3 gu3" *purple-horse drum*, for instance, a purple-horse and a drum appeared on opposite sides of the robot (step 1, figure 11).

The experimenter asked the participant to name both items, as they served as inputs for the trisyllabic word. Next, the experimenter pressed a button to play an animation of the robot (jiggling) to combine the purple-horse and the drum into a new object, i.e., a purple-horse drum (step 2, figure 11). The experimenter then asked the participant to name the trisyllabic novel compound.

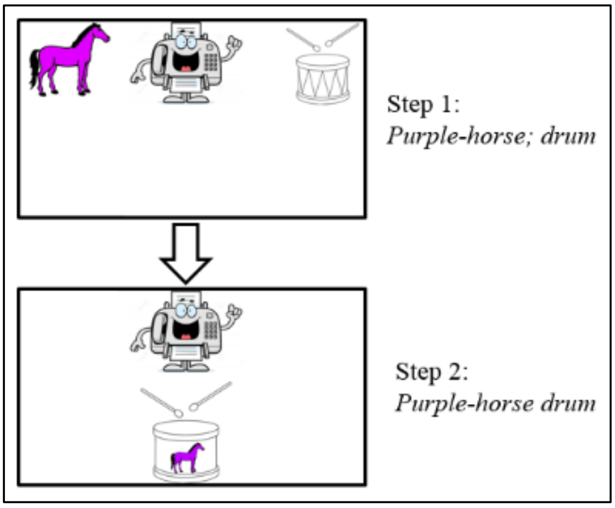


Figure 11. To elicit the trisyllabic compound "purple-horse drum", a purple horse and a drum were firstly presented on opposite sides of the robot. Participants were asked to produce both nouns. The purple-horse and the drum were then combined into a new item by the robot, and participants were then asked to produce the novel compound "purple-horse drum".

Each participant produced one token per item. A total of 382 tokens were collected, with 22 tokens excluded from the analysis due to the poor acoustic quality. This resulted in 360 tokens included in the analysis (NH children: 170 tokens; CI children: 190 tokens).

Results

Group comparison

Age at implantation

Figure 12 illustrates the pitch contour of the trisyllabic tone sandhi compounds from the NH children 3-year-olds and the CI groups at different ages of implantation. The NH children correctly applied the tone sandhi process in response to the tonal context, i.e., applying the tone sandhi process once for double-T3 sequences (T3T3T1/2/4) and twice for the triple-T3 sequence (T3T3T3) resulting in the surface tones **T2**T3T1/2/4 and **T2T2**T3 respectively[18]. The early implantation group (CI_Imp 1-2) differentiated between the two contexts, but did not produce the pitch contour of tone sandhi syllables in the same manner as the NH children. The other CI groups, however, generally produced a level pitch contour for tone sandhi syllables irrespective of the tonal context.

[18] Note that in the phonetic implementation of T2T2T3 sequences, i.e., the surface tone of T3T3T3 items, the rising pitch component of syllable 2 is much reduced due to the coarticulation effect, as suggested by Xu (1994).

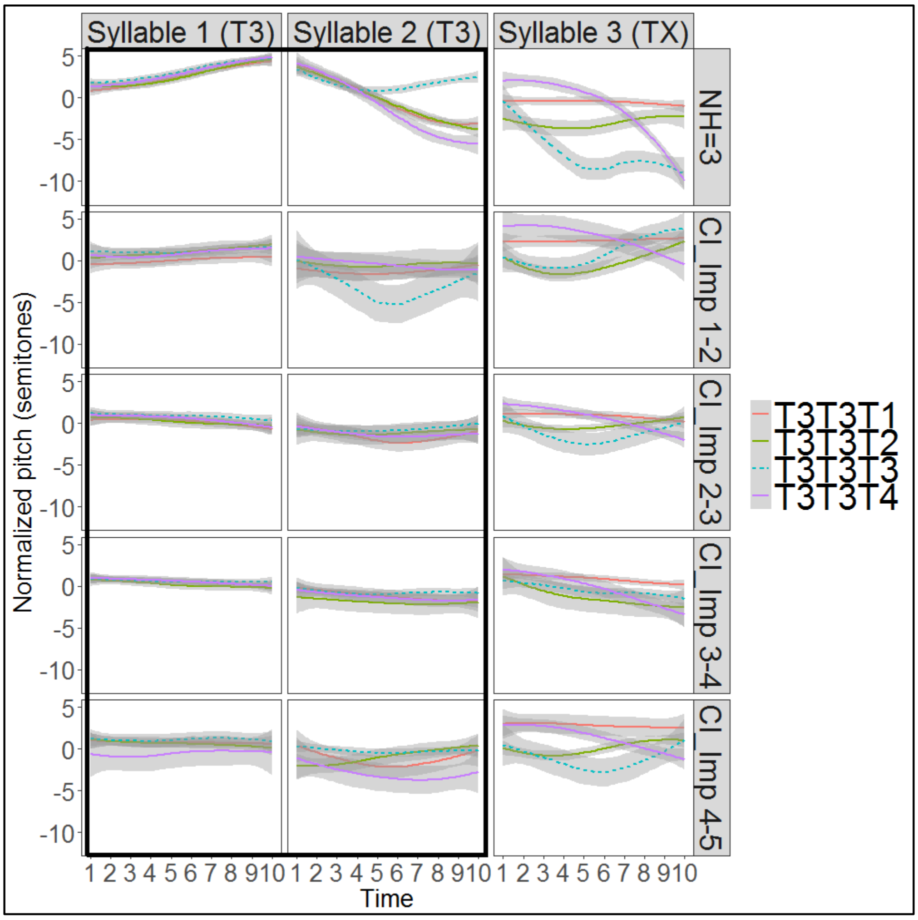


Figure 12. Pitch contours for the trisyllabic tone sandhi productions from NH 3-year-olds and CI groups at different ages of implantation. The target surface tones are **T2T3**T1/2/4 for double-T3 (T3T3T1/2/4) items and **T2T2**T3 for triple-T3 (T3T3T3) items. The first and second syllables (as marked by the black line) were the target syllables included in the analysis.

To investigate the pitch realization of the tone sandhi productions from children with different ages of implantation, a linear mixed-effects model was performed on the pitch contour (slope and curvature) of children's productions of syllables 1 and 2, Two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Context" (double-T3 and triple-T3), a covariate "Chronological age" and two random factors "Participant" and "Item" (four tone sandhi items) were entered into the model. The results showed that there were significant interactions of "Group × Context" on the pitch slope of Syllable 1 (F (4 5753) = 5.66, $p < 0.001$) and pitch slope and curvature of Syllable 2 (slope: F (4, 5734) = 5.56, $p < 0.001$; curvature: F (4, 5722) = 3.14, $p < 0.05$; Supplementary material S32). Tukey-HSD post-hoc tests were performed to compare the pitch slope of Syllables 1 and 2 (1) between contexts and (2) between groups. The results of the two tests show that: (1) NH children correctly changed the surface tone of Syllables 1 and 2 in response to the tonal context, i.e., producing **T2T2**T3 for T3T3T3 items and **T2**T3T1/2/4 for T3T3T1/2/4 items; (2) CI groups generally produced a level or dipping pitch (CI_Imp 1-2's Syllable 2 productions) for these two syllables irrespective of the tonal context (Supplementary materials S33 and S34).

CI experience

Figure 13 illustrates the pitch contour of trisyllabic tone sandhi compounds from NH children (3-year-olds) and CI groups with different years of CI experience. The CI groups generally produced a level pitch contour for the tone sandhi syllables 1 and 2, irrespective of the tonal context.

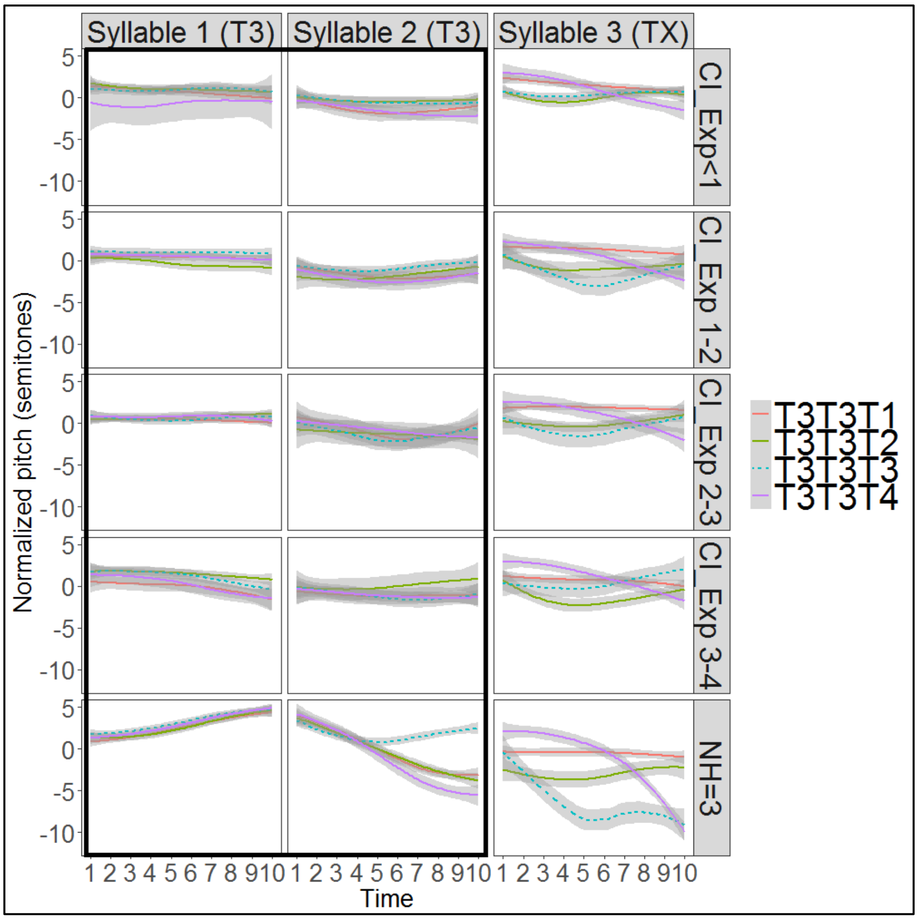


Figure 13. Pitch contours of trisyllabic tone sandhi productions from NH 3-year-olds and CI groups with different years of CI experience. The target surface tones are **T2T3**T1/2/4 for double-T3 (T3T3T1/2/4) items and **T2T2**T3 for triple-T3 (T3T3T3) items. The first and second syllables (as marked by the black line) were the target syllables included in the analysis.

To investigate the pitch realization of tone sandhi productions from children with different years of CI experiences, a linear mixed-effects model was performed on the pitch contour (slope and curvature) of children's productions of syllables 1 and 2 with each word, as this syllable underwent pitch variations across tonal contexts, i.e., **low-falling** or **rising** in T3T3T1/2/4 (surface tones **T2**T3T1/2/4) or T3T3T3 (surface tones **T2T2**T3) contexts. Two fixed factors "Group" (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3) and "Context" (before T1/2/4 and before T3), a covariate

“Chronological age” and two random factors “Participant” and “Item” (four tone sandhi items) were entered into the model. The results show that there was a significant interaction of “Group × Context” on the pitch slope for both syllables (Syllable 1: $F (4, 5754) = 5.55$, $p < 0.001$; Syllable 2: $F (4, 5752) = 3.36$, $p < 0.01$; Supplementary material S35). The Tukey-HSD post-hoc tests on this interaction show that only the NH children correctly changed the surface tone of syllables 1 and 2 in response to the tonal context, while other CI groups generally produced a level/falling pitch for both syllables in both contexts (Supplementary materials S36 and S37).

Regression analysis

Linear regression models were performed on the pitch slope of CI groups’ production of trisyllabic tone sandhi syllables (syllables 1 and 2 for each word), with five predictors: “context (double-T3 and triple-T3)”, “age at implantation”, “CI experience”, “chronological age” and “residual hearing”. The results showed that, “context”, “age at implantation”, “CI experience” and “chronological age” were all significant predictors in predicting these children’s trisyllabic tone sandhi productions. Among these predictors, “age at implantation” exhibited larger effects than “CI experience” and “chronological age”, as reflected by the larger standardized coefficients (β: age at implantation = -0.3 and -0.08 for syllables 1 and 2, CI experience = -0.14 and 0.05 for syllables 1 and 2, chronological age = 0.12 and -0.01 for syllables 1 and 2; see Supplementary material S38 for more detail).

Summary

The results of Experiment 3b showed that children with CIs were not able to correctly apply the tone sandhi process to trisyllabic compounds, generally producing a level or dipping pitch for tone sandhi syllables irrespective of the tonal context.

Even the early age at implantation group (CI_Imp 1-2) mis-applied the tone sandhi process on these compounds and produced (atypical) dipping or level tones.

General discussion

This study examined the acquisition of the Mandarin tonal system by Mandarin-learning children using CIs, including lexical tones in single words and the neutral tone and tone sandhi processes in connected speech. We asked whether earlier age at implantation and/or longer CI experience would help children with CIs to produce normal-like lexical tone productions, and master the neutral tone category and tone sandhi processes. The group comparison between NH and CI groups showed that only those who were implanted earlier, i.e., between 1 to 2 years of age, produced normal-like lexical tones, and generally had contextual tones approximating those of NH 3-year-olds, at least in some conditions, i.e., (familiar) neutral tone and (short) disyllabic tone sandhi items. The other CI groups, including those with 3-4 years of CI experience (most of these children were not early implanted), had not developed a normal-like tonal system, producing flatter pitch contours for lexical tones and unchanged pitch realizations for contextual tones. The results of the regression analysis also showed that the effect of the age at implantation was larger than that of the length of CI experience in predicting the tonal abilities of these children with CIs. These results provide acoustic evidence to confirm that, for Mandarin-learning children using CIs, earlier age at implantation is critical for developing the typical tonal system, including both lexical tones and contextual tones.

Our results confirm previous findings that children with CIs generally face challenges in acquiring and producing normal-like **lexical tones** (Tan, Dowell & Vogel, 2016). Relative to NH children, the children with CIs in the present study generally produced flatter lexical tone pitch contours, consistent with findings from previous acoustic investigations (Xu, 2004), indicating that these pre-schoolers have not developed normal-like lexical tone categories. Our results also extend previous findings suggesting that both early implantation and longer CI experience help children with CIs to improve the accuracy of their tonal productions (e.g. Han et al., 2007). However, we found that, relative to longer CI experience, early implantation is more important in helping children with CIs to develop normal-like lexical tone categories. The critical role of early implantation for the acquisition of lexical tones might be related to the fact that lexical tone acquisition typically occurs at an early stage during language development in children with NH, where lexical tones are acquired at the single word stage of development (Li & Thompson, 1977). However, it should also be noted that the lexical tone productions from the early implanted children in this study (implanted before age 2; mean age: 4;7, SD: 10 months) were still more variable than the NH 3-year-old controls (mean age: 3;8, SD: 3 months). This shows that these early implanted children were still behind their NH peers in their tonal development, suggesting that even earlier implantation may be required, i.e., within the first year of life.

Nonetheless, the effect of CI experience was significant in predicting the tonal productions from children with CIs, albeit to a smaller extent than age at implantation. It is possible that, after a longer period of CI experience, children with CIs may still acquire normal-like lexical tone productions. A recent study from Li et al. (2017) suggests that 8-18-year-olds with 6 to 15 years CI experience who were implanted at 1 to 5 years of ages, can reach a relatively high accuracy in tone production (90.08%, SD = 6.64%). In the present study, the length of participants' CI use was up to 4 years. It is possible that 4 years of CI experience is not enough for these children to develop normal-like lexical tone categories. Future studies will need to test children with longer CI experience to further explore the effect of the longer CI experience on the acquisition of normal-like lexical tone productions.

However, caution should be taken in interpreting the effect of early implantation and long CI experience on the tonal acquisition, as these two factors are often correlated. This was true in the current study, in which there was a strong correlation between age at implantation and length of CI experience: $r = -0.53$, $p < 0.001$, and the combined effects of these two factors were significant in the regression analysis in all experiments. The seven early-implanted children also had a relatively long experience of CI use, e.g., two had more than 3 years of CI experience, and the other five had 2 to 3 years CI experience. It was therefore hard to disentangle the effects of early implantation on tonal acquisition from that of longer CI experience. For example, in the longer CI experience group (i.e., CI_Exp 3-4), only two of the seven children had early implantation (between 1 to 2 years); the other five were implanted between 2 to 3 years. Thus, without early implantation, longer CI experience alone may not be sufficient for children with CIs to acquire the tonal system. Perhaps, then, it is the combined effect of both early age at implantation and longer CI experience that predict better outcomes in children's tonal acquisition. It has also been suggested that, if children with CIs were implanted before 2.5 yrs., the benefit of early implantation would combine with the impact of increased CI experience to provide an additional advantage for language development; however, this added advantage diminishes systematically with increasing age at implantation (Connor et al., 2006). To gain a better understanding of how these two factors contribute to and/or interact with the development of tone by children with CIs, it would be helpful in future studies to test more children with early implantation *and* longer CI experience.

Our results on **neutral tone** and **tone sandhi** have extended previous research on children with CIs from *lexical tones in single words* to *contextual tone processes in connected speech*. However, it was found that the children with CIs generally produced both neutral tone and tone sandhi with unchanged/fixed pitch patterns, irrespective of the tonal context; only the early implantation group (CI_Imp 1-2), who had already developed the correct lexical tone categories, were able to correctly produce contextual tones, and then only in the "simple" conditions, i.e., **familiar** neutral tone words and **short** (disyllabic) tone sandhi compounds.

These findings suggested that building lexical tone categories is an essential prerequisite for acquiring contextual tones (see also Demuth, 1993). Only with well-established lexical tone categories can tonal processes then be learned, including an understanding of which tonal contexts trigger contextually conditioned pitch realizations. It is therefore of interest that even the early-implanted children, who had generally good lexical tone realizations, could not produce contextual tones for the **unfamiliar** neutral tone words or the **longer** (trisyllabic) tone sandhi compounds. This suggests that these early-implanted children, even with 2.5-4 years of CI experience, are not yet performing like their NH 3-year-old peers (cf. Tang et al. (2018), Tang et al. (submitted)), failing to generalize the neutral tone category to novel items they have not heard before, and failing to apply the tone sandhi process to larger (potentially harder) prosodic units. This suggested that, to fully master contextual tones and build robust tonal representations, children with CIs may need to receive implantation even earlier, perhaps before age 1. Several studies from other non-tonal languages have demonstrated that children receiving implantation within the first year of life showed greater advantages in language development than those who were implanted later than age 1 (Colletti et al., 2005; Waltzman & Roland, 2005; Dettman et al., 2007; Leigh et al., 2013; Ching et al., 2018). In China, however, only about 5% of children with CIs have been implanted before the age of 1 (Liang & Mason, 2013). Given the importance of both lexical tone and tonal process in connected speech in Mandarin Chinese, more effort is needed to facilitate the implementation of early newborn hearing screening and implantation for Mandarin-learning children with profound hearing loss.

Our results also raise many questions about the language input to the children in this study, and how this might influence both their tonal development and language development more generally. Relative to the NH children in this study, the children with CIs generally produced lengthened neutral tone durations. This is not surprising as it is known that children with CIs generally produce speech at a slower speaking rate, with longer vowel, word and sentence durations (Uchanski & Geers, 2003). For Mandarin-learning children with CIs, however, the lengthened neutral tone duration might also be attributed to the teachers/therapists' exaggerated manner of speaking, in which the weaker syllables carrying neutral tone might be emphasized and lengthened. It was observed, during data collection for this study, that these children also tended to speak slowly or even syllable-by-syllable, rather than fluent running speech. This might result in children with CIs' misapplication of tone sandhi in trisyllabic compounds, even for the early implanted children[19]. If the slow speaking

[19] We found that the speaking rate (seconds per syllable) of children with CIs' trisyllabic tone sandhi productions was slower than that of NH 3-year-old children (CI: 0.69 second/syllable; NH: 0.37 second/syllable; CI vs. NH: $t(88) = 7.671$, $p < 0.001$), and this was not correlated with the age at implantation (Pearson correlation: $r = -0.19$, $p = 0.104$) or the length of CI experience ($r = 0.11$, $p = 0.463$), suggesting that even children with early implantation and with longer CI experience still exhibit slower speaking rate. Moreover, the speaking rate of children with CIs' was moderately but

rate is due to the nature of the input from teachers/therapists, this might delay the learning of connected speech processes. This suggests that future studies might want to examine therapist/child interactions, to better understand the nature of the input they receive, and the potential effects this might have for learning not only lexical tones, but also tones in context.

Finally, there are some limitations in the present study. Due to children's limited attention span, only one token was included for each lexical tone[20], and a few for each of the various tonal conditions tested; further study could focus on only one of the experiments presented in this study, but in more depth. In addition, given that most children with CIs in China are not implanted early, this was a population study; future study could focus on a more homogeneous, early-implanted group, with equal numbers of children in each group. Both will eventually be needed to confirm the reliability and generalizability of the current results. This could also address some of the inconsistency in results between the current acoustic investigation and previous research using perceptual coding, where we found a lack of effect of longer CI experience on children's lexical tone acquisition. Future study could, for example, conduct a follow-up study using perceptual coding on the current data, to examine the reliability and consistency between acoustic approaches and perceptual coding.

significantly correlated with the slope of Syllable 1 of these words ($r = 0.33$, $p < 0.001$), suggesting that the slow speaking rate might explain their mis-application of tone sandhi to some degree.

[20] To compensate for the limited number of lexical tone items, we conducted an additional analysis on Syllable 1 (carrying T1-4) of neutral tone words, to validate the results of Experiment 1. The results are consistent with that in Experiment 1 (see Supplementary material S39).

Conclusion

The present study explored the acquisition of the Mandarin tonal system by children with CIs, including both lexical tones and tones in context (neutral tone and tone sandhi). Acoustic analysis showed that only those children implanted before 2 years of age had developed a normal-like tonal system, producing pitch contours for both lexical and contextual tones that were comparable to that of NH 3-year-olds. However, their lexical tone productions were more variable than those of the NH children, and they had not fully acquired tonal processes. Children who were implanted later, even those with 3 to 5 years of CI experience, did not produce typical tonal productions. These findings provide acoustic evidence confirming that earlier age at implantation helps children with CIs develop not only typical *lexical* tone categories, but also facilitates their implementation of tones in context.

Acknowledgements

We thank Robert Cowan and the reviewers for providing useful comments on the manuscript and Xin Cheng for helping with the reliability check. We also thank the help from China Rehabilitation Research Center for Hearing and Speech Impairment (CRRCHSI), Beijing Piaoliangmama (漂亮妈妈) Rehabilitation Center for Hearing and Speech Impairment and Beijing Xiaojudeng (小桔灯) Rehabilitation Center for Hearing and Speech Impairment in data collection. The authors acknowledge the financial support of the HEARing CRC, established under the Australian Government's Cooperative Research Centres (CRC) Program. The CRC Program supports industry-led collaborations between industry, researchers and the community. This research was also supported by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE1101021), ARC

FL130100014, and, in part, by a Macquarie University iMQRES scholarship to the first author. The equipment was funded by MQSIS 9201501719.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear Mixed-effects Models sing Eigen and S4*. R package version, 1 (7), 1-23.

Boersma, P., & Weenink, D. (2016). *Praat: Doing Phonetics by Computer*. [Computer program]. Version 6.0. 19.

Chen, M. Y. (2000). *Tone Sandhi: Patterns Across Chinese Dialects* (Vol. 92). Cambridge University Press. https://doi.org/10.1017/CBO9780511486364

Chen, Y., & Wong, L. L. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology*, 56(sup2), S7-S16. https://doi.org/10.1080/14992027.2017.1300694

Ching, T. Y., Dillon, H., Leigh, G., & Cupples, L. (2018). Learning from the Longitudinal Outcomes of Children with Hearing Impairment (LOCHI) study: Summary of 5-year findings and implications. *International Journal of Audiology*, 57(sup2), https://doi.org/105-111. 10.1080/14992027.2017.1385865

Colletti, V., Carner, M., Miorelli, V., Guida, M., Colletti, L., & Fiorino, F. G. (2005). Cochlear implantation at under 12 months: report on 10 patients. *The Laryngoscope*, 115(3), 445-449. https://doi.org/10.1097/01.mlg.0000157838.61497.e7

Connor, C. M., Craig, H. K., Raudenbush, S. W., Heavner, K., & Zwolan, T. A. (2006). The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there an added value for early implantation?. *Ear and Hearing*, 27(6), 628-644. https://doi.org/10.1097/01.aud.0000240640.59205.42

Demuth, K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language*, 20(2), 275-301. https://doi.org/10.1017/S030500090000828X

Deng, X. & Yip, V. (2018). A multimedia corpus of child Mandarin: The Tong corpus. *Journal of Chinese Linguistics*, 46(1): 69-92. https://doi.org/10.21415/T5PC7Q

Dettman, S. J., Pinder, D., Briggs, R. J., Dowell, R. C., & Leigh, J. R. (2007). Communication development in children who receive the cochlear implant younger than 12 months: risks versus benefits. *Ear and Hearing*, 28(2), 11S-18S. https://doi.org/10.1097/AUD.0b013e31803153f8

Duanmu, S. (2012). Word-length preferences in Chinese: a corpus study. *Journal of East Asian Linguistics*, 21(1), 89-114. https://doi.org/10.1007/s10831-011-9087-y

Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., & Xu, L. (2007). Tone production of Mandarin Chinese speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 71(6), 875-880. https://doi.org/10.1016/j.ijporl.2007.02.008

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, 27(1), 3-42. https://doi.org/10.1017/S030500099900402X

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R Package Version*, 2(0).

Leigh, J., Dettman, S., Dowell, R., & Briggs, R. (2013). Communication development in children who receive a cochlear implant by 12 months of age. *Otology & Neurotology*, 34(3), 443-450. https://doi.org/10.1097/MAO.0b013e3182814d2c

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1-33. https://doi.org/10.18637/jss.v069.i01.

Liang, Q., & Mason, B. (2013). Enter the dragon–China's journey to the hearing world. *Cochlear Implants International*, 14(sup1), 26-31. https://doi.org/10.1179/1467010013Z.00000000080

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199. https://doi.org/10.1017/S0305000900001598

Li, Y. L., Lin, Y. H., Yang, H. M., Chen, Y. J., & Wu, J. L. (2018). Tone production and perception and intelligibility of produced speech in Mandarin-speaking cochlear implanted children. *International Journal of Audiology*, 57(2), 135-142. https://doi.org/10.1080/14992027.2017.1374566

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494-1502. https://doi.org/10.3758/s13428-016-0809-y

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Chapman & Hall/CRC.

Peters, J., Hanssen, J., & Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics*, 46, 185-209. https://doi.org/10.1016/j.wocn.2014.07.004

R Core Team (2016). "R: A language and environment for statistical computing" [Computer program]. Version 3.3.1, https://www.R-project.org/ (Last viewed June 21, 2016).

Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, 20, 81-124.

Tan, J., Dowell, R., & Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American Journal of Audiology*, 25(3), 246-256. https://doi.org /10.1044/2016_AJA-15-0069

Tang, P. (2014). A study of prosodic errors of Chinese neutral tone by advanced Japanese students (Chinese version). *TCSOL Studies*, 56(4), 39-47. https://doi.org/10.16131/j.cnki.cn44-1669/g4.2014.04.005

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017a). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(2), 493-503. https://doi.org/10.1121/1.4995998

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017b). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(5), 2823-2835. https://doi.org/10.1121/1.5008372

Tang P, Yuen I, Xu Rattanasone N, Gao L, Demuth K (2018). Acquisition of weak syllables in tonal languages: acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, 1–27. https://doi.org/10.1017/S0305000918000296

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (submitted) The acquisition of phonological alternations: the case of the Mandarin tone sandhi process.

Uchanski, R. M., & Geers, A. E. (2003). Acoustic characteristics of the speech of young cochlear implant users: A comparison with normal-hearing age-mates. *Ear and Hearing*, 24(1), 90S-105S. https://doi.org/10.1097/01.AUD.0000051744.24290.C1

Vandali, A. E., & van Hoesel, R. J. (2012). Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking. *The Journal of the Acoustical Society of America*, 132(1), 392-402. https://doi.org/10.1121/1.4718452

Waltzman, S. B., & Roland, J. T. (2005). Cochlear implantation in children younger than 12 months. *Pediatrics*, 116(4), e487-e493. https://doi.org/10.1542/peds.2005-0282

Xu, L., Li, Y., Hao, J., Chen, X., Xue, S. A., & Han, D. (2004). Tone production in Mandarin-speaking children with cochlear implants: a preliminary study. *Acta Oto-laryngologica*, 124(4), 363-367. https://doi.org/10.1080/00016480410016351

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4), 2240-2253. https://doi.org/10.1121/1.408684

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83. https://doi.org/10.1006/jpho.1996.0034

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' Acoustic Realization of Mandarin Tone Sandhi and Lexical Tones in Context are Not Yet Fully Adult-like. *Frontiers in Psychology*, 9, 817.

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Zhou, N., Huang, J., Chen, X., & Xu, L. (2013). Relationship between tone perception and production in prelingually-deafened children with cochlear implants. *Otol Neurotol*, 34(3), 499-506. https://doi.org/10.1097/MAO.0b013e318287ca86

Supplementary materials

S1. Participant demographic information.

Participant number	Gender	Residual hearing (L/R in dB)	Type of Device(s)	CI Brand	Age at implantation (month)	CI experience (month)	Age (months)	Age at using hearing aids (month)	Hearing aid experience before CI surgery (month)
1	M	95/95	CI+HA	MEDEL	13	29	42		
2	M	110/120	CI+HA	MEDEL	15	35	50		
3	M	100/100	CI+HA	MEDEL	18	39	58		
4	M	110/110	CI+HA	Cochlear	18	48	67		
5	M	110/110	CI+HA	Advanced Bionics	18	31	50	17	1
6	M	80/100	CI+HA	MEDEL	22	33	55		
7	M	100/110	CI+HA	Cochlear	23	28	51		
8	F	90/90	CI+HA	MEDEL	25	13	39		
9	M	110/110	CI+HA	Cochlear	25	25	50	18	7
10	M	97/97	CI+HA	MEDEL	26	10	37		
11	F	100/100	CI+HA	MEDEL	27	49	76		
12	M	110/110	CI+HA	Cochlear	27	20	47		

13	F	90/100	CI+HA	MEDEL	27	28	55		
14	M	100/110	CI+HA	Cochlear	28	42	70		
15	M	100/85	CI+HA	MEDEL	28	22	50	28	0
16	M	80/90	CI+HA	Advanced Bionics	29	42	71		
17	F	100/110	CI+HA	Advanced Bionics	29	43	73		
18	M	100/100	CI+HA	MEDEL	29	21	50		
19	F	95/95	CI+CI	Advanced Bionics	29	17	47		
20	M	97/97	CI+HA	MEDEL	30	14	45		
21	M	100/100	CI+CI	Mindni	30	7	37		
22	F	110/110	CI+HA	MEDEL	30	18	48		
23	M	105/95	CI+HA	MEDEL	30	22	52		
24	M	110/110	CI+HA	Cochlear	30	22	52		
25	F	95/95	CI+CI	MEDEL	30	34	64		
26	F	110/100	CI+HA	Cochlear	30	28	59	30	0
27	F	106/80	CI+HA	MEDEL	31	29	60		
28	F	96/96	CI+HA	MEDEL	31	16	48		
29	M	100/70	CI+HA	Cochlear	31	25	57		
30	M	110/110	CI+HA	MEDEL	32	14	47		

31	M	97/97	CI+HA	MEDEL	32	26	58		
32	M	97/97	CI+HA	MEDEL	33	11	44		
33	M	97/97	CI+CI	Cochlear	33	18	52		
34	M	115/115	CI+HA	MEDEL	33	26	59	33	0
35	M	100/100	CI+HA	Cochlear	34	18	53		
36	F	97/97	CI+HA	MEDEL	34	10	45		
37	F	110/100	CI+HA	Cochlear	34	20	54	34	0
38	M	100/105	CI+HA	MEDEL	35	35	70		
39	F	95/100	CI+HA	MEDEL	36	22	58		
40	M	98/110	CI+HA	MEDEL	36	22	59		
41	F	97/97	CI+HA	MEDEL	37	9	46		
42	M	100/100	CI+HA	MEDEL	37	24	62	37	0
43	M	110/110	CI+HA	MEDEL	37	28	66	18	19
44	F	100/100	CI+HA	Cochlear	38	9	47		
45	M	100/100	CI+HA	MEDEL	38	23	61		
46	M	110/110	CI+HA	Cochlear	38	28	66		
47	F	100/100	CI+HA	MEDEL	39	19	59	31	8
48	F	100/100	CI+HA	MEDEL	40	11	52		
49	M	80/90	CI+HA	Cochlear	40	29	70		
50	F	100/90	CI+HA	Cochlear	40	16	56		

51	M	110/110	CI+HA	Cochlear	41	20	62		
52	F	80/110	CI+HA	MEDEL	41	25	67	24	17
53	F	100/100	CI+CI	MEDEL	42	20	63		
54	M	60/90	CI+HA	Cochlear	43	11	55		
55	M	90/90	CI+HA	Cochlear	43	16	59		
56	F	96/96	CI+HA	MEDEL	44	6	50		
57	F	100/100	CI+CI	Cochlear	45	20	66		
58	M	100/110	CI+HA	Cochlear	47	29	76		
59	M	95/95	CI+HA	Cochlear	47	13	60	29	18
60	F	96/96	CI+HA	Cochlear	48	20	69		
61	F	90/110	CI+HA	MEDEL	49	7	56	40	9
62	F	100/90	CI+HA	MEDEL	52	14	66		
63	F	100/100	CI+HA	MEDEL	53	22	76		
64	M	100/100	CI+HA	MEDEL	53	22	76		
65	M	100/105	CI+HA	MEDEL	54	21	76		
66	F	110/110	CI+HA	Cochlear	54	2	56		
67	M	90/110	CI+HA	Cochlear	55	16	71		
68	F	110/110	CI+CI	Advanced Bionics	55	11	66		
69	F	100/100	CI+HA	Cochlear	55	10	65		

70	F	80/90	CI+HA	Cochlear	58	10	68
71	M	80/90	CI+HA	Cochlear	61	7	68
72	M	80/90	CI+HA	Cochlear	74	6	80

S2. Additional analysis on children with bilateral CIs

There were seven children fitted with two CIs in this study, including four children from CI_Imp 2-3, two children from CI_Imp 3-4 and one child from CI_Imp 4-5. To explore whether the implantation modality (i.e., bilateral CI+CI vs. bimodal CI+HA) would impact children's tonal acquisition, we compared the lexical tone pitch contours of the four children with unimodal implantation from the CI_Imp 2-3 group with other children with bimodal implantation from the same group (Figure 1). These children were selected to gain a better control on their ages at implantation.

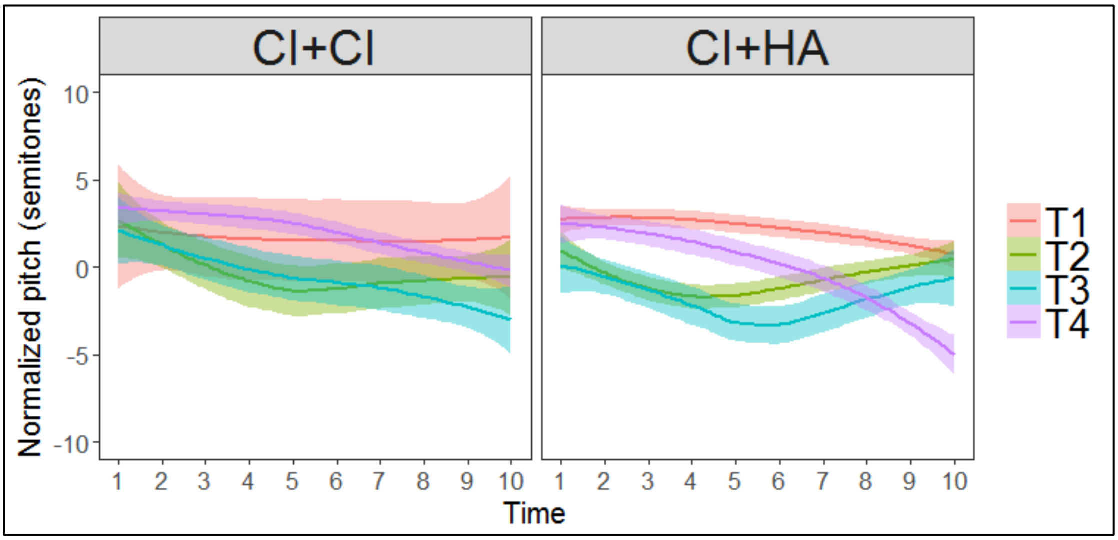


Figure 1. Pitch contours of lexical tones from bilateral (CI+CI) and bimodal (CI+HA) groups.

A linear mixed model was performed on pitch slopes and curvatures with two fixed factors Modality (CI+CI and CI+HA) and Tone (T1, T2, T3 and T4) and a random factor Participant. The results showed that there was a significant interaction of Modality × Tone on the pitch slope, suggesting that these two groups differed in the pitch slope across tones (Table 1).

Table 1. Results of linear mixed-regression model on pitch shape (pitch slope and curvature) of lexical tones, with two fixed factors Modality (CI+CI vs. CI+HA) and Tone (T1, T2, T3 and T4).

Parameter	**Factors**	**df 1**	**df 2**	**F**	***p***
Pitch slope	Modality	1	1137	0.29	0.591
	Tone	**3**	**1137**	**4.31**	**< 0.01****
	Modality × Tone	**3**	**1137**	**3.41**	**< 0.05***
Pitch Curvature	Modality	1	1137	0.00	0.982
	Tone	**3**	**1137**	**3.40**	**< 0.05***
	Modality × Tone	3	1137	1.04	0.372

The post-hoc test shows that the two groups were not different in the pitch slope of T1, T2 and T3, but the bimodal group produced T4 with a more falling pitch slope than the bilateral group (Table 2). This result therefore suggests that the CI+HA bimodal implantation might benefit the tonal acquisition by children with hearing impairment. However, it should be noted that this small group is not representative of most children with CIs in China and we do not have enough children with bilateral implantation to draw a strong conclusion about bimodal benefit. Future studies with a balanced design between group is needed to explore the effect of modality on the tonal acquisition by children with hearing impairment.

Table 2. Pairwise comparisons of pitch slopes of lexical tones between children with bilateral (CI+CI) and bimodal (CI+HA) implantation in the CI_Imp 2-3 group.

Comparison	Tone	***β***	**SE**	**df**	***t***	***p***
Unimodal - Bimodal	T1	-0.32	1.02	51.02	-0.31	0.755
	T2	0.35	1.01	50.85	0.34	0.734
	T3	1.00	1.01	50.85	0.98	0.330
	T4	**2.04**	**1.01**	**50.85**	**2.01**	**< 0.05***

S3. Additional analysis of duration of hearing aid (HA) use before CI surgery on tonal acquisition

Twelve children with CIs also had a record of HA use before CI surgery (age at implantation: 1 child was implanted before age 2, 5 children were implanted from 2 to 3 years, 5 children were implanted from 3 to 4 years and the remaining 1 children was implanted from 4 to 5 years; CI experience: 7 children had 2 to 3 years' CI experience, 4 children had 1 to 2 years' CI experience and 2 children had less than 1 year of CI experience). Six of these children received HA and CI at the same time, and thus their HA experience before CI is 0 month; the other six children received the HAs at 1 to 19 months before the CI surgery.

To explore the effect of the length of HA experience before CI on these children's tonal acquisition, we conducted regression analyses with HA experience before CI and other predictors on their tonal productions. The results are presented in Tables 1 to 5, showing that HA experience before CI was a significant predictor in predicting lexical tone, neutral tone and trisyllabic tone sandhi pitch productions from children with CIs, but with smaller effects than age at implantation, as reflected by the standardized coefficients. This suggests that long HA experience before CI might benefit children with CIs' tonal acquisition, but early implantation is critical for them to acquire typical tone productions.

Table 1. Results of linear regression model of lexical tone pitch slopes with different predictors.

Predictor	***ΔR²***	***B***	***SE (B)***	***β***	***p***
Tone category	0.54	--	--	--	--
HA experience before CI	0.08	0.010	0.003	0.074	<0.05*
Age at implantation	0.28	0.001	0.005	0.306	<0.01**
CI experience	0.09	0.030	0.005	-0.120	<0.05*
Chronological age	0.06	-0.001	0.006	-0.007	<0.05*
Residual hearing	0.04	-0.009	0.003	0.001	0.091

Table 2. Results of linear regression model of neutral tone pitch slope with different predictors.

Predictor	***ΔR²***	***B***	***SE (B)***	***β***	***p***
Context	0.03	0.239	0.042	0.172	<0.05*
Type	0.03	-0.221	0.038	-0.180	<0.05*
HA experience before CI	0.04	0.017	0.002	0.108	<0.05*
Age at implantation	0.03	-0.021	0.004	-0.286	<0.05*
CI experience	0.01	-0.004	0.004	-0.277	0.285
Chronological age	0.04	0.013	0.005	0.108	<0.05*
Residual hearing	0.02	-0.012	0.003	-0.161	<0.05*

Table 3. Results of linear regression model of neutral tone duration with different predictors.

Predictor	***ΔR²***	***B***	***SE (B)***	***β***	***p***
Context	0.03	0.224	0.121	0.149	0.072
Type	0.16	-0.480	0.112	-0.361	<0.001***
HA experience before CI	0.05	0.002	0.010	0.025	0.828
Age at implantation	0.16	0.312	0.163	4.624	<0.05*
CI experience	0.06	0.338	0.163	4.361	<0.05*
Chronological age	0.04	-0.260	0.152	-1.980	0.087
Residual hearing	0.06	0.019	0.011	0.226	<0.05*

Table 4. Results of linear regression model of pitch slope of disyllabic tone sandhi words (Syllable 1) with different predictors.

Predictor	***ΔR²***	***B***	***SE (B)***	***β***	***p***
Context	0.06	0.353	0.023	0.243	<0.001***
HA experience before CI	0.01	0.001	0.001	0.014	0.539
Age at implantation	0.09	0.011	0.002	0.147	<0.001***
CI experience	0.04	0.001	0.002	0.118	<0.001***
Chronological age	0.03	-0.010	0.002	-0.095	<0.001***
Residual hearing	0.03	0.008	0.001	0.002	0.935

Table 5. Results of linear regression model of pitch slope of trisyllabic tone sandhi words (Syllables 1 and 2) with different predictors.

Predictor	**Syllable**	***ΔR²***	***B***	***SE (B)***	***β***	***p***
Context	1	0.01	0.029	0.042	0.03	0.492
	2	0.08	0.355	0.048	0.3	<0.001***
HA experience before CI	1	0.03	0.008	0.003	0.203	<0.001***
	2	0.01	-0.159	0.003	-0.308	<0.001***
Age at implantation	1	0.08	-0.007	0.004	-0.283	<0.05*
	2	0.2	0.042	0.004	0.664	<0.001***
CI experience	1	0.01	0.005	0.003	-0.265	0.154
	2	0.01	-0.007	0.004	-0.085	0.069
Chronological age	1	0.07	-0.023	0.004	-0.131	<0.001***
	2	0.01	-0.009	0.005	-0.098	0.050
Residual hearing	1	0.05	-0.013	0.003	0.072	<0.001***
	2	0.01	0.001	0.003	0.016	0.732

S4. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's lexical tone productions, with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Tones" (T1, T2, T3 and T4) and a covariate "Chronological age".

Parameter	Factors	df 1	df 2	F	*p*
Pitch slope	Group	4	145	0.83	0.508
	Tone	**3**	**5882**	**4.47**	**<0.01****
	Chronological age	1	145	3.05	0.083
	Group × Tone	**12**	**5890**	**3.64**	**<0.001*****
	Group × Chronological age	4	145	0.65	0.628
	Tone × Chronological age	3	5882	0.82	0.480
	Group × Tone × Chronological age	12	5889	1.19	0.282
Pitch curvature	**Group**	**4**	**151**	**3.10**	**<0.05***
	Tone	**3**	**5882**	**3.73**	**<0.05***
	Chronological age	1	152	2.35	0.127
	Group × Tone	**12**	**5893**	**3.91**	**<0.001*****
	Group × Chronological age	4	151	2.21	0.071
	Tone × Chronological age	3	5882	1.50	0.214
	Group × Tone × Chronological age	12	5891	1.23	0.256

S5. Pairwise comparisons of lexical tone pitch contours (slope and curvature) between groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) for each tone.

Tone	Group difference	Parameter	*β*	SE	df	t	*p*
T1	NH=3 - CI_Imp 1-2	Slope	-0.54	1.20	593	-0.45	0.992
		Curvature	0.06	1.14	777	0.05	1.000
	NH=3 - CI_Imp 2-3	Slope	0.49	0.71	621	0.17	0.979
		Curvature	0.26	0.68	811	0.39	0.995
	NH=3 - CI_Imp 3-4	Slope	0.72	1.06	593	0.68	0.961
		Curvature	-0.32	1.01	777	-0.31	0.998
	NH=3 - CI_Imp 4-5	Slope	-1.64	2.87	592	-0.57	0.979
		Curvature	0.40	2.73	775	0.15	1.000
	CI_Imp 1-2 - CI_Imp 2-3	Slope	2.02	1.32	599	1.54	0.538
		Curvature	0.20	1.25	784	0.16	1.000
	CI_Imp 1-2 - CI_Imp 3-4	Slope	1.26	1.53	591	0.82	0.924
		Curvature	-0.37	1.45	775	-0.26	0.999
	CI_Imp 1-2 - CI_Imp 4-5	Slope	-1.10	3.08	591	-0.36	0.997
		Curvature	0.34	2.92	775	0.12	1.000
	CI_Imp 2-3 - CI_Imp 3-4	Slope	-0.76	1.19	600	-0.64	0.968
		Curvature	-0.58	1.13	786	-0.51	0.986
	CI_Imp 2-3 - CI_Imp 4-5	Slope	-3.12	2.93	593	-1.07	0.823

		Curvature	0.14	2.77	777	0.05	1.000
	CI_Imp 3-4 - CI_Imp 4-5	Slope	-2.36	3.03	591	-0.78	0.937
		Curvature	0.71	2.87	775	0.25	0.999
T2	NH=3 - CI_Imp 1-2	Slope	-0.83	1.20	592	-0.70	0.958
		Curvature	-0.30	1.14	776	-0.27	0.999
	NH=3 - CI_Imp 2-3	**Slope**	**2.69**	**0.70**	**594**	**3.82**	**<0.01****
		Curvature	-0.30	0.67	778	-0.45	0.992
	NH=3 - CI_Imp 3-4	**Slope**	**5.57**	**1.06**	**593**	**5.25**	**<0.001*****
		Curvature	-1.54	1.01	776	-1.53	0.542
	NH=3 - CI_Imp 4-5	**Slope**	**2.02**	**1.87**	**591**	**2.82**	**<0.01****
		Curvature	-0.19	2.73	775	-0.07	1.000
	CI_Imp 1-2 - CI_Imp 2-3	**Slope**	**3.52**	**1.31**	**591**	**2.69**	**<0.05***
		Curvature	0.00	1.24	775	0.00	1.000
	CI_Imp 1-2 - CI_Imp 3-4	**Slope**	**6.40**	**1.53**	**591**	**4.18**	**<0.001*****
		Curvature	-1.24	1.45	775	-0.85	0.914
	CI_Imp 1-2 - CI_Imp 4-5	Slope	0.86	3.08	591	0.28	0.999
		Curvature	0.11	2.92	775	0.04	1.000
	CI_Imp 2-3 - CI_Imp 3-4	**Slope**	**2.88**	**1.19**	**591**	**2.42**	**<0.05***
		Curvature	-1.24	1.13	775	-1.11	0.804

	CI_Imp 2-3 - CI_Imp 4-5	Slope	-2.66	2.92	591	-0.91	0.893
		Curvature	0.11	2.77	775	0.04	1.000
	CI_Imp 3-4 - CI_Imp 4-5	Slope	-5.54	3.03	591	-1.83	0.358
		Curvature	1.35	2.87	775	0.47	0.990
T3	NH=3 - CI_Imp 1-2	Slope	1.46	1.20	595	1.05	0.147
		Curvature	1.43	1.14	780	1.18	0.119
	NH=3 - CI_Imp 2-3	Slope	1.11	0.71	603	1.57	0.517
		Curvature	**4.48**	**0.67**	**789**	**6.68**	**<0.001*** **
	NH=3 - CI_Imp 3-4	Slope	0.36	1.06	596	0.34	0.997
		Curvature	**6.46**	**1.01**	**781**	**6.41**	**<0.001*** **
	NH=3 - CI_Imp 4-5	Slope	4.47	2.88	592	1.56	0.527
		Curvature	**9.73**	**2.73**	**776**	**3.57**	**<0.01** **
	CI_Imp 1-2 - CI_Imp 2-3	**Slope**	**3.57**	**1.31**	**591**	**2.72**	**<0.05***
		Curvature	0.05	1.24	775	0.04	1.000
	CI_Imp 1-2 - CI_Imp 3-4	**Slope**	**2.82**	**1.53**	**591**	**1.84**	**<0.05***
		Curvature	2.03	1.45	775	1.40	0.631
	CI_Imp 1-2 - CI_Imp 4-5	Slope	-2.01	3.08	591	-0.65	0.966
		Curvature	5.30	2.92	775	1.82	0.366
	CI_Imp 2-3 - CI_Imp 3-4	Slope	-0.75	1.19	591	-0.63	0.970

		Curvature	1.98	1.13	775	1.76	0.401
	CI_Imp 2-3 - CI_Imp 4-5	**Slope**	**-5.58**	**2.92**	**591**	**-1.91**	**<0.05***
		Curvature	**5.25**	**2.77**	**775**	**1.89**	**<0.05***
	CI_Imp 3-4 - CI_Imp 4-5	Slope	-4.83	3.03	591	-1.60	0.502
		Curvature	3.27	2.87	775	1.14	0.786
T4	NH=3 - CI_Imp 1-2	Slope	-4.30	2.87	591	-1.49	0.567
		Curvature	-0.92	1.14	776	-0.81	0.927
	NH=3 - CI_Imp 2-3	**Slope**	**-2.15**	**0.70**	**593**	**-3.05**	**<0.01****
		Curvature	**-1.38**	**0.67**	**777**	**-2.06**	**<0.05***
	NH=3 - CI_Imp 3-4	**Slope**	**-2.62**	**1.06**	**592**	**-2.47**	**<0.05***
		Curvature	-1.76	1.01	776	-1.75	0.403
	NH=3 - CI_Imp 4-5	**Slope**	**-3.62**	**1.20**	**592**	**-3.03**	**<0.01****
		Curvature	0.83	2.73	775	0.31	0.998
	CI_Imp 1-2 - CI_Imp 2-3	Slope	1.47	1.31	591	1.13	0.793
		Curvature	-0.45	1.24	775	-0.37	0.996
	CI_Imp 1-2 - CI_Imp 3-4	Slope	1.01	1.53	591	0.66	0.966
		Curvature	-0.84	1.45	775	-0.58	0.978
	CI_Imp 1-2 - CI_Imp 4-5	Slope	-0.67	3.08	591	-0.22	1.000
		Curvature	1.76	2.92	775	0.60	0.975

CI_Imp 2-3 - CI_Imp 3-4	Slope	-0.47	1.19	591	-0.40	0.995
	Curvature	-0.39	1.13	775	-0.34	0.997
CI_Imp 2-3 - CI_Imp 4-5	Slope	-2.15	2.92	591	-0.73	0.949
	Curvature	2.21	2.77	775	0.80	0.932
CI_Imp 3-4 - CI_Imp 4-5	Slope	-1.68	3.03	591	-0.55	0.982
	Curvature	2.60	2.87	775	0.90	0.896

S6. Results of one-way ANOVA on pitch range in children's lexical tone productions, with an independent variable "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and a covariate "Chronological age".

Factor	**Df**	**F**	***p***
Group	**4**	**3.22**	**<0.05***
Chronological age	1	2.80	0.096
Group × Chronological age	4	0.01	1.000
Residuals	141		

S7. Pairwise comparison of lexical tone pitch range between groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5).

Group difference	***β***	**SE**	**df**	**t**	***p***
NH=3 - CI_Imp 1-2	1.54	2.66	141	0.55	0.129
NH=3 - CI_Imp 2-3	**3.36**	**1.61**	**141**	**3.23**	**<0.05***
NH=3 - CI_Imp 3-4	**2.96**	**2.30**	**141**	**2.84**	**<0.05***
NH=3 - CI_Imp 4-5	**4.45**	**6.22**	**141**	**3.63**	**<0.01****
CI_Imp 1-2 - CI_Imp 2-3	-0.48	2.91	141	-0.17	1.000
CI_Imp 1-2 - CI_Imp 3-4	-0.88	3.34	141	-0.26	0.999
CI_Imp 1-2 - CI_Imp 4-5	1.01	6.67	141	0.39	0.995
CI_Imp 2-3 - CI_Imp 3-4	-0.40	2.59	141	-0.15	1.000
CI_Imp 2-3 - CI_Imp 4-5	1.09	6.33	141	0.49	0.988
CI_Imp 3-4 - CI_Imp 4-5	1.48	6.54	141	0.53	0.984

S8. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's lexical tone productions, with two fixed factors "Group" (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) and "Tones" (T1, T2, T3 and T4) and a covariate "Chronological age".

Parameter	**Factors**	**df 1**	**df 2**	**F**	***p***
Pitch slope	**Group**	**4**	**146**	**3.75**	**<0.01****
	Tone	**3**	**5890**	**12.39**	**<0.001*****
	Chronological age	1	146	3.21	0.075
	Group × Tone	**12**	**5897**	**2.20**	**<0.05***
	Group × Chronological age	4	146	1.70	0.153
	Tone × Chronological age	3	5882	0.89	0.445
	Group × Tone × Chronological age	12	5894	1.53	0.106
Pitch curvature	**Group**	**4**	**152**	**2.83**	**<0.05***
	Tone	**3**	**5892**	**16.27**	**<0.001*****
	Chronological age	1	151	2.15	0.145
	Group × Tone	**12**	**5900**	**2.02**	**<0.05***
	Group × Chronological age	4	152	2.2	0.072
	Tone × Chronological age	3	5882	1.46	0.223
	Group × Tone × Chronological age	12	5897	1.69	0.062

S9. Pairwise comparisons of lexical tone pitch contours (slope and curvature) between groups (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) for each tone.

Tone	Group difference	Parameter	*β*	SE	df	t	*p*
T1	NH=3 - CI_Exp<1	Slope	0.12	1.37	612	0.09	1.000
		Curvature	0.66	1.28	841	0.52	0.986
	NH=3 - CI_Exp 1-2	Slope	-0.40	1.45	609	-0.28	0.999
		Curvature	0.75	1.36	837	0.55	0.982
	NH=3 - CI_Exp 2-3	Slope	-1.73	3.21	586	-0.54	0.983
		Curvature	0.74	3.01	808	0.24	0.999
	NH=3 - CI_Exp 3-4	Slope	-0.67	1.23	622	-0.55	0.982
		Curvature	0.48	1.15	854	0.42	0.994
	CI_Exp<1 - CI_Exp 1-2	Slope	-0.52	1.08	580	-0.48	0.989
		Curvature	0.09	1.02	801	0.09	1.000
	CI_Exp<1 - CI_Exp 2-3	Slope	-1.85	3.07	580	-0.60	0.975
		Curvature	0.07	2.87	801	0.03	1.000
	CI_Exp<1 - CI_Exp 3-4	Slope	-0.79	0.76	584	-1.04	0.838
		Curvature	-0.18	0.72	806	-0.26	0.999
	CI_Exp 1-2 - CI_Exp 2-3	Slope	-1.33	3.10	580	-0.43	0.993
		Curvature	-0.02	2.91	801	-0.01	1.000

	CI_Exp 1-2 - CI_Exp 3-4	Slope	-0.28	0.90	583	-0.31	0.998
		Curvature	-0.27	0.84	804	-0.32	0.998
	CI_Exp 2-3 - CI_Exp 3-4	Slope	1.05	3.01	580	0.35	0.997
		Curvature	-0.26	2.82	801	-0.09	1.000
T2	**NH=3 - CI_Exp<1**	**Slope**	**2.61**	**1.26**	**801**	**2.07**	**<0.05***
		Curvature	0.52	1.34	580	0.39	0.995
	NH=3 - CI_Exp 1-2	**Slope**	**2.94**	**1.33**	**801**	**2.20**	**<0.05***
		Curvature	2.54	1.42	580	1.78	0.387
	NH=3 - CI_Exp 2-3	**Slope**	**2.85**	**3.00**	**801**	**2.62**	**<0.01****
		Curvature	0.12	3.20	580	0.04	1.000
	NH=3 - CI_Exp 3-4	**Slope**	**3.48**	**1.20**	**581**	**2.90**	**<0.01****
		Curvature	2.13	1.12	802	1.90	0.318
	CI_Exp<1 - CI_Exp 1-2	Slope	2.01	1.08	580	1.86	0.342
		Curvature	0.33	1.02	801	0.32	0.998
	CI_Exp<1 - CI_Exp 2-3	Slope	0.41	3.07	580	0.13	1.000
		Curvature	4.46	2.87	801	1.55	0.528
	CI_Exp<1 - CI_Exp 3-4	**Slope**	**2.96**	**0.76**	**583**	**3.87**	**<0.001*****
		Curvature	0.47	0.72	804	0.66	0.964
	CI_Exp 1-2 - CI_Exp 2-3	Slope	2.42	3.10	580	0.78	0.937

		Curvature	4.79	2.91	801	1.65	0.467
	CI_Exp 1-2 - CI_Exp 3-4	Slope	0.94	0.90	582	1.05	0.830
		Curvature	0.80	0.84	803	0.96	0.874
	CI_Exp 2-3 - CI_Exp 3-4	Slope	3.36	3.01	580	1.12	0.797
		Curvature	3.99	2.82	801	1.42	0.618
	NH=3 - CI_Exp<1	**Slope**	**1.94**	**1.34**	**580**	**1.84**	**<0.05***
		Curvature	1.05	1.26	801	0.83	0.921
	NH=3 - CI_Exp 1-2	Slope	-0.73	1.42	580	-0.52	0.986
		Curvature	**6.08**	**1.33**	**801**	**2.81**	**<0.01****
	NH=3 - CI_Exp 2-3	Slope	2.28	3.20	580	0.71	0.954
		Curvature	**8.27**	**3.00**	**801**	**2.75**	**<0.05***
T3	**NH=3 - CI_Exp 3-4**	Slope	0.73	1.20	584	0.61	0.974
		Curvature	**5.37**	**1.13**	**806**	**4.77**	**<0.001*****
	CI_Exp<1 - CI_Exp 1-2	**Slope**	**-2.67**	**1.08**	**580**	**-2.47**	**<0.05***
		Curvature	-2.13	1.02	801	-2.10	0.222
	CI_Exp<1 - CI_Exp 2-3	Slope	0.34	3.07	580	0.11	1.000
		Curvature	**-9.31**	**2.87**	**801**	**-3.24**	**<0.01****
	CI_Exp<1 - CI_Exp 3-4	Slope	-1.21	0.77	590	-1.58	0.514
		Curvature	**-6.41**	**0.72**	**813**	**-8.93**	**<0.001*****

	CI_Exp 1-2 - CI_Exp 2-3	Slope	3.01	3.10	580	0.97	0.868
		Curvature	**-7.18**	**2.91**	**801**	**-2.47**	**<0.05***
	CI_Exp 1-2 - CI_Exp 3-4	Slope	1.46	0.90	587	1.63	0.481
		Curvature	**-4.29**	**0.84**	**810**	**-5.08**	**<0.001*****
	CI_Exp 2-3 - CI_Exp 3-4	Slope	-1.55	3.01	581	-0.52	0.986
		Curvature	2.90	2.82	801	1.03	0.842
T4	**NH=3 - CI_Exp<1**	**Slope**	**-2.70**	**1.34**	**580**	**-1.82**	**<0.05***
		Curvature	0.26	1.26	801	0.21	1.000
	NH=3 - CI_Exp 1-2	**Slope**	**-3.27**	**1.42**	**580**	**-2.19**	**<0.05***
		Curvature	-0.45	1.33	801	-0.34	0.997
	NH=3 - CI_Exp 2-3	**Slope**	**-4.50**	**3.20**	**580**	**-2.03**	**<0.05***
		Curvature	3.60	3.00	801	1.20	0.751
	NH=3 - CI_Exp 3-4	**Slope**	**-2.66**	**1.20**	**581**	**-2.22**	**<0.05***
		Curvature	1.90	1.12	801	1.70	0.438
	CI_Exp<1 - CI_Exp 1-2	Slope	-0.96	1.08	580	-0.89	0.901
		Curvature	0.18	1.02	801	0.18	1.000
	CI_Exp<1 - CI_Exp 2-3	**Slope**	**2.20**	**3.07**	**580**	**2.35**	**<0.05***
		Curvature	3.34	2.87	801	1.16	0.773

CI_Exp<1 - CI_Exp 3-4	**Slope**	**3.36**	**0.76**	**582**	**4.40**	**<0.001*** **
	Curvature	1.64	0.71	803	2.29	0.148
CI_Exp 1-2 - CI_Exp 2-3	**Slope**	**2.24**	**3.10**	**580**	**2.01**	**<0.05***
	Curvature	3.16	2.91	801	1.09	0.814
CI_Exp 1-2 - CI_Exp 3-4	**Slope**	**2.39**	**0.90**	**581**	**2.67**	**<0.05***
	Curvature	1.46	0.84	802	1.73	0.414
CI_Exp 2-3 - CI_Exp 3-4	Slope	-3.84	3.01	580	-1.28	0.704
	Curvature	-1.70	2.82	801	-0.60	0.975

S10. Results of one-way ANOVA on pitch range in children's lexical tone productions, with an independent variable "Group" (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) and a covariate "Chronological age".

Factor	**Df**	**F**	***p***
Group	**4**	**2.93**	**<0.05***
Chronological age	1	1.44	0.233
Group × Chronological age	4	0.34	0.852
Residuals	141		

S11. Pairwise comparison of lexical tone pitch range between groups (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4).

Group difference	***β***	**SE**	**df**	**t**	***p***
NH=3 - CI_Exp<1	**3.57**	**2.75**	**141**	**2.80**	**<0.05***
NH=3 - CI_Exp 1-2	**4.48**	**1.69**	**141**	**3.65**	**<0.01****
NH=3 - CI_Exp 2-3	**3.39**	**1.95**	**141**	**2.73**	**<0.05***
NH=3 - CI_Exp 3-4	**3.01**	**6.46**	**141**	**2.66**	**<0.05***
CI_Exp<1 - CI_Exp 1-2	0.91	3.04	141	0.30	0.998
CI_Exp<1 - CI_Exp 2-3	-0.18	3.20	141	-0.06	1.000
CI_Exp<1 - CI_Exp 3-4	-2.58	6.94	141	-0.66	0.964
CI_Exp 1-2 - CI_Exp 2-3	-1.09	2.35	141	-0.47	0.990
CI_Exp 1-2 - CI_Exp 3-4	-2.50	6.59	141	-0.83	0.920
CI_Exp 2-3 - CI_Exp 3-4	-2.40	6.66	141	-0.66	0.964

S12. Results of linear regression model of lexical tone pitch slopes with different predictors: tone category (T1-4), age at implantation, CI experience, chronological age and residual hearing. The predictors were entered into the model stepwise and the best model included all predictors.

Predictor	ΔR^2	***B***	***SE (B)***	***β***	***p***
Tone category	0.31	--	--	--	--
Age at implantation	0.15	0.001	0.002	0.022	<0.01**
CI experience	0.10	0.004	0.002	0.009	<0.05*
Chronological age	0.09	-0.004	0.001	-0.007	<0.05*
Residual hearing	0.09	-0.005	0.003	-0.007	0.091

S13. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's neutral tone productions, with three fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5), "Context" (after T1/2/4 and after T3) and "Type" (Reduplicatives and Possessives) and a covariate "Chronological age".

Parameter	**Factors**	**df 1**	**df 2**	**F**	***p***
Pitch slope	Group	4	232	1.46	0.215
	Context	**1**	**12050**	**4.66**	**<0.01***
	Type	1	11862	2.99	0.084
	Chronological age	1	312	0.65	0.421
	Group × Context	4	11950	0.52	0.719
	Group × Type	4	11746	1.47	0.210
	Context × Type	**1**	**12050**	**3.39**	**<0.05***
	Group × Chronological age	4	232	1.20	0.310
	Context × Chronological age	1	12010	0.10	0.750
	Type × Chronological age	1	11605	0.85	0.357
	Group × Context × Type	**4**	**11977**	**3.09**	**<0.01*****
	Group × Context × Chronological age	4	11939	0.27	0.900
	Group × Type × Chronological age	4	11710	1.84	0.117
	Context × Type × Chronological age	1	12009	0.19	0.659
	Group × Context × Type × Chronological age	4	11938	2.19	0.067

Pitch curvature	Group	4	381	0.92	0.450
	Context	1	11964	2.19	0.139
	Type	1	11848	0.07	0.793
	Chronological age	1	568	0.03	0.870
	Group × Context	4	11937	0.08	0.988
	Group × Type	4	11792	0.40	0.810
	Context × Type	1	11964	0.00	0.994
	Group × Chronological age	4	388	0.38	0.820
	Context × Chronological age	1	11943	0.00	0.981
	Type × Chronological age	1	11709	0.10	0.749
	Group × Context × Type	4	11937	0.12	0.974
	Group × Context × Chronological age	4	11931	0.26	0.901
	Group × Type × Chronological age	4	11775	0.39	0.818
	Context × Type × Chronological age	1	11943	0.06	0.803
	Group × Context × Type × Chronological age	4	11931	0.14	0.967

S14. Pairwise comparisons of neutral tone pitch slope between contexts (after T1/2/4 vs. after T3) across types (Reduplicative and Possessive) and groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5).

Group	Type	*β*	SE	df	*t*	*p*
NH=3	**Reduplicative**	**-9.79**	**2.65**	**4**	**-3.70**	**<0.01****
	Possessive	**-11.49**	**2.65**	**4**	**-4.34**	**<0.01****
CI_Imp 1-2	**Reduplicative**	**-6.86**	**3.42**	**11**	**-2.01**	**<0.05***
	Possessive	-2.29	3.68	15	-0.62	0.543
CI_Imp 2-3	Reduplicative	-1.16	2.76	5	-0.42	0.693
	Possessive	-1.38	2.80	5	-0.49	0.642
CI_Imp 3-4	Reduplicative	-1.99	2.97	6	-0.67	0.527
	Possessive	-0.59	2.99	7	-0.20	0.850
CI_Imp 4-5	Reduplicative	-2.18	3.98	21	-0.55	0.589
	Possessive	-2.41	3.97	21	-0.61	0.551

S15. Pairwise comparisons of neutral tone pitch slope between groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) across types (Reduplicative and Possessive) and contexts (after T1/2/4 and after T3).

Type	Context	Group difference	*β*	SE	df	t	*p*
Reduplicative	After T1/2/4	NH=3-CI_Imp1-2	-1.75	1.06	521	-1.50	0.069
		NH=3-CI_Imp2-3	**-2.81**	**1.22**	**485**	**-2.66**	**<0.05***
		NH=3-CI_Imp3-4	**-2.78**	**1.85**	**405**	**-2.50**	**<0.05***
		NH=3-CI_Imp4-5	**-2.30**	**0.97**	**536**	**-2.31**	**<0.05***
		CI_Imp1-2-CI_Imp2-3	-0.06	0.92	415	-0.06	1.000
		CI_Imp1-2-CI_Imp3-4	-1.03	1.67	373	-1.21	0.743
		CI_Imp1-2-CI_Imp4-5	-0.44	0.54	418	-0.82	0.926
		CI_Imp2-3-CI_Imp3-4	-1.97	1.78	376	-1.11	0.801
		CI_Imp2-3-CI_Imp4-5	-0.50	0.81	405	-0.62	0.972
		CI_Imp3-4-CI_Imp4-5	-2.48	1.62	368	-1.53	0.543
	After T3	NH=3-CI_Imp1-2	1.81	1.82	2333	1.55	0.532
		NH=3-CI_Imp2-3	**3.55**	**2.00**	**2132**	**2.27**	**<0.05***
		NH=3-CI_Imp3-4	**3.09**	**2.88**	**1795**	**2.38**	**<0.05***
		NH=3-CI_Imp4-5	**3.28**	**1.71**	**2460**	**2.16**	**<0.05***
		CI_Imp1-2-CI_Imp2-3	**3.27**	**1.29**	**1513**	**3.21**	**<0.01****
		CI_Imp1-2-CI_Imp3-4	**3.91**	**2.44**	**1534**	**3.60**	**<0.01****

		CI_Imp1-2-CI_Imp4-5	**3.09**	**0.77**	**1551**	**3.00**	**<0.01****
		CI_Imp2-3-CI_Imp3-4	1.64	2.58	1524	1.41	0.620
		CI_Imp2-3-CI_Imp4-5	**2.83**	**1.14**	**1492**	**2.48**	**<0.05***
		CI_Imp3-4-CI_Imp4-5	0.81	2.36	1530	0.35	0.997
Possessive	After T1/2/4	NH=3-CI_Imp1-2	-1.09	1.43	747	-1.26	0.104
		NH=3-CI_Imp2-3	**-2.41**	**1.53**	**656**	**-2.26**	**<0.05***
		NH=3-CI_Imp3-4	**-2.88**	**2.08**	**486**	**-2.42**	**<0.05***
		NH=3-CI_Imp4-5	**-2.35**	**1.34**	**789**	**-2.26**	**<0.05***
		CI_Imp1-2-CI_Imp2-3	-0.31	0.94	410	-0.34	0.997
		CI_Imp1-2-CI_Imp3-4	-0.79	1.69	372	-0.47	0.990
		CI_Imp1-2-CI_Imp4-5	-0.26	0.58	452	-0.46	0.991
		CI_Imp2-3-CI_Imp3-4	-0.48	1.78	368	0.27	0.999
		CI_Imp2-3-CI_Imp4-5	-0.05	0.81	386	-0.06	1.000
		CI_Imp3-4-CI_Imp4-5	-0.53	1.62	364	-0.32	0.998
	After T3	**NH=3-CI_Imp1-2**	**2.14**	**2.11**	**3055**	**2.07**	**<0.05***
		NH=3-CI_Imp2-3	**2.47**	**2.26**	**2703**	**2.10**	**<0.05***
		NH=3-CI_Imp3-4	**2.42**	**3.04**	**2039**	**2.80**	**<0.05***
		NH=3-CI_Imp4-5	**4.32**	**1.98**	**3182**	**4.18**	**<0.001*****
		CI_Imp1-2-CI_Imp2-3	2.61	1.37	1770	1.90	0.318

	CI_Imp1-2-CI_Imp3-4	2.56	2.46	1577	1.04	0.836
	CI_Imp1-2-CI_Imp4-5	**4.18**	**0.85**	**1982**	**4.89**	**<0.001*****
	CI_Imp2-3-CI_Imp3-4	0.05	2.59	1551	0.02	1.000
	CI_Imp2-3-CI_Imp4-5	**6.79**	**1.18**	**1630**	**5.76**	**<0.001*****
	CI_Imp3-4-CI_Imp4-5	**6.74**	**2.36**	**1531**	**2.86**	**<0.05***

S16. Results of linear mixed-effects model on normalized duration of neutral tone syllables, with three fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5), "Context" (after T1/2/4 and after T3) and "Type" (Reduplicatives and Possessives) and a covariate "Chronological age".

Factors	**df 1**	**df 2**	**F**	***p***
Group	4	206	1.72	0.148
Context	**1**	**1063**	**3.80**	**<0.05***
Type	**1**	**1080**	**7.93**	**<0.01****
Chronological age	1	271	0.21	0.651
Group × Context	4	1049	1.45	0.215
Group × Type	**4**	**1046**	**8.864**	**<0.001*****
Context × Type	**1**	**1063**	**9.43**	**<0.01****
Group × Chronological age	4	208	1.35	0.253
Context × Chronological age	1	1071	0.99	0.321
Type × Chronological age	1	1092	2.29	0.131
Group × Context × Type	4	1049	2.14	0.074
Group × Context × Chronological age	4	1050	1.17	0.320
Group × Type × Chronological age	4	1069	1.98	0.095
Context × Type × Chronological age	1	1071	1.69	0.194
Group × Context × Type × Chronological age	4	1050	2.00	0.092

S17. Pairwise comparisons of normalized duration of neutral tone syllables across types (Reduplicative and Possessive) and groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5).

Type	Group difference	*β*	SE	df	t	*p*
Reduplicative	NH=3 - CI_Imp 1-2	-0.26	0.22	304	-1.17	0.771
	NH=3 - CI_Imp 2-3	**-0.60**	**0.10**	**224**	**-6.19**	**<0.001*******
	NH=3 - CI_Imp 3-4	**-0.29**	**0.14**	**195**	**-2.08**	**<0.05***
	NH=3 - CI_Imp 4-5	**-0.33**	**0.28**	**189**	**-2.79**	**<0.05***
	CI_Imp 1-2 - CI_Imp 2-3	-0.24	0.23	297	-1.48	0.580
	CI_Imp 1-2 - CI_Imp 3-4	-0.04	0.25	270	-0.14	1.000
	CI_Imp 1-2 - CI_Imp 4-5	0.03	0.36	225	0.09	1.000
	CI_Imp 2-3 - CI_Imp 3-4	0.31	0.16	206	1.93	0.307
	CI_Imp 2-3 - CI_Imp 4-5	0.38	0.30	193	1.27	0.708
	CI_Imp 3-4 - CI_Imp 4-5	0.07	0.31	190	0.21	1.000
Possessive	**NH=3 - CI_Imp 1-2**	**-0.63**	**0.18**	**239**	**-3.47**	**<0.01****
	NH=3 - CI_Imp 2-3	**-0.64**	**0.09**	**201**	**-6.83**	**<0.001***

NH=3 - CI_Imp 3-4	**-0.90**	**0.14**	**196**	**-6.39**	**<0.001*** **
NH=3 - CI_Imp 4-5	**-0.33**	**0.28**	**188**	**-2.75**	**<0.05***
CI_Imp 1-2 - CI_Imp 2-3	-0.01	0.20	235	-0.04	1.000
CI_Imp 1-2 - CI_Imp 3-4	-0.27	0.22	225	-1.20	0.750
CI_Imp 1-2 - CI_Imp 4-5	0.31	0.33	202	0.92	0.889
CI_Imp 2-3 - CI_Imp 3-4	-0.26	0.16	199	-1.64	0.474
CI_Imp 2-3 - CI_Imp 4-5	0.31	0.29	190	1.07	0.822
CI_Imp 3-4 - CI_Imp 4-5	0.57	0.31	190	1.84	0.352

S18. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's neutral tone productions, with three fixed factors "Group" (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4), "Context" (after T1/2/4 and after T3) and "Type" (Reduplicatives and Possessives) and a covariate "Chronological age".

Parameter	Factors	df 1	df 2	F	*p*
Pitch slope	Group	4	198	1.80	0.130
	Context	**1**	**11988**	**9.12**	**<0.01****
	Type	**1**	**12032**	**5.28**	**<0.05***
	Chronological age	1	195	2.05	0.153
	Group × Context	**4**	**11981**	**3.54**	**<0.01****
	Group × Type	4	11712	1.58	0.177
	Context × Type	1	11985	0.98	0.322
	Group × Chronological age	4	197	1.46	0.216
	Context × Chronological age	1	11983	2.78	0.095
	Type × Chronological age	1	12009	2.66	0.103
	Group × Context × Type	4	11979	1.10	0.353
	Group × Context × Chronological age	4	11980	1.94	0.101
	Group × Type × Chronological age	4	11791	1.48	0.205
	Context × Type × Chronological age	1	11982	0.65	0.421
	Group × Context × Type × Chronological age	4	11979	0.85	0.493

Pitch curvature	Group	4	344	1.23	0.299
	Context	1	11959	0.30	0.586
	Type	1	11939	0.31	0.576
	Chronological age	1	337	1.34	0.248
	Group × Context	4	11957	0.11	0.980
	Group × Type	4	11738	0.84	0.499
	Context × Type	1	11959	1.09	0.297
	Group × Chronological age	4	341	0.88	0.476
	Context × Chronological age	1	11958	0.12	0.731
	Type × Chronological age	1	11920	0.36	0.550
	Group × Context × Type	4	11957	0.21	0.931
	Group × Context × Chronological age	4	11957	0.28	0.893
	Group × Type × Chronological age	4	11786	0.78	0.537
	Context × Type × Chronological age	1	11958	0.72	0.395
	Group × Context × Type × Chronological age	4	11957	0.15	0.962

S19. Pairwise comparisons of neutral tone pitch slope between contexts (after T1/2/4 vs. after T3) across groups (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3).

Group	***β***	SE	df	*t*	*p*
NH=3	**-10.64**	**0.82**	**11955**	**3.04**	**<0.01*** **
CI_Exp<1	-1.26	0.61	11987	0.74	0.23
CI_Exp 1-2	-1.40	0.68	11992	0.98	0.164
CI_Exp 2-3	-1.60	2.55	12044	0.81	0.209
CI_Exp 3-4	-1.32	0.25	11972	0.86	0.207

S20. Pairwise comparisons of neutral tone pitch slope between groups (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) across contexts (after T1/2/4 and after T3).

Context	Group difference	*β*	SE	df	t	*p*
After T1/2/4	NH=3 - CI_Exp<1	-0.89	0.80	207	-1.10	0.806
	NH=3 - CI_Exp 1-2	-0.20	0.85	203	-0.24	0.999
	NH=3 - CI_Exp 2-3	-0.18	2.09	221	-0.09	1.000
	NH=3 - CI_Exp 3-4	-0.64	0.69	195	-0.94	0.882
	CI_Exp<1 - CI_Exp 1-2	0.68	0.71	223	0.96	0.873
	CI_Exp<1 - CI_Exp 2-3	1.07	2.03	225	0.53	0.985
	CI_Exp<1 - CI_Exp 3-4	0.24	0.50	226	0.48	0.989
	CI_Exp 1-2 - CI_Exp 2-3	0.39	2.05	224	0.19	1.000
	CI_Exp 1-2 - CI_Exp 3-4	-0.44	0.57	213	-0.77	0.940
	CI_Exp 2-3 - CI_Exp 3-4	-0.82	1.99	224	-0.41	0.994
After T3	**NH=3 - CI_Exp<1**	**2.13**	**1.08**	**647**	**2.05**	**<0.05***
	NH=3 - CI_Exp 1-2	**2.69**	**1.12**	**616**	**2.40**	**<0.01****

NH=3 - CI_Exp 2-3	**2.27**	**2.80**	**658**	**2.19**	**<0.05***
NH=3 - CI_Exp 3-4	**4.18**	**0.91**	**601**	**4.58**	**<0.001*****
CI_Exp<1 - CI_Exp 1-2	**3.82**	**0.95**	**683**	**4.04**	**<0.001*****
CI_Exp<1 - CI_Exp 2-3	1.40	2.73	669	1.25	0.725
CI_Exp<1 - CI_Exp 3-4	**5.30**	**0.68**	**722**	**7.80**	**<0.001*****
CI_Exp 1-2 - CI_Exp 2-3	0.42	2.75	663	0.15	1.000
CI_Exp 1-2 - CI_Exp 3-4	**2.48**	**0.75**	**632**	**2.27**	**<0.05***
CI_Exp 2-3 - CI_Exp 3-4	1.90	2.67	664	0.71	0.954

S21. Results of linear mixed-effects model on normalized duration of neutral tone syllables, with three fixed factors "Group" (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4), "Context" (after T1/2/4 and after T3) and "Type" (Reduplicatives and Possessives) and a covariate "Chronological age".

Factors	df 1	df 2	F	*p*
Group	**4**	**180**	**4.54**	**<0.01****
Context	1	1039	2.14	0.144
Type	**1**	**1060**	**4.77**	**<0.05***
Chronological age	1	178	0.06	0.806
Group × Context	4	1039	1.14	0.337
Group × Type	4	1071	1.87	0.114
Context × Type	**1**	**1039**	**7.70**	**<0.01****
Group × Chronological age	4	179	0.48	0.751
Context × Chronological age	1	1039	2.54	0.111
Type × Chronological age	1	1064	0.66	0.417
Group × Context × Type	4	1038	2.29	0.058
Group × Context × Chronological age	4	1038	0.77	0.545
Group × Type × Chronological age	4	1068	1.54	0.190
Context × Type × Chronological age	1	1039	1.64	0.201
Group × Context × Type × Chronological age	4	1038	2.28	0.059

S22. Pairwise comparisons of normalized duration of neutral tone syllables groups (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4).

Group difference	*β*	SE	df	t	*p*
NH=3 - CI_Exp<1	**-0.35**	**0.16**	**169**	**-2.28**	**<0.05***
NH=3 - CI_Exp 1-2	**-0.41**	**0.16**	**165**	**-2.16**	**<0.05***
NH=3 - CI_Exp 2-3	**-0.43**	**0.40**	**176**	**-2.60**	**<0.05***
NH=3 - CI_Exp 3-4	**-0.80**	**0.13**	**162**	**-5.98**	**<0.001*****
CI_Exp<1 - CI_Exp 1-2	0.07	0.14	175	0.54	0.984
CI_Exp<1 - CI_Exp 2-3	0.11	0.39	178	0.28	0.999
CI_Exp<1 - CI_Exp 3-4	0.25	0.10	181	1.63	0.052
CI_Exp 1-2 - CI_Exp 2-3	0.04	0.39	177	0.09	1.000
CI_Exp 1-2 - CI_Exp 3-4	0.27	0.11	167	1.40	0.082
CI_Exp 2-3 - CI_Exp 3-4	0.24	0.38	178	1.18	0.120

S23. Results of linear regression model of neutral tone pitch slopes with different parameters (Context: after T1/2/4 and after T3; Type: Reduplicative and Possessive). The predictors were entered into the model stepwise and the best model included all predictors.

Predictor	ΔR^2	***B***	***SE (B)***	***β***	***p***
Context	0.16	0.188	0.025	0.120	<0.001***
Type	0.04	-0.287	0.022	-0.209	<0.001***
Age at implantation	0.05	0.006	0.002	0.069	<0.001***
CI experience	0.01	-0.002	0.002	-0.002	0.908
Chronological age	0.01	0.006	0.001	0.010	0.641
Residual hearing	0.01	0.003	0.003	0.025	0.149

S24. Results of linear regression model of neutral tone duration with different parameters (Context: after T1/2/4 and after T3; Type: Reduplicative and Possessive). The predictors were entered into the model stepwise and the best model included all predictors.

Predictor	***ΔR²***	***B***	***SE (B)***	***β***	***p***
Context	0.01	0.140	0.101	0.073	0.144
Type	0.10	-0.517	0.082	-0.306	<0.001***
Age at implantation	0.09	-0.019	0.093	-0.228	<0.05*
CI experience	0.02	0.011	0.090	0.138	<0.05*
Chronological age	0.02	0.010	0.093	0.123	<0.05*
Residual hearing	0.01	-0.014	0.011	-0.119	0.831

S25. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's disyllabic Context sandhi productions (Syllable 1), with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Context" (full sandhi vs. half sandhi) and a covariate "Chronological age".

Parameter	Factors	df 1	df 2	F	*p*
Pitch slope	Group	4	102	0.55	0.700
	Context	**1**	**3756**	**22.10**	**<0.001*****
	Chronological age	1	100	0.10	0.752
	Group × Context	**4**	**14715**	**4.53**	**<0.01****
	Group × Chronological age	4	102	0.48	0.749
	Context × Chronological age	1	14710	2.95	0.086
	Group × Context × Chronological age	4	14712	1.35	0.250
Pitch curvature	Group	4	139	1.32	0.266
	Context	**1**	**12116**	**3.10**	**<0.05***
	Chronological age	1	134	1.38	0.243
	Group × Context	4	14727	1.12	0.346
	Group × Chronological age	4	138	1.09	0.365
	Context × Chronological age	1	14718	1.53	0.216
	Group × Context × Chronological age	4	14723	0.79	0.534

S26. Pairwise comparisons of pitch slope of disyllabic tone sandhi compounds (Syllable 1) between two sandhi contexts (full sandhi vs. half sandhi) for each group (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5).

Group	*β*	SE	df	*t*	*p*
NH=3	**12.12**	**0.50**	**20**	**24.43**	**<0.001*****
CI_Imp 1-2	**3.63**	**0.88**	**185**	**4.14**	**<0.001*****
CI_Imp 2-3	1.37	0.60	42	1.27	0.106
CI_Imp 3-4	0.99	0.66	61	1.01	0.141
CI_Imp 4-5	0.46	0.76	104	0.60	0.55

S27. Pairwise comparisons of pitch slope of disyllabic tone sandhi compounds (Syllable 1) between groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) for each sandhi context (full sandhi and half sandhi).

Context	Group difference	*β*	SE	df	t	*p*
Full sandhi	**NH=3-CI_Imp1-2**	**2.93**	**1.17**	**203**	**2.79**	**<0.05***
	NH=3-CI_Imp2-3	**4.55**	**0.73**	**204**	**6.23**	**<0.001*****
	NH=3-CI_Imp3-4	**4.75**	**0.84**	**200**	**5.68**	**<0.001*****
	NH=3-CI_Imp4-5	**3.17**	**1.00**	**196**	**3.19**	**<0.05***
	CI_Imp1-2-CI_Imp2-3	**3.62**	**1.24**	**209**	**2.91**	**<0.05***
	CI_Imp1-2-CI_Imp3-4	**3.82**	**1.31**	**207**	**2.92**	**<0.05***
	CI_Imp1-2-CI_Imp4-5	2.25	1.42	204	1.59	0.508
	CI_Imp2-3-CI_Imp3-4	0.20	0.94	211	0.21	1.000
	CI_Imp2-3-CI_Imp4-5	-1.37	1.08	204	-1.27	0.711
	CI_Imp3-4-CI_Imp4-5	-1.57	1.16	202	-1.36	0.653
Half sandhi	**NH=3-CI_Imp1-2**	**-7.56**	**0.98**	**102**	**-7.69**	**<0.001*****
	NH=3-CI_Imp2-3	**-6.21**	**0.61**	**103**	**-6.11**	**<0.001*****
	NH=3-CI_Imp3-4	**-6.39**	**0.71**	**103**	**-7.02**	**<0.001*****

NH=3-CI_Imp4-5	**-8.49**	**0.85**	**102**	**-9.04**	**<0.001*** **
CI_Imp1-2-CI_Imp2-3	1.36	1.04	103	1.30	0.690
CI_Imp1-2-CI_Imp3-4	1.18	1.10	103	1.08	0.819
CI_Imp1-2-CI_Imp4-5	-0.93	1.19	102	-0.78	0.936
CI_Imp2-3-CI_Imp3-4	-0.18	0.78	104	-0.22	0.999
CI_Imp2-3-CI_Imp4-5	-2.28	0.91	102	-2.51	0.097
CI_Imp3-4-CI_Imp4-5	-2.11	0.98	103	-2.16	0.203

S28. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's disyllabic tone sandhi productions (Syllable 1), with two fixed factors "Group" (NH=3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) and "Context" (full sandhi vs. half sandhi) and a covariate "Chronological age".

Parameter	Factors	df 1	df 2	F	*p*
Pitch slope	Group	4	102	1.69	0.157
	Context	**1**	**4647**	**13.12**	**<0.001*** **
	Chronological age	1	101	0.18	0.676
	Group × Context	**4**	**14716**	**5.21**	**<0.001*** **
	Group × Chronological age	4	101	1.62	0.174
	Context × Chronological age	1	14712	2.04	0.153
	Group × Context × Chronological age	4	14713	2.10	0.077
Pitch curvature	Group	4	141	0.92	0.453
	Context	1	12742	2.12	0.145
	Chronological age	1	137	0.10	0.752
	Group × Context	4	14733	0.91	0.458
	Group × Chronological age	4	139	1.16	0.333
	Context × Chronological age	1	14725	1.19	0.274
	Group × Context × Chronological age	4	14728	0.83	0.506

S29. Pairwise comparisons of pitch slope of disyllabic tone sandhi compounds (Syllable 1) between two sandhi contexts (full sandhi vs. half sandhi) for each group (NH=3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4).

Group	***β***	SE	df	***t***	***p***
NH=3	**12.12**	**0.50**	**20**	**24.43**	**<0.001*** **
CI_Exp<1	0.71	0.80	131	0.89	0.377
CI_Exp 1-2	1.36	0.66	61	1.05	0.149
CI_Exp 2-3	1.99	0.75	100	1.33	0.093
CI_Exp 3-4	0.65	2.07	226	0.32	0.751

S30. Pairwise comparisons of pitch slope of disyllabic tone sandhi compounds (syllable 2) between groups (NH=3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) for each sandhi context (full sandhi and half sandhi).

Context	**Group difference**	***β***	**SE**	**df**	**t**	***p***
Full sandhi	**NH=3 - CI_Exp<1**	**3.75**	**1.07**	**193**	**3.50**	**<0.01****
	NH=3 - CI_Exp 1-2	**4.36**	**0.76**	**190**	**5.72**	**<0.001*****
	NH=3 - CI_Exp 2-3	**3.12**	**0.82**	**195**	**3.80**	**<0.01****
	NH=3 - CI_Exp 3-4	**4.64**	**1.14**	**200**	**4.08**	**<0.001*****
	CI_Exp<1 - CI_Exp 1-2	0.61	1.17	197	0.52	0.985
	CI_Exp<1 - CI_Exp 2-3	-0.63	1.21	199	-0.52	0.985
	CI_Exp<1 - CI_Exp 3-4	0.89	1.44	201	0.62	0.972
	CI_Exp 1-2 - CI_Exp 2-3	-1.24	0.94	201	-1.32	0.682
	CI_Exp 1-2 - CI_Exp 3-4	0.28	1.23	203	0.23	0.999
	CI_Exp 2-3 - CI_Exp 3-4	1.52	1.26	204	1.20	0.751
Half sandhi	**NH=3 - CI_Exp<1**	**-7.67**	**0.91**	**102**	**-8.39**	**<0.001*****
	NH=3 - CI_Exp 1-2	**-6.68**	**0.65**	**102**	**-10.22**	**<0.001*****

NH=3 - CI_Exp 2-3	**-6.30**	**0.70**	**102**	**-9.01**	**<0.001*****
NH=3 - CI_Exp 3-4	**-7.51**	**0.96**	**102**	**-7.81**	**<0.001*****
CI_Exp<1 - CI_Exp 1-2	1.00	0.99	102	1.01	0.852
CI_Exp<1 - CI_Exp 2-3	1.37	1.02	102	1.35	0.664
CI_Exp<1 - CI_Exp 3-4	0.16	1.22	102	0.13	1.000
CI_Exp 1-2 - CI_Exp 2-3	0.38	0.80	103	0.48	0.990
CI_Exp 1-2 - CI_Exp 3-4	-0.83	1.03	103	-0.81	0.928
CI_Exp 2-3 - CI_Exp 3-4	-1.21	1.06	103	-1.14	0.785

S31. Results of linear regression model of pitch slope of disyllabic tone sandhi compounds (Syllable 1) with different predictors. The predictors were entered into the model stepwise and the best model included all predictors.

Predictor	ΔR^2	***B***	***SE (B)***	β	***p***
Context	0.17	0.172	0.015	0.129	<0.001***
Age at implantation	0.17	0.007	0.001	0.123	<0.001***
CI experience	0.14	0.001	0.001	0.113	<0.001***
Chronological age	0.06	-0.003	0.001	-0.065	<0.001***
Residual hearing	0.03	0.007	0.001	0.018	0.204

S32. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's trisyllabic tone sandhi productions (Syllables 1 and 2), with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Context" (double-T3 and triple-T3) and a covariate "Chronological age".

Syllable	Parameter	Factors	df 1	df 2	F	*p*
Syllable 1	Pitch slope	**Group**	**4**	**174**	**6.44**	**<0.001*** **
		Context	1	5722	0.33	0.568
		Chronological age	1	174	2.03	0.156
		Group × Context	**4**	**5733**	**5.66**	**<0.001*** **
		Group × Chronological age	4	174	1.20	0.312
		Context × Chronological age	1	5733	0.33	0.565
		Group × Context × Chronological age	4	5733	2.15	0.072
	Pitch curvature	Group	4	2145	0.77	0.544
		Context	1	5596	0.31	0.576
		Chronological age	1	2146	1.17	0.279
		Group × Context	4	5727	0.42	0.794
		Group × Chronological age	4	2145	1.07	0.370
		Context × Chronological age	1	5727	0.24	0.624
		Group × Context × Chronological age	4	5727	0.41	0.802
Syllable 2	Pitch	**Group**	**4**	**177**	**3.56**	**<0.01** **

slope	**Context**	**1**	**1405**	**4.68**	**<0.05***
	Chronological age	1	177	0.23	0.631
	Group × Context	**4**	**5734**	**5.56**	**<0.001*****
	Group × Chronological age	4	177	0.74	0.564
	Context × Chronological age	1	5728	0.55	0.457
	Group × Context × Chronological age	4	5728	0.24	0.919
Pitch curvature	Group	4	742	0.37	0.828
	Context	1	4698	0.53	0.467
	Chronological age	1	743	1.61	0.204
	Group × Context	**4**	**5722**	**3.14**	**<0.05***
	Group × Chronological age	4	742	0.40	0.806
	Context × Chronological age	1	5721	0.31	0.575
	Group × Context × Chronological age	4	5721	0.44	0.779

S33. Pairwise comparisons of pitch parameters of trisyllabic tone sandhi compounds (Syllables 1 and 2) between contexts (double-T3 vs. triple-T3) for each group (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5).

Group	Syllable	Parameter	*β*	SE	df	t	*p*
NH=3	Syllable 1	Slope	0.81	0.18	5	1.89	0.059
	Syllable 2	**Slope**	**9.22**	**0.71**	**5**	**12.99**	**<0.001*** **
		Curvature	**-1.19**	**0.41**	**5**	**-2.90**	**<0.05***
CI_Imp 1-2	Syllable 1	Slope	0.77	0.72	80	1.08	0.283
	Syllable 2	Slope	1.58	1.46	80	1.09	0.281
		Curvature	**-4.28**	**1.33**	**80**	**-3.22**	**<0.001*** **
CI_Imp 2-3	Syllable 1	Slope	0.16	0.44	19	0.38	0.706
	Syllable 2	Slope	-1.18	1.02	19	-1.16	0.26
		Curvature	0.01	0.83	19	0.01	0.991
CI_Imp 3-4	Syllable 1	Slope	-0.35	0.52	30	-0.68	0.498
	Syllable 2	Slope	-0.73	1.14	30	-0.64	0.527
		Curvature	0.13	0.98	30	0.13	0.893
CI_Imp 4-5	Syllable 1	Slope	-0.2	0.62	50	-0.32	0.751
	Syllable 2	Slope	0.52	1.29	50	0.4	0.688
		Curvature	0.91	1.15	50	0.79	0.429

S34. Pairwise comparisons of pitch slope of trisyllabic tone sandhi compounds (Syllables 1 and 2) between groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) for each sandhi context (double-T3 and triple-T3).

Context	Group difference	Syllable	Parameter	*β*	SE	df	t	*p*
Double-T3	**NH=3 - CI_Imp 1-2**	**Syllable 1**	**Slope**	**2.11**	**0.78**	**188**	**2.72**	**<0.05***
		Syllable 2	**Slope**	**-8.87**	**1.23**	**195**	**-7.18**	**<0.001*****
			Curvature	0.34	0.74	850	0.45	0.991
	NH=3 - CI_Imp 2-3	**Syllable 1**	**Slope**	**4.39**	**0.49**	**175**	**9.01**	**<0.001*****
		Syllable 2	**Slope**	**-8.48**	**0.77**	**179**	**-11.04**	**<0.001*****
			Curvature	0.19	0.45	738	0.43	0.993
	NH=3 - CI_Imp 3-4	**Syllable 1**	**Slope**	**4.57**	**0.58**	**177**	**7.94**	**<0.001*****
		Syllable 2	**Slope**	**-7.9**	**0.91**	**181**	**-8.69**	**<0.001*****
			Curvature	0.35	0.54	756	0.66	0.965
	NH=3 - CI_Imp 4-5	**Syllable 1**	**Slope**	**3.7**	**0.68**	**172**	**5.48**	**<0.001*****
		Syllable 2	**Slope**	**-9.35**	**1.06**	**175**	**-8.78**	**<0.001*****
			Curvature	-0.78	0.62	706	-1.26	0.715
	CI_Imp 1-2 - CI_Imp 2-3	**Syllable 1**	**Slope**	**2.27**	**0.88**	**185**	**2.59**	**<0.05***
		Syllable 2	Slope	0.38	1.39	191	0.27	0.999
			Curvature	-0.14	0.84	827	-0.17	1.000
	CI_Imp 1-2 - CI_Imp 3-4	**Syllable 1**	**Slope**	**2.46**	**0.93**	**185**	**2.64**	**<0.05***

		Syllable 2	Slope	0.97	1.47	191	0.66	0.965
			Curvature	0.02	0.88	824	0.02	1.000
	CI_Imp 1-2 - CI_Imp 4-5	Syllable 1	Slope	1.59	1	182	1.6	0.502
		Syllable 2	Slope	-0.48	1.58	187	-0.31	0.998
			Curvature	-1.12	0.94	791	-1.19	0.756
	CI_Imp 2-3 - CI_Imp 3-4	Syllable 1	Slope	0.19	0.71	176	0.26	0.999
		Syllable 2	Slope	0.59	1.11	180	0.53	0.985
			Curvature	0.16	0.66	752	0.25	0.999
	CI_Imp 2-3 - CI_Imp 4-5	Syllable 1	Slope	-0.68	0.79	173	-0.87	0.909
		Syllable 2	Slope	-0.86	1.24	176	-0.69	0.958
			Curvature	-0.97	0.73	716	-1.34	0.664
	CI_Imp 3-4 - CI_Imp 4-5	Syllable 1	Slope	-0.87	0.85	174	-1.03	0.843
		Syllable 2	Slope	-1.45	1.34	178	-1.09	0.814
			Curvature	-1.14	0.78	727	-1.45	0.594
Triple-T3	**NH=3 - CI_Imp 1-2**	**Syllable 1**	**Slope**	**2.58**	**0.91**	**360**	**2.82**	**<0.05***
		Syllable 2	Slope	1.94	1.52	447	1.28	0.705
			Curvature	-2.75	1.16	3104	-2.36	**<0.05***
	NH=3 - CI_Imp 2-3	**Syllable 1**	**Slope**	**4.24**	**0.59**	**379**	**7.16**	**<0.001*****
		Syllable 2	Slope	-0.44	0.99	471	-0.45	0.992

		Curvature	1.40	0.76	3183	1.84	0.353
NH=3 - CI_Imp 3-4	**Syllable 1**	**Slope**	**3.91**	**0.69**	**360**	**5.68**	**<0.001*****
	Syllable 2	Slope	0.6	1.14	447	0.53	0.985
		Curvature	1.68	0.88	3106	1.91	0.310
NH=3 - CI_Imp 4-5	**Syllable 1**	**Slope**	**3.2**	**0.81**	**360**	**3.93**	**<0.001*****
	Syllable 2	Slope	0.4	1.35	447	0.3	0.998
		Curvature	1.33	1.04	3105	1.28	0.704
CI_Imp 1-2 - CI_Imp 2-3	Syllable 1	Slope	1.66	1.04	365	1.6	0.5
	Syllable 2	Slope	-2.38	1.73	453	-1.38	0.643
		Curvature	**4.15**	**1.33**	**3127**	**3.12**	**<0.05***
CI_Imp 1-2 - CI_Imp 3-4	Syllable 1	Slope	1.33	1.1	360	1.21	0.743
	Syllable 2	Slope	-1.34	1.82	446	-0.74	0.948
		Curvature	**4.43**	**1.40**	**3102**	**3.17**	**<0.05***
CI_Imp 1-2 - CI_Imp 4-5	Syllable 1	Slope	0.62	1.18	360	0.53	0.985
	Syllable 2	Slope	-1.54	1.96	446	-0.79	0.934
		Curvature	**4.07**	**1.50**	**3102**	**2.71**	**<0.05***
CI_Imp 2-3 - CI_Imp 3-4	Syllable 1	Slope	-0.33	0.85	368	-0.39	0.995
	Syllable 2	Slope	1.04	1.41	457	0.74	0.948
		Curvature	0.28	1.09	3139	0.26	0.999

CI_Imp 2-3 - CI_Imp 4-5	Syllable 1	Slope	-1.04	0.95	366	-1.09	0.81
	Syllable 2	Slope	0.84	1.58	455	0.53	0.984
		Curvature	-0.07	1.22	3131	-0.06	1.000
CI_Imp 3-4 - CI_Imp 4-5	Syllable 1	Slope	-0.71	1.02	360	-0.7	0.956
	Syllable 2	Slope	-0.2	1.69	446	-0.12	1
		Curvature	-0.35	1.29	3102	-0.27	0.999

S35. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's trisyllabic tone sandhi productions (Syllables 1 and 2), with two fixed factors "Group" (NH=3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) and "Context" (double-T3 and triple-T3) and a covariate "Chronological age".

Syllable	Parameter	Factors	df 1	df 2	F	*p*
Syllable 1	Pitch slope	**Group**	**4**	**179**	**3.70**	**<0.01****
		Context	1	5803	0.37	0.544
		Chronological age	1	185	2.98	0.086
		Group × Context	**4**	**5754**	**5.55**	**<0.001*****
		Group × Chronological age	4	178	1.80	0.131
		Context × Chronological age	1	5794	0.34	0.558
		Group × Context × Chronological age	4	5757	0.43	0.789
	Pitch curvature	Group	4	2398	1.79	0.128
		Context	1	5692	0.04	0.836
		Chronological age	1	2561	2.93	0.087
		Group × Context	4	5726	0.20	0.938
		Group × Chronological age	4	2388	1.97	0.096
		Context × Chronological age	1	5727	0.03	0.874
		Group × Context × Chronological age	4	5727	0.18	0.950
Syllable 2	Pitch	Group	4	186	1.90	0.113

slope	Context	1	3139	0.44	0.509
	Chronological age	1	193	1.00	0.317
	Group × Context	**4**	**5752**	**3.36**	**<0.01****
	Group × Chronological age	4	185	1.12	0.351
	Context × Chronological age	1	5808	0.05	0.816
	Group × Context × Chronological age	4	5757	0.34	0.854
Pitch curvature	**Group**	**4**	**775**	**2.99**	**<0.05***
	Context	1	5351	0.99	0.319
	Chronological age	1	846	2.79	0.095
	Group × Context	4	5721	0.30	0.881
	Group × Chronological age	4	771	1.90	0.109
	Context × Chronological age	1	5724	0.92	0.339
	Group × Context × Chronological age	4	5721	0.22	0.929

S36. Pairwise comparisons of pitch slope of trisyllabic tone sandhi compounds (Syllables 1 and 2) between contexts (double-T3 vs. triple-T3) for each group (NH=3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4).

Group	Syllable	*β*	SE	df	t	*p*
NH=3	Syllable 1	0.81	0.18	5	1.89	0.059
	Syllable 2	**9.22**	**0.71**	**5**	**12.99**	**<0.001*** **
CI_Exp<1	Syllable 1	-0.58	0.70	75	-0.83	0.407
	Syllable 2	-0.34	1.43	75	-0.24	0.814
CI_Exp 1-2	Syllable 1	0.18	0.43	19	0.42	0.678
	Syllable 2	-0.45	1.02	19	-0.44	0.663
CI_Exp 2-3	Syllable 1	-0.06	0.47	23	-0.14	0.890
	Syllable 2	-0.91	1.06	23	-0.85	0.403
CI_Exp 3-4	Syllable 1	0.54	0.77	104	0.70	0.483
	Syllable 2	1.52	1.55	104	0.98	0.328

S37. Pairwise comparisons of pitch slope of trisyllabic tone sandhi compounds (Syllables 1 and 2) between groups (NH=3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) for each sandhi context (double-T3 and triple-T3).

Tone	Group difference	Parameter	*β*	SE	df	t	*p*
Double-T3	**NH=3 - CI_Exp<1**	**Syllable 1**	**4.13**	**0.77**	**171**	**5.39**	**<0.001*******
		Syllable 2	**-8.00**	**1.19**	**176**	**-6.71**	**<0.001*******
	NH=3 - CI_Exp 1-2	**Syllable 1**	**3.97**	**0.49**	**172**	**8.10**	**<0.001*******
		Syllable 2	**-9.25**	**0.76**	**178**	**-12.12**	**<0.001*******
	NH=3 - CI_Exp 2-3	**Syllable 1**	**3.49**	**0.53**	**178**	**6.57**	**<0.001*******
		Syllable 2	**-7.53**	**0.83**	**186**	**-9.07**	**<0.001*******
	NH=3 - CI_Exp 3-4	**Syllable 1**	**5.21**	**0.84**	**183**	**6.20**	**<0.001*******
		Syllable 2	**-9.58**	**1.32**	**192**	**-7.28**	**<0.001*******
	CI_Exp<1 - CI_Exp 1-2	Syllable 1	-0.16	0.87	171	-0.19	1.000
		Syllable 2	-1.25	1.35	176	-0.93	0.886
	CI_Exp<1 - CI_Exp 2-3	Syllable 1	-0.64	0.89	173	-0.71	0.953
		Syllable 2	0.46	1.39	179	0.33	0.997

	CI_Exp<1 - CI_Exp 3-4	Syllable 1	1.08	1.10	178	0.98	0.865
		Syllable 2	-1.58	1.73	185	-0.92	0.890
	CI_Exp 1-2 - CI_Exp 2-3	Syllable 1	-0.47	0.67	176	-0.71	0.955
		Syllable 2	1.72	1.05	183	1.64	0.473
	CI_Exp 1-2 - CI_Exp 3-4	Syllable 1	1.24	0.94	181	1.33	0.673
		Syllable 2	-0.33	1.46	189	-0.22	0.999
	CI_Exp 2-3 - CI_Exp 3-4	Syllable 1	1.72	0.96	183	1.79	0.381
		Syllable 2	-2.05	1.50	191	-1.37	0.650
Triple-T3	**NH=3 - CI_Exp<1**	**Syllable 1**	**3.24**	**0.92**	**353**	**3.52**	**<0.01****
		Syllable 2	0.89	1.52	455	0.59	0.977
	NH=3 - CI_Exp 1-2	**Syllable 1**	**3.84**	**0.60**	**371**	**6.44**	**<0.001*****
		Syllable 2	-0.48	0.98	480	-0.48	0.989
	NH=3 - CI_Exp 2-3	**Syllable 1**	**3.12**	**0.63**	**354**	**4.93**	**<0.001*****
		Syllable 2	0.79	1.04	456	0.76	0.943

NH=3 - CI_Exp 3-4	**Syllable 1**	**5.44**	**0.99**	**353**	**5.50**	**<0.001*** **
	Syllable 2	1.17	1.63	455	0.72	0.952
CI_Exp<1 - CI_Exp 1-2	Syllable 1	0.60	1.05	358	0.57	0.979
	Syllable 2	-1.37	1.73	462	-0.79	0.933
CI_Exp<1 - CI_Exp 2-3	Syllable 1	-0.12	1.07	353	-0.12	1.000
	Syllable 2	-0.10	1.76	454	-0.06	1.000
CI_Exp<1 - CI_Exp 3-4	Syllable 1	2.20	1.31	353	1.68	0.448
	Syllable 2	0.28	2.16	454	0.13	1.000
CI_Exp 1-2 - CI_Exp 2-3	Syllable 1	-0.72	0.81	362	-0.89	0.900
	Syllable 2	1.26	1.33	467	0.95	0.878
CI_Exp 1-2 - CI_Exp 3-4	Syllable 1	1.60	1.11	358	1.45	0.597
	Syllable 2	1.65	1.83	461	0.90	0.897
CI_Exp 2-3 - CI_Exp 3-4	Syllable 1	1.32	1.13	353	1.06	0.240
	Syllable 2	0.38	1.86	454	0.21	1.000

S38. Results of linear regression models of pitch slope of trisyllabic tone sandhi compounds (Syllables 1 and 2) with different predictors (Context: double-T3 context T3T3T1/2/4 and triple-T3 context T3T3T3). The predictors were entered into the model stepwise and the best model included all predictors.

Predictor	Syllable	ΔR^2	*B*	*SE (B)*	*β*	*p*
Context	1	0.03	-0.008	0.017	-0.011	0.612
	2	0.19	0.047	0.025	0.053	<0.05*
Age at implantation	1	0.17	0.004	0.001	-0.301	<0.001***
	2	0.06	0.003	0.002	-0.081	<0.001***
CI experience	1	0.13	0.001	0.001	-0.140	<0.001***
	2	0.06	-0.005	0.002	0.051	<0.05*
Chronological age	1	0.07	-0.009	0.001	0.123	<0.001***
	2	0.01	0.001	0.001	-0.012	0.090
Residual hearing	1	0.03	-0.004	0.001	0.012	0.820
	2	0.01	-0.001	0.001	-0.011	0.789

S39. Additional analysis of Syllable 1 (carrying T1-4) in each neutral tone word of experiment 2.

Given that in experiment 1 (lexical tones) there were only four test items (one token for each lexical tone), we performed an additional analysis on syllable 1 of the neutral tone words used to validate the results of experiment 1. Recall that the only early implanted children produced lexical tone pitch contours comparable to those of NH children, while other children, including those with longer CI experience, produced flatter pitch for all tones.

A linear mixed-effects model was performed on the pitch contour parameters (slope and curvature) of Syllable 1 of neutral tone words, with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Tones" (T1, T2, T3 and T4), a covariant "Chronological age" (age in months; to control for the different chronological ages across participants) and a random factor "Participant". The results showed significant interactions of "Group × Tone" on pitch slope ($F(12, 12019) = 18.71$, $p < 0.001$) and curvature ($F(12, 11999) = 5.15$, $p < 0.001$; Table 1). Tukey-HSD post-hoc tests showed that the CI_Imp 1-2 group produced pitch contours comparable to those of the NH children for all lexical tones, while other CI groups differed from the NH children in T2, T3 and T4 productions, with flatter pitch contours (Table 2).

Table 1. Results of linear mixed-effects model on pitch shape (slope and curvature) in children's productions of Syllable 1 (lexical tones) of neutral tone words, with two fixed factors "Group" (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5) and "Tones" (T1, T2, T3 and T4) and a covariate "Chronological age".

Parameter	Factors	df 1	df 2	F	*p*
Pitch slope	**Group**	**4**	**188**	**4.17**	**<0.01****
	Tone	**3**	**8903**	**15.03**	**<0.001*****
	Chronological age	1	227	0.91	0.341
	Group × Tone	**12**	**12019**	**18.71**	**<0.001*****
	Group × Chronological age	4	191	1.18	0.321
	Tone × Chronological age	3	7009	1.76	0.152
	Group × Tone × Chronological age	12	7772	1.68	0.064
Pitch curvature	**Group**	**4**	**1548**	**4.47**	**<0.01****
	Tone	**3**	**11868**	**5.66**	**<0.001*****
	Chronological age	1	2053	1.07	0.301
	Group × Tone	**12**	**11999**	**5.15**	**<0.001*****

Group × Chronological age	4	1571	1.43	0.220
Tone × Chronological age	3	11760	0.82	0.484
Group × Tone × Chronological age	12	11730	0.72	0.733

Table 2. Pairwise comparisons of pitch contours of syllable 1 (TX) of neutral tone words between groups (NH = 3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5).

Tone	Group difference	Parameter	*β*	SE	df	t	*p*
T1	NH=3 - CI_Imp 1-2	Slope	1.43	1.07	661	1.34	0.666
		Curvature	0.70	0.83	8508	0.84	0.917
	NH=3 - CI_Imp 2-3	Slope	1.31	0.57	514	2.32	0.142
		Curvature	-0.17	0.43	7740	-0.39	0.995
	NH=3 - CI_Imp 3-4	Slope	1.55	0.66	482	2.36	0.130
		Curvature	-0.24	0.49	7545	-0.48	0.989
	NH=3 - CI_Imp 4-5	Slope	1.08	0.86	504	1.26	0.716
		Curvature	-0.09	0.65	7607	-0.14	1.000
	CI_Imp 1-2 - CI_Imp 2-3	Slope	-0.12	1.16	643	-0.10	1.000
		Curvature	-0.86	0.90	8434	-0.96	0.873
	CI_Imp 1-2 - CI_Imp 3-4	Slope	0.12	1.21	619	0.10	1.000
		Curvature	-0.93	0.93	8330	-1.00	0.854

	CI_Imp 1-2 - CI_Imp 4-5	Slope	-0.35	1.33	603	-0.26	0.999
		Curvature	-0.79	1.02	8221	-0.77	0.939
	CI_Imp 2-3 - CI_Imp 3-4	Slope	0.23	0.80	505	0.29	0.998
		Curvature	-0.07	0.60	7703	-0.12	1.000
	CI_Imp 2-3 - CI_Imp 4-5	Slope	-0.23	0.97	515	-0.24	0.999
		Curvature	0.08	0.73	7699	0.10	1.000
	CI_Imp 3-4 - CI_Imp 4-5	Slope	-0.46	1.03	501	-0.45	0.991
		Curvature	0.15	0.77	7625	0.19	1.000
T2	NH=3 - CI_Imp 1-2	Slope	0.37	1.04	735	0.35	0.997
		Curvature	1.64	0.83	9076	1.98	0.277
	NH=3 - CI_Imp 2-3	**Slope**	**1.26**	**0.58**	**579**	**2.17**	**<0.05***
		Curvature	**2.64**	**0.45**	**8249**	**5.90**	**<0.001*****
	NH=3 - CI_Imp 3-4	**Slope**	**2.21**	**0.70**	**601**	**3.15**	**<0.05***
		Curvature	**2.12**	**0.54**	**8339**	**3.91**	**<0.001*****

NH=3 - CI_Imp 4-5	**Slope**	**2.10**	**0.87**	**554**	**2.42**	**<0.05***
	Curvature	**2.30**	**0.66**	**8012**	**3.47**	**<0.01****
CI_Imp 1-2 - CI_Imp 2-3	**Slope**	**2.89**	**1.14**	**721**	**1.78**	**<0.05***
	Curvature	1.00	0.91	9012	1.10	0.806
CI_Imp 1-2 - CI_Imp 3-4	**Slope**	**1.84**	**1.21**	**714**	**1.52**	**0.547**
	Curvature	0.48	0.96	8963	0.50	0.987
CI_Imp 1-2 - CI_Imp 4-5	Slope	1.74	1.32	668	1.32	0.679
	Curvature	0.66	1.03	8737	0.64	0.968
CI_Imp 2-3 - CI_Imp 3-4	**Slope**	**2.95**	**0.85**	**619**	**2.12**	**<0.05***
	Curvature	-0.52	0.66	8451	-0.79	0.933
CI_Imp 2-3 - CI_Imp 4-5	Slope	0.84	0.99	576	0.85	0.914
	Curvature	-0.33	0.76	8171	-0.44	0.992
CI_Imp 3-4 - CI_Imp 4-5	Slope	-0.11	1.07	586	-0.10	1.000
	Curvature	0.19	0.82	8221	0.23	0.999

	NH=3 - CI_Imp 1-2	Slope	-1.67	1.04	594	-1.61	0.490
		Curvature	0.82	0.79	8071	1.03	0.841
	NH=3 - CI_Imp 2-3	**Slope**	**-3.61**	**0.57**	**514**	**-6.36**	**<0.001*** **
		Curvature	-0.58	0.43	7740	-1.36	0.657
	NH=3 - CI_Imp 3-4	**Slope**	**-3.90**	**0.66**	**482**	**-5.94**	**<0.001*** **
		Curvature	-0.22	0.49	7545	-0.44	0.992
T3	**NH=3 - CI_Imp 4-5**	**Slope**	**-4.11**	**0.86**	**504**	**-4.78**	**<0.001*** **
		Curvature	-0.22	0.65	7607	-0.34	0.997
	CI_Imp 1-2 - CI_Imp 2-3	**Slope**	**-1.94**	**1.13**	**587**	**-1.91**	**<0.05***
		Curvature	-1.40	0.87	8063	-1.61	0.488
	CI_Imp 1-2 - CI_Imp 3-4	**Slope**	**-2.23**	**1.18**	**569**	**-1.89**	**<0.05***
		Curvature	-1.03	0.90	7980	-1.15	0.781
	CI_Imp 1-2 - CI_Imp 4-5	Slope	-2.44	1.30	563	-1.87	0.334
		Curvature	-1.04	0.99	7928	-1.05	0.834

	CI_Imp 2-3 - CI_Imp 3-4	Slope	-0.29	0.80	505	-0.37	0.996
		Curvature	0.36	0.60	7703	0.60	0.975
	CI_Imp 2-3 - CI_Imp 4-5	Slope	-0.51	0.97	515	-0.52	0.985
		Curvature	0.36	0.73	7699	0.49	0.988
	CI_Imp 3-4 - CI_Imp 4-5	Slope	-0.21	1.03	501	-0.21	1.000
		Curvature	0.00	0.77	7625	-0.01	1.000
T4	NH=3 - CI_Imp 1-2	Slope	-2.06	1.04	595	-1.99	0.274
		Curvature	-0.84	0.79	8081	-1.06	0.826
	NH=3 - CI_Imp 2-3	**Slope**	**-2.40**	**0.57**	**507**	**-4.25**	**<0.001*** **
		Curvature	**-0.87**	**0.43**	**7679**	**-2.04**	**<0.05***
	NH=3 - CI_Imp 3-4	**Slope**	**-2.66**	**0.66**	**485**	**-4.04**	**<0.001*** **
		Curvature	-0.91	0.49	7573	-1.84	0.351
	NH=3 - CI_Imp 4-5	**Slope**	**-2.14**	**0.86**	**506**	**-2.49**	**<0.05***
		Curvature	-0.29	0.65	7623	-0.45	0.992

CI_Imp 1-2 - CI_Imp 2-3	Slope	-0.34	1.13	584	-0.30	0.998
	Curvature	-0.03	0.86	8040	-0.03	1.000
CI_Imp 1-2 - CI_Imp 3-4	Slope	-0.60	1.18	569	-0.51	0.987
	Curvature	-0.07	0.90	7980	-0.07	1.000
CI_Imp 1-2 - CI_Imp 4-5	Slope	-0.08	1.30	563	-0.06	1.000
	Curvature	0.55	0.99	7928	0.56	0.981
CI_Imp 2-3 - CI_Imp 3-4	Slope	-0.26	0.80	499	-0.33	0.998
	Curvature	-0.04	0.60	7654	-0.07	1.000
CI_Imp 2-3 - CI_Imp 4-5	Slope	0.26	0.97	511	0.27	0.999
	Curvature	0.58	0.73	7666	0.79	0.934
CI_Imp 3-4 - CI_Imp 4-5	Slope	0.52	1.03	501	0.51	0.987
	Curvature	0.62	0.77	7625	0.80	0.931

We then explored the effect of CI experience on the pitch realization of children's productions of Syllable 1 of neutral tone words. A linear mixed-effects model was performed on the pitch contour (slope and curvature) for each token, with two fixed factors “Group” (CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3, CI_Exp 3-4, NH=3) and “Tones” (T1, T2, T3 and T4), a covariant “Chronological age” and a random factor “Participant”. The results showed significant interactions of “Group × Tone” on pitch slope ($F (12, 11989) = 19.38$, $p < 0.001$) and curvature ($F (12, 11998) = 5.17$, $p < 0.001$; Table 3). Tukey-HSD post-hoc tests revealed that, relative to the NH children, all CI groups generally produced flatter pitch contours for T2, T3 and T4 (Table 4).

Table 3. Results of linear mixed-effects model on pitch shape (slope and curvature) in children’s productions of Syllable 1 (lexical tones) of neutral tone words, with two fixed factors “Group” (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4) and “Tones” (T1, T2, T3 and T4) and a covariate "Chronological age".

Parameter	**Factors**	**df 1**	**df 2**	**F**	***p***
Pitch slope	**Group**	**4**	**175**	**4.25**	**<0.01****
	Tone	**3**	**12006**	**13.06**	**<0.001*****
	Chronological age	1	173	0.01	0.942
	Group × Tone	**12**	**11989**	**19.38**	**<0.001*****
	Group × Chronological age	4	174	0.19	0.945
	Tone × Chronological age	3	12009	0.42	0.741
	Group × Tone × Chronological age	12	11979	0.90	0.544

Pitch curvature	**Group**	**4**	**1381**	**3.11**	**<0.05***
	Tone	**3**	**11948**	**3.46**	**<0.05***
	Chronological age	1	1353	0.34	0.558
	Group × Tone	**12**	**11998**	**5.17**	**<0.001*****
	Group × Chronological age	4	1353	0.13	0.970
	Tone × Chronological age	3	11949	0.12	0.947
	Group × Tone × Chronological age	12	11948	0.21	0.998

Table 4. Pairwise comparisons of pitch contours of syllable 1 (TX) of neutral tone words between groups (NH = 3, CI_Exp<1, CI_Exp 1-2, CI_Exp 2-3 and CI_Exp 3-4).

Tone	Group difference	Parameter	*β*	SE	df	t	*p*
T1	NH=3 - CI_Exp<1	Slope	-0.32	0.95	502	-0.34	0.997
		Curvature	0.14	0.72	7771	0.20	1.000
	NH=3 - CI_Exp 1-2	Slope	0.11	0.97	498	0.11	1.000
		Curvature	0.64	0.74	7769	0.87	0.909
	NH=3 - CI_Exp 2-3	Slope	0.29	1.31	576	0.22	1.000
		Curvature	0.36	1.00	8013	0.36	0.996
	NH=3 - CI_Exp 3-4	Slope	-1.41	0.82	478	-1.74	0.414
		Curvature	0.39	0.61	7593	0.64	0.968
	CI_Exp<1 - CI_Exp 1-2	Slope	0.43	0.79	542	0.54	0.983
		Curvature	0.50	0.61	8078	0.82	0.925
	CI_Exp<1 - CI_Exp 2-3	Slope	0.61	1.18	622	0.51	0.986
		Curvature	0.22	0.91	8203	0.24	0.999
	CI_Exp<1 - CI_Exp 3-4	Slope	-1.10	0.58	538	-1.88	0.332

		Curvature	0.25	0.45	8000	0.56	0.981
	CI_Exp 1-2 - CI_Exp 2-3	Slope	0.18	1.20	613	0.15	1.000
		Curvature	-0.28	0.93	8187	-0.30	0.998
	CI_Exp 1-2 - CI_Exp 3-4	Slope	-1.53	0.63	521	-2.43	0.109
		Curvature	-0.25	0.48	7965	-0.51	0.986
	CI_Exp 2-3 - CI_Exp 3-4	Slope	-1.70	1.08	630	-1.58	0.514
		Curvature	0.03	0.83	8189	0.04	1.000
T2	**NH=3 - CI_Exp<1**	**Slope**	**-1.73**	**1.00**	**617**	**-1.92**	**<0.05***
		Curvature	0.84	0.78	8496	1.08	0.819
	NH=3 - CI_Exp 1-2	Slope	-0.85	1.03	595	-0.82	0.924
		Curvature	**2.52**	**0.80**	**8365**	**2.66**	**<0.05***
	NH=3 - CI_Exp 2-3	Slope	-0.65	1.26	603	-0.52	0.986
		Curvature	**2.64**	**0.98**	**8502**	**2.65**	**<0.05***
	NH=3 - CI_Exp 3-4	**Slope**	**-2.59**	**0.86**	**574**	**-3.00**	**<0.05***

		Curvature	**-1.76**	**0.66**	**8238**	**-2.64**	**<0.05***
	CI_Exp<1 - CI_Exp 1-2	Slope	0.88	0.83	652	1.06	0.828
		Curvature	-0.32	0.65	8694	-0.49	0.989
	CI_Exp<1 - CI_Exp 2-3	Slope	1.08	1.10	638	0.98	0.865
		Curvature	-0.20	0.86	8728	-0.23	0.999
	CI_Exp<1 - CI_Exp 3-4	Slope	-0.87	0.61	655	-1.42	0.617
		Curvature	**-2.60**	**0.48**	**8725**	**-5.41**	**<0.001*** **
	CI_Exp 1-2 - CI_Exp 2-3	Slope	0.20	1.12	618	0.18	1.000
		Curvature	0.12	0.88	8612	0.13	1.000
	CI_Exp 1-2 - CI_Exp 3-4	**Slope**	**-1.75**	**0.66**	**595**	**-2.66**	**<0.05***
		Curvature	**-2.28**	**0.51**	**8378**	**-4.49**	**<0.001*** **
	CI_Exp 2-3 - CI_Exp 3-4	**Slope**	**-1.94**	**0.97**	**608**	**-2.00**	**<0.05***
		Curvature	**-2.40**	**0.76**	**8597**	**-3.15**	**<0.05***
T3	**NH=3 - CI_Exp<1**	**Slope**	**-2.17**	**0.95**	**502**	**-2.24**	**<0.05***

	Curvature	-0.15	0.72	7771	-0.21	1.000
NH=3 - CI_Exp 1-2	**Slope**	**2.26**	**0.97**	**486**	**2.33**	**<0.05***
	Curvature	0.92	0.73	7661	1.26	0.718
NH=3 - CI_Exp 2-3	**Slope**	**2.17**	**1.34**	**625**	**2.17**	**<0.05***
	Curvature	0.25	1.04	8338	0.24	0.999
NH=3 - CI_Exp 3-4	**Slope**	**4.00**	**0.82**	**478**	**4.90**	**<0.001*****
	Curvature	0.54	0.61	7594	0.88	0.904
CI_Exp<1 - CI_Exp 1-2	**Slope**	**3.43**	**0.78**	**523**	**4.38**	**<0.001*****
	Curvature	1.07	0.60	7923	1.80	0.377
CI_Exp<1 - CI_Exp 2-3	Slope	2.35	1.22	687	1.93	0.302
	Curvature	0.41	0.95	8580	0.43	0.993
CI_Exp<1 - CI_Exp 3-4	**Slope**	**5.17**	**0.58**	**538**	**8.85**	**<0.001*****
	Curvature	0.69	0.45	8000	1.56	0.526
CI_Exp 1-2 - CI_Exp 2-3	Slope	-1.09	1.23	665	-0.88	0.904

		Curvature	-0.67	0.96	8497	-0.70	0.958
	CI_Exp 1-2 - CI_Exp 3-4	Slope	1.74	0.62	491	2.81	0.041
		Curvature	-0.38	0.47	7710	-0.81	0.929
	CI_Exp 2-3 - CI_Exp 3-4	Slope	2.83	1.12	709	2.53	0.085
		Curvature	0.29	0.87	8636	0.33	0.997
T4	**NH=3 - CI_Exp<1**	**Slope**	**-2.60**	**0.95**	**502**	**-2.63**	**<0.05***
		Curvature	-0.08	0.72	7771	-0.12	1.000
	NH=3 - CI_Exp 1-2	**Slope**	**-2.21**	**0.97**	**486**	**-2.22**	**<0.05***
		Curvature	-0.29	0.73	7661	-0.40	0.995
	NH=3 - CI_Exp 2-3	**Slope**	**2.04**	**1.31**	**576**	**-2.03**	**<0.05***
		Curvature	-0.90	1.00	8013	-0.90	0.898
	NH=3 - CI_Exp 3-4	**Slope**	**2.09**	**0.82**	**480**	**2.56**	**<0.05***
		Curvature	0.57	0.62	7611	0.92	0.890
	CI_Exp<1 - CI_Exp 1-2	Slope	0.39	0.78	523	0.50	0.987

	Curvature	-0.21	0.60	7923	-0.35	0.997
CI_Exp<1 - CI_Exp 2-3	Slope	0.56	1.18	622	0.47	0.990
	Curvature	-0.81	0.91	8203	-0.90	0.899
CI_Exp<1 - CI_Exp 3-4	**Slope**	**2.69**	**0.59**	**542**	**4.60**	**<0.001*** **
	Curvature	0.65	0.45	8032	1.45	0.594
CI_Exp 1-2 - CI_Exp 2-3	Slope	0.17	1.20	603	0.14	1.000
	Curvature	-0.61	0.92	8123	-0.66	0.965
CI_Exp 1-2 - CI_Exp 3-4	**Slope**	**2.30**	**0.62**	**495**	**3.71**	**<0.01** **
	Curvature	0.86	0.47	7740	1.83	0.355
CI_Exp 2-3 - CI_Exp 3-4	Slope	2.13	1.08	632	1.97	0.280
	Curvature	1.46	0.83	8198	1.76	0.395

Chapter Seven: The Representation of Allophonic Variants of Tone Sandhi in the Developing Lexicon

This chapter is based on the following paper:

Tang, P., Xu Rattanasone, N., Yuen, I., Gao, L., Demuth, K. (submitted). The representation of allophonic variants of tone sandhi in the developing lexicon. *Developmental Psychology.*

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors and collaborators) when needed.

Abstract

Phonological processes result in surface variants of the same words across phonological contexts, posing potential word learning challenges for learners. Mandarin tone sandhi is a tonal process changing tone 3 (T3) in different tonal and syntactic contexts, resulting in allophonic variants of T3 in connected speech. A previous study found that 3-5-year-olds have robust representations of *familiar* tone sandhi words (Wewalaarachchi & Singh, 2016), but these words were lexicalized in children's mental lexicon; it therefore remains unclear how children might perform on *novel* items involving productive tone sandhi processes. This study examined Mandarin-learning children's representation of allophonic variants of T3 using novel compounds as stimuli. Ninety-four 3-5-year-olds and 29 adults were tested. Sensitivity to allophonic mispronunciations of T3 syllables in novel tone sandhi compounds was measured using a visual fixation procedure. The results showed that children, like adults, treated tone sandhi mispronunciations as target-like. Thus, in recognizing novel tone sandhi words, Mandarin-speaking children exhibit flexibility in accommodating the allophonic variants of T3, suggesting that they have developed an abstract T3 category in their mental lexicon. The findings reveal the effect of phonological processes in shaping children's phonological representations.

Key words: phonological representation; phonological processes; Mandarin Chinese; tone sandhi; mispronunciation; language acquisition

Introduction

During language acquisition, children must develop accurate phonological representations for words in their mental lexicon and use these to produce and comprehend speech. However, these representations can be complex, because phonological processes can result in multiple surface realizations of the same word stem across phonological contexts. For instance, the *final-devoicing process* in Dutch prohibits voiced obstruents such as /d/ in the word /bɛ**d**/ 'bed' from being voiced, leading to [t] instead. However, the devoicing process does not apply word-medially, resulting in the plural /bɛ**d**en/ 'beds'. The Dutch devoicing process thus results in multiple surface realizations of the lexical stem across phonological contexts, with /d/ realized as [**t**] or [d], depending on the phonological environment. Acquiring these allophonic variants and productively using them in novel words is still on-going in Dutch-learning 4-year-olds (Zamuner, Kerkhoff & Fikkert, 2012).

Several studies have investigated how children process and/or represent surface variants of phonological processes in different languages (e.g., Dutch (Kerkhoff, 2007; Buckler & Fikkert, 2016), English (Skoruppa, Mani & Peperkamp, 2013), French (Skoruppa, Mani & Peperkamp, 2013) and German (van de Vijver & Baer-Henney, 2013)). Most of these have focussed on consonants and vowels. Few studies have attempted to go beyond segments to investigate the effect of allophonic variation in tonal processes, despite the fact that most of worlds' languages are tonal languages (Yip, 2002). Mandarin Chinese is a tonal language spoken by a vast majority of population in mainland China, with four lexical tones and a well-known tonal process, tone sandhi. The aim of the current study was therefore to explore how 3-5-year-olds represent the surface variants of *tone sandhi* processes in Mandarin Chinese.

Lexical tones in Mandarin Chinese primarily contrast in pitch contour, i.e., level tone 1 (T1), rising tone 2 (T2), dipping tone 3 (T3) and falling tone 4 (T4; figure 1). These four lexical tones are used to differentiate word meanings, e.g., /ma1/ *mother*, /ma2/ *hemp*, /ma3/ *horse* and /ma4/ *scold*, and are acquired early, i.e., before age 3 (Li & Thompson, 1977).

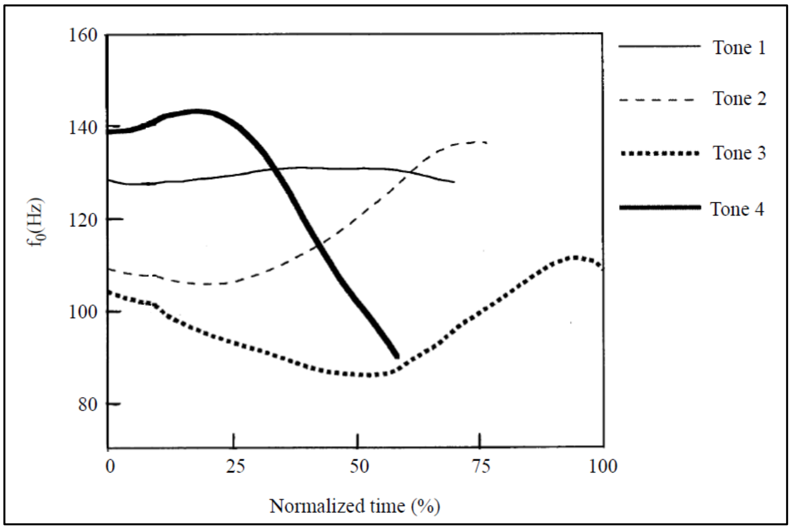


Figure 1. Pitch contours of Mandarin lexical tones, i.e., T1 (level), T2 (rising), T3 (dipping) and T4 (falling).

Among the four lexical tones, T3 is subject to tone sandhi which results in multiple surface realizations. T3 is acoustically realized with a dipping contour when it occurs as a monosyllable in citation form or in word-final position of a disyllabic word. In word-initial position of a disyllabic word, however, it undergoes the tone sandhi process. Depending on the following lexical tones, the tone sandhi process generates two allophonic variants of T3 (Chen, 2000). In a T3+T3 sequence, where **full sandhi** occurs, the initial T3 is realized as the *rising* tone similar with T2 (referred as *T2 to differentiate this from the underlying lexical T2): **T3**+T3→ ***T2**+T3. In a T3+T1/T2/T4 sequence, where **half sandhi** occurs, T3 is realized with a *low-falling* contour (referred to as **half-T3**): **T3**+T1/2/4 → **half-T3**+T1/2/4.

Mandarin-speaking 3-year-olds have been reported to correctly produce both *T2 and half-T3 variants in appropriate full and half sandhi contexts in known words (Li & Thompson, 1977; Hua & Dodd, 2000; Xu Rattanasone, et al. 2018). These results seem to suggest an early acquisition of tone sandhi processes. However, it is unclear from these studies using *known* words whether children can generalise their knowledge productively to *new* words. A recent study using a novel compound formation method showed that 3-5-year-olds can productively apply *both* tone sandhi processes in *production* (Tang et al., submitted). This raises the question as to how children at this age might represent these allophonic variants in their mental lexicon, i.e., is each allophonic variant linked independently to its respective sandhi context (i.e., *T2 to full sandhi and half-T3 to half sandhi), or are both forms linked to the lexical T3 category? Understanding this issue will deepen our understanding of children's ability to accommodate the surface variants resulting from tone sandhi and their acquisition of phonological processes in general.

To date, there has been only one study examining Mandarin-speaking 3-5-year-olds' perception of T3 variants, which used known/lexicalized tone sandhi words, e.g., "yu3 san3" *umbrella* realised as *T2T3 (Wewalaarachchi & Singh, 2016). Using the intermodal preferential looking (IPL) paradigm (Golinkoff, et al., 1987), they examined children's sensitivity to tonal mispronunciation of the full sandhi form (*T2T3). Children listened to an underlying T3T3 sandhi word (e.g., *umbrella*, "yu3 san3") with correct surface realization ***T2**T3 or a mispronounced realized with **half-T3**T3[21]. The results showed that children rejected the half-T3T3, suggesting that they have robust *T2T3 representations for these words. However, these words were

[21] Note that the authors argue that they used **T3**T3 for this mispronounced form, but the pitch contours they provide suggest that this was realized as **half-T3**T3.

familiar words, always realised as *T2T3 in the input children hear. Children might therefore have lexicalised them as holistic T2T3 units and not realise that T2 is really *T2, and thus underlyingly T3.

This leaves open the question as to what allophonic representations children might have for T3. There is some evidence from the adult literature showing that Mandarin speakers represent both *T2 and half-T3 as variants of underlying T3 (Chen, Shen and Schiller, 2011). Given that 3-5-year-olds are already productively applying both tone sandhi processes in production (Tang et al., submitted), it is possible that they might also represent these two allophonic variants in the lexicon, like adults. However, to test this, children must be given disyllabic compounds where they know the first syllable is underlyingly T3.

The aim of the present study was therefore to examine 3-5-year-olds' perceptual representation of the allophonic variants of T3 resulting from tone sandhi processes. We used novel compounds with underlying T3T3 (the context for full sandhi to apply) and the IPL procedure. Adults were included as controls. We asked if children and adults would reject **half-T3**T3 'mispronunciations' for underlying T3T3 novel compounds, in line with Wewalaarachchi and Singh (2016), or if children might represent both *T2 and half-T3 as allophonic variants of T3, as observed in adults (Chen, Shen and Schiller, 2011).

Method

Participants

Ninety-four Mandarin-speaking 3-, 4- and 5-year-olds and 29 adult controls were recruited (Table 1). All participants spoke Mandarin as their first language and were born and raised in Beijing, China. According to reports from the kindergarten, the children did not have any speech, hearing, language or intellectual difficulty. The study was conducted in accordance with the ethics protocol approved by [left blank for reviewing] University's Human Ethics Panel.

Table 1. Number of participants in each age group

Age group	Male	Female	Total
3 yrs. (3;1-3;12, mean: 3;8, SD: 4 months)	14	20	34
4 yrs. (4;1-4;12, mean: 4;5, SD: 4 months)	18	21	39
5 yrs. (5;1-6;2, mean: 5;7, SD: 5 months)	9	12	21
Adults (19-25 yrs., mean: 20, SD: 2 years)	7	22	29

Stimuli

Two types of disyllabic "animal + object" novel compounds were used as stimuli, including eight *non-sandhi* T2T1 control items, e.g., "niu2 deng1" *cow-bulb*, and eight *tone sandhi* (underlying T3T3) target items, e.g., "ma3 gu3" *horse-drum*, with the surface sandhi realization ***T2**T3. Following the methods used by Wewalaarachchi and Singh (2016), tonal mispronunciations were obtained by changing the target syllable (underlying T2 and surface *T2) to half-T3, resulting in eight *non-sandhi* and eight *sandhi* mispronunciations (see Table 2).

Table 2. Novel compound tonal stimuli used in this study

	Non-sandhi (underlying T2T1)	**Sandhi** (underlying T3T3)
Target surface tones	T2T1	*T2T3
Mispronounced surface tones	half-T3T1	half-T3T3

Note that tonal realization of Syllable 1 (underlined) were acoustically identical in both the non-sandhi and sandhi conditions (see Figure 2). In the *non-sandhi* condition, Syllable 1 is underlyingly a lexical T2 without any tone sandhi process. Thus, the half-T3 mispronunciation is not a possible allophone for this syllable. On the other hand, in the sandhi condition, Syllable 1 is underlyingly a lexical T3. There are two possible surface (allophonic) variants for T3: *T2 or half-T3. If both allophones are accessed during the tone sandhi process, both might be treated as acceptable in a mispronunciation task.

Audio stimuli

All the target speech stimuli were recorded by a female Beijing-Mandarin speaker in a child-directed speech style. For the sandhi mispronunciations, a list of half-T3T1 compounds sharing the same Syllable 1 with the target sandhi compounds was also recorded. Using Praat software (Boersma & Weenink, 2016), the pitch contour of Syllable 1 was then extracted and superimposed on that of the target tone sandhi compounds *T2T3 to create half-T3T3 (see Hallé, Chang & Best (2004) for similar methods). The resulting Syllable 1 target and mispronounced tones are shown in Figure 2 (see Appendix 2 for acoustic details). Note, critically, that the Syllable 1 non-sandhi control target and mispronounced tones are acoustically identical to the target and mispronounced sandhi forms. Thus, any differences between these two conditions will be due to phonological, not phonetic, differences.

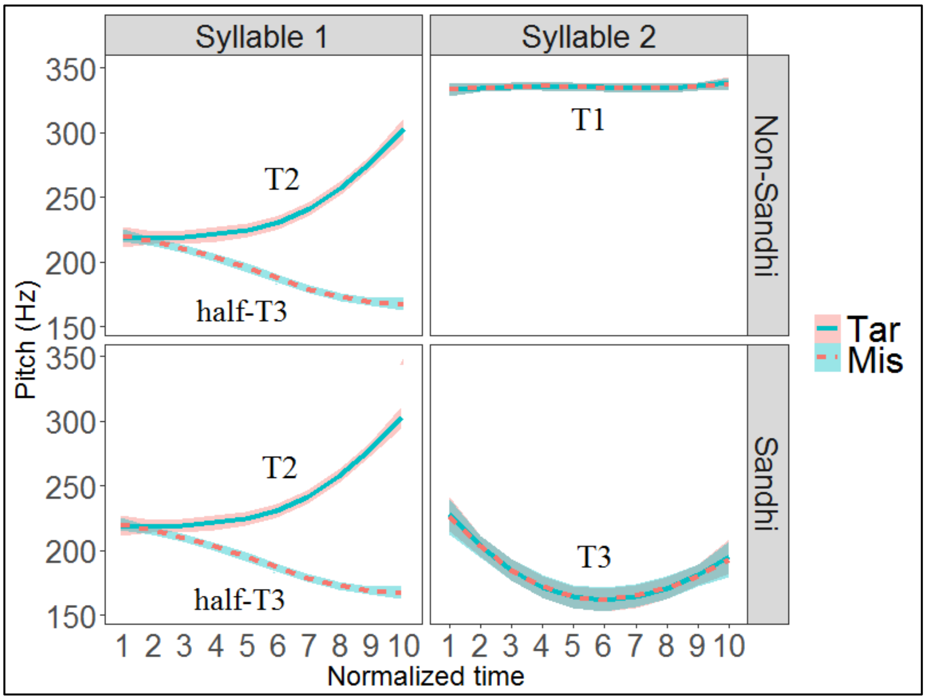


Figure 2. Pitch contours of syllables 1 and 2 within each stimulus for target pronunciations (Tar) and mispronunciations (Mis) in Non-Sandhi and Sandhi conditions. The tonal mispronunciation was made on the first syllable, changing it from *T2/T2 to half-T3.

Visual stimuli

Each audio stimulus was matched with a pair of yoked familiar vs. novel pictures (9 cm * 9 cm). The familiar pictures were always "animal + object" combinations, e.g., a drum with a horse on it for the stimulus "ma3 gu3" *horse-drum*; the novel pictures were always "novel animal + object" combinations with similar shading and visual complexity. During the experiment, the yoked familiar-novel pictures were displayed side-by-side horizontally on a 15.6 inch (38.7 cm * 25.9 cm) white screen.

Apparatus

Audio stimuli were played via AKG K612 Pro headphones at approximately 70 dB. The SMI Red-250 portable eye-tracker was used to record participants' visual fixations (sampling rate: 250 Hz; position: around 70 cm in front of the participant) in the current IPL task. The experiment was programmed using SMI Experimenter Centre software with a 9-point calibration procedure prior to testing.

Procedure

All participants were tested individually in a quiet room at the affiliated kindergarten of Beijing Language and Culture University (for children) or Beijing Language and Culture University (for Adults). Prior to testing, a picture naming task had been conducted to ensure that all participants knew the underlying tone of each word used to form the novel compounds in the perception experiment (cf. Tang et al., submitted).

The perception experiment began with three practice trials to familiarize the participant with the procedure (data collected did not undergo further analysis), followed by 19 test trials in pseudo-randomised order. The 19 test trials included four non-sandhi target trials, four non-sandhi mispronunciation trials, four sandhi target trials, four sandhi mispronunciation trials and three *word*-mispronunciation trials (where Syllable 1 was replaced with a syllable that differed from the target in segments and tone). The *word*-mispronunciation trials were used as a screener to exclude any participant who did not understand the task. Six 3-year-olds who did not respond to 2 of the 3 word-mispronunciation trials were excluded from further analysis (see the Data analysis section below for more detail).

In each trial, a fixation image (Mickey Mouse) first flashed in the middle of the screen for 2 seconds to attract the participant's attention. This was followed by a pre-naming phase (4 seconds), a post-naming (4 seconds) phase, and a 'dancing' phase designed to help the participant stay engaged (2 seconds). In the pre-naming phase, participants were first presented with the yoked familiar-novel pictures. After 1.25 seconds, each picture flashed once (with a 500 ms interval between the two flashes), ensuring that children looked at both pictures. The order of the flashes was randomized across participants. The audio directive "ni3 kan4!" *look!* (800 ms) was then played, followed by silence (200 ms). The boundary between the pre- and post-naming phases was aligned with the onset of syllable 2 of each novel compound word (Figure 3). After the post-naming phase, the 'dancing' phase started, during which there was animation (jiggling) of the familiar picture (for correct pronunciation trials) or the novel picture (for mispronunciation trials).

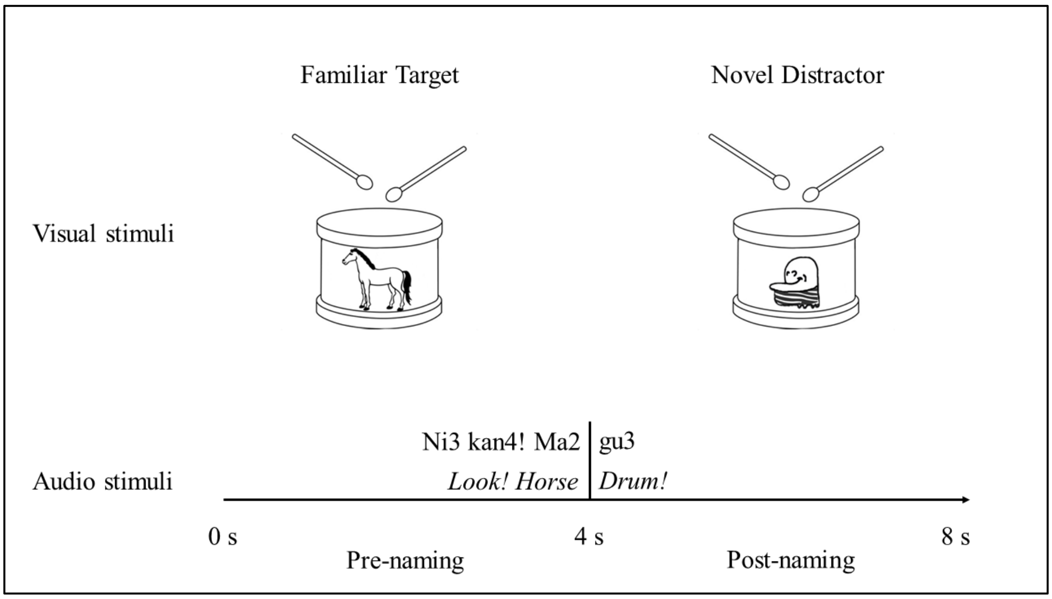


Figure 3. Sequence of events in each trial.

Data analysis

Analyses were conducted using the BeGaze software (version 3.7). Areas of interest (AOIs) were defined as two 11×11 cm squares covering the two pictures in each trial, given the typical 0.5° - 1° accuracy for mobile eye-trackers. Only fixations on the two AOIs underwent further analysis.

Two types of analysis were used: (1) the difference ratio of looks to the familiar picture and (2) the time course analysis. The difference ratio of looks to the familiar picture was computed by subtracting the proportion of the fixation times on the familiar picture during the pre-naming phase from that during the post-naming phase. This analysis provides information about the overall effect of the auditory stimuli on participants' preference between the familiar and novel pictures. A significant increase in looking proportion toward the familiar picture between the pre- and post-naming phases typically suggests that the participant has mapped the auditory stimulus onto the familiar picture, and vice versa; no change in fixation to the familiar picture suggests uncertainty in how the auditory stimulus corresponds to the visual targets. Linear mixed-effect models were used to compare the difference ratio across conditions and groups, using the "*anova*" function in the R package "lmerTest", providing statistical significance of omnibus main effects and interactions with F-statistics (Kuznetsova, Brockhoff, & Christensen, 2015; R Core Team, 2016). When a multi-level main effect or interaction reached significance ($p < 0.05$), post-hoc comparisons were then performed using the "lsmeans" package (Lenth, 2016).

The time course analysis provides additional fine-grained information regarding fixation patterns that may be obscured by the aggregate difference measures, revealing the efficiency with which word selection is made upon hearing the audio stimuli (Fernald et al., 2001). In the current study, the time course of fixation was binned into 40 windows (200 ms for each window) for each trial. A cluster-based permutation test was then used to estimate the divergence period between the looking time courses to the two objects (Maris & Oostenveld, 2007). Relative to the traditional approach finding the divergence window using adjusted multiple comparisons, the cluster-based permutation test provides a better way to minimize false-alarm rates without sacrificing sensitivity. This method first identifies the time-bins with *t*-statistics exceeding a critical threshold (e.g., $p < 0.05$), and then groups these into time-clusters on the basis of adjacency. The maximum cluster-level test statistics (the sum of all individual *t*-values within a cluster) are then computed to generate permutation distributions based on 1000 random partitions. The significance of a cluster is then determined in terms of the highest or the lowest 2.5th percentile of the corresponding distribution.

A total of 1968 trials were originally obtained from the 123 participants. Participants were then screened using the three *word*-mispronunciation trials, for which we expected negative difference ratios of looking to the familiar picture in at least two trials. Six participants (all 3-year-olds) were excluded from further analysis for not reaching this criterion. For the remaining participants, 116 trials with greater than 50% track-loss data were also excluded. This resulted in a total of 117 participants (88 children and 29 adults) and 1756 trials used in the final analysis.

Results

Difference ratio

A linear-mixed effects model was performed to compare the difference ratio of looks to the familiar picture between the target pronunciation and mispronunciation trials across conditions and groups (Figure 4), with three fixed factors "Pronunciation" (Target and Mispronunciation), "Condition" (Non-Sandhi and Sandhi) and "Group" (3-, 4-, 5-year-olds and adults) and a random factor "Participant". The results are presented in Appendix 3, showing that there was a significant interaction of "Pronunciation × Condition" ($F(1, 223) = 32.16$, $p < 0.001$). This indicates that target pronunciation and mispronunciation differed in Non-Sandhi and Sandhi conditions. In the post-hoc analysis of the non-Sandhi condition, all groups showed significantly smaller difference ratios in the mispronunciation trials than in the target pronunciation trials. This shows that all participants looked less to the familiar picture and more to the competitor picture indicating a rejection of the mispronunciation. In the Sandhi condition, however, the difference ratios did not differ between the target pronunciation and mispronunciation trials for any group. This shows that participants did not reduce their looks to the familiar picture, irrespective of whether a correct pronunciation or a mispronunciation was presented (Appendix 4).

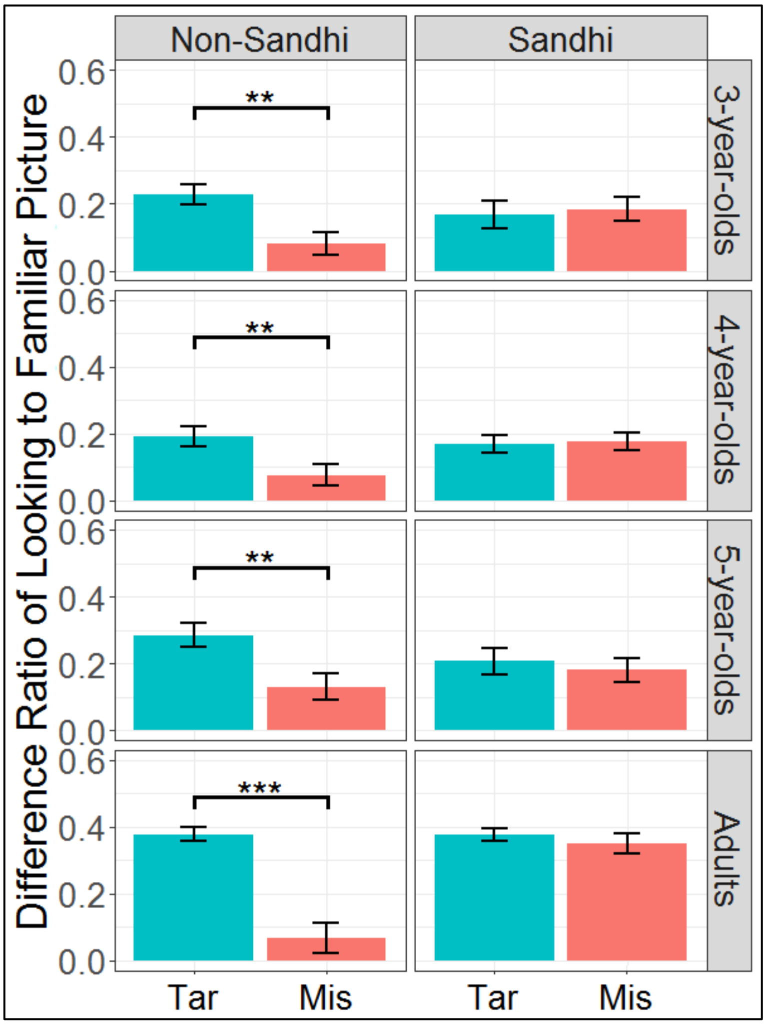


Figure 4. Difference ratios of looking to familiar pictures for target pronunciations (Tar) and mispronunciations (Mis) across conditions (Non-Sandhi and Sandhi) and groups: 3-, 4-, 5-year-olds and adults. Asterisks indicate significant differences: $p < 0.01$** and $p < 0.001$***.

Time course analysis

The second analysis was the time course analysis, exploring a more nuanced gaze pattern upon hearing the target pronunciations and mispronunciations. Cluster-based permutation tests were conducted to find the divergence period between the target pronunciation and mispronunciation trials across conditions and groups (Figure 5 and Appendix 5). In the Non-Sandhi condition, the divergence between the two time-courses (target and mispronunciation trials) appeared after the offset of the target disyllabic stimuli for all groups (the onset and offset of the target stimuli are marked by the dash lines). In the Sandhi condition, however, the two time-courses did not diverge for any child group, and only diverged for a short period (from 6000 ms to 6800 ms) for the adult group. These results suggested that, similar to the difference ratio analysis, children and adults did not reject the mispronunciation in the sandhi condition, though adults showed some uncertainty as to the acceptability of the mispronunciation.

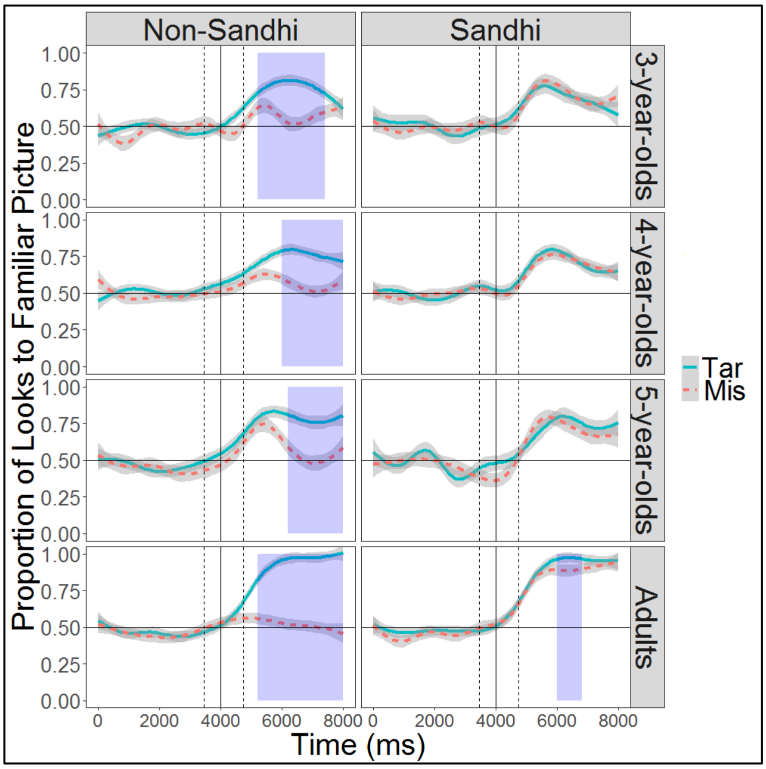


Figure 5. Time-courses of looking proportion toward target of target pronunciations and mispronunciations for Non-Sandhi and Sandhi conditions across groups. The horizontal solid lines indicate the change level (proportion = 0.5) of looking toward target, and the vertical solid lines indicate the onset of the second syllable within each stimulus, marked as the boundary between pre- and post-naming stages. Dash lines mark the averaged onset and offset points of stimuli. Blue shadows indicate the significant divergence between two curves based on cluster-based permutation tests.

Discussion

The current study examined how allophonic variants of T3 resulting from the tone sandhi process are represented in Mandarin-speaking 3-5-year-olds' mental lexicon. We asked if children would represent both *T2 and half-T3 as allophonic variants of T3, and as possible outcomes of the tone sandhi process. The results showed that children, like adults, accepted both the *T2T3 target pronunciation and the half-T3T3 mispronunciation as possible surface realizations for underlying T3T3 items, suggesting that both allophonic variants were accessed in recognizing T3 syllables in novel tone sandhi compounds. Thus, our results suggest that both *T2 and half-T3 are stored in 3-5-year-olds' mental lexicon, perhaps together under the same abstract T3 category.

Our results did not support the hypothesis that children would reject half-T3T3 mispronunciations, as previously found in Wewalaarachchi and Singh (2016) using familiar words. The current findings suggest that familiar items such as "yu3 san3" *umbrella* are lexicalised as a single lexical unit with the surface form T2T3. This is consistent with the form children hear in their language input. For these words it was therefore unclear whether children realise that T2T3 is really underlyingly T3T3, and hence a derived *T2T3 form. In the current study, when children understood that novel compounds were underlyingly T3T3, where tone sandhi should apply, they, like adults, were willing to accept both allophonic variants from the sandhi process. This suggests that even 3-year-olds use knowledge about tone sandhi when learning new words. This is consistent with the recent findings that 3-year-olds were able to productively apply tone sandhi to novel compound words (Tang et al., submitted).

These conclusions are further strengthened by the results of the mispronunciations in the non-Sandhi **control condition**. When children and adults understood that the novel compound in the control condition was underlyingly T2T1, with no involvement of tone sandhi, half-T3 could not be a possible allophonic variant, and was rejected. This is similar to the finding that children rejected half-T3 mispronunciations in Wewalaarachchi and Singh (2016), because the familiar lexicalized disyllabic compounds were treated as lexicalized T2T3 words. Note that that study had no adult controls; it is therefore unclear how adults might have performed on this task.

Our results therefore indicate that, when recognizing novel tone sandhi compounds which do not have robust lexical representations, children exhibit flexibility in accepting the allophonic mispronunciations of T3 syllables. This indicates that phonological processes such as tone sandhi shape children's phonological representations, leading to an abstract T3 category and its tone sandhi allophones being stored in their mental lexicon. According to the Multi-Level Representation model (Zhou and Marslen-Wilson, 1994), when compounds are processed in Mandarin Chinese, the representations of each constituent syllable is initially activated before the entire compound is activated as speech unfolds. Such abstract representations for T3 might therefore be advantageous for children in accommodating the surface variation of T3 syllables (i.e., sometimes rising and sometimes low-falling), in connected speech, and in learning new words. It is perhaps for the same reason that reports of T2/T3 perceptual confusion by young children abound (Li & Thompson, 1977; Wong, Schwartz & Jenkins, 2005; Shi et al., 2017); since underlying T2 is acoustically identical to derived *T2, children might think that T2 as actually an allophone of T3.

One of the central questions in the acquisition of phonological alternations is how children represent the surface variants. It has been suggested that in many languages children initially prefer invariant (non-alternating) forms, with adult-like mastery of phonological alternations emerging later (Albright & Hayes, 2011). For example, Kerrhoff (2007) and Zamuner, Kerrhoff, and Fikkert (2012) show that Dutch-learning 4-year-olds have problems in both recognizing and applying *final devoicing* in the singulars of nouns, while they perform much better on non-alternating items. Song, Shattuck-Hufnagel and Demuth (2015) also examined American English-speaking 2-year-olds productions of allophonic variants of alveolar stops /t, d/ in spontaneous speech, finding that children produced variant forms less often than their mothers did in child-directed speech. Thus, it seems that young children start out by abstracting and using invariant forms first, build canonical phonological representations in their mental lexicon first, before learning to incorporate phonological variants. In contrast, our results show that Mandarin-learning 3-year-olds have already developed an abstract T3 category along with its allophonic variants, rather than simply the canonical T3 form. This is also supported by evidence from production, showing that 3-year-olds use both T3 allophonic variants in appropriate phonological contexts (Tang et al., submitted).

These findings raise several questions for further research. Firstly, how might younger children (i.e., 2-year-olds) perform on the current task? When do children develop both allophonic variants for T3? Secondly, is the direction of accepting the mispronunciations between *T2 and half-T3 symmetric? That is, will children also accept the *T2 mispronunciation for the target half-T3 form? According to the Featurally Underspecified Lexicon (FUL) model, listeners are sensitive to mispronunciations from the underspecified form to fully specified forms but not vice versa (Lahari & Reetz, 2002). Given that half-T3 is not a tone category in Mandarin, exploring the perceptual (as)symmetry in mispronunciation tasks between *T2 and half-T3 will further inform the nature of the allophonic variants of T3.

In summary, our results suggest that, in recognizing novel tone sandhi words, Mandarin-speaking 3-5-year-olds exhibit flexibility in accepting allophonic mispronunciations of T3 syllables. This suggests that children have developed an abstract phonological representation for T3 in their mental lexicon, including its allophonic variants. Our findings therefore highlight the role of phonological processes on shaping children's phonological representations, deepening our understanding of children's ability to accommodate surface phonological variants in learning new words. The novel items used in the present study thus provide a framework in examining children's phonological representations when productive processes were involved, providing further insight into children's ability to contend with phonological processes in learning new words.

Acknowledgements

We thank the Child Language Lab, the Phonetics Lab, and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and supports. We also thank the help from the Affiliated Kindergarten of Beijing Language and Culture University in data collection. This research was supported, in part, by a Macquarie University iMQRES scholarship to the first author, and the following grants: ARC CE110001021, ARC FL130100014. The equipment was funded by MQSIS 9201501719.

References

Albright, A., & Bruce, H. (2011). Learning and learnability in phonology. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The Handbook of Phonological Theory*. 2nd edn. Oxford: Wiley-Blackwell. 661–690.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. [Computer program]. Version 6.0. 19.

Chen, M. Y. (2000). *Tone Sandhi: Patterns Across Chinese Dialects* (Vol. 92). Cambridge University Press.

Chen, Y., Shen, R., & Schiller, N. O. (2011). Representation of allophonic tone sandhi variants. In *Proceedings of Psycholinguistics Representation of Tone*. Satellite Workshop to ICPhS, Hongkong (pp. 38-41).

Fernald, A., McRoberts, G. W., and Swingley, D. (2001). "Infants' developing competence in recognizing and understanding words in fluent speech," in *Approaches to Bootstrapping: Phonological, Syntactic and Neurophysiological Aspects of Early Language Acquisition*, eds J. Weissenborn and B. Hoehle (Amsterdam: John Benjamins), 97–124.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, 27(1), 3-42.

Kerkhoff, A. O. (2007). *Acquisition of Morpho-phonology: The Dutch Voicing Alternation* (Doctoral dissertation, LOT).

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology*, 7, 637-675.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Softwar*e, 69(1), 1-33.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23-45.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395-421.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: tests in linear mixed effects models. R Package, version 2.0-20. Vienna: R Foundation for Statistical Computing.

R Core Team (2016). "R: A language and environment for statistical computing" [Computer program]. Version 3.3.1, https://www.R-project.org/ (Last viewed June 21, 2016). R Package, version 2(0).

Shi, R., Gao, J., Achim, A., & Li, A. (2017). Perception and representation of lexical tones in native Mandarin-learning infants and toddlers. *Frontiers in Psychology*, 8, 1117.

Skoruppa, K., Mani, N., & Peperkamp, S. (2013). Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child Development*, 84(1), 313-330.

Song, J. Y., Shattuck-Hufnagel, S., & Demuth, K. (2015). Development of phonetic variants (allophones) in 2-year-olds learning American English: A study of alveolar stop/t, d/codas. *Journal of Phonetics*, 52, 152-169.

Tang, P., Yuen, I., Xu Rattanasone, N., & Demuth, K. (submitted). The acquisition of phonological alternations: the case of the Mandarin tone sandhi process.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(5), 2823-2835.

Van de Vijver, R. & Baer-Henney, D. (2013). On the development of the productivity of plural suffixes in German. In Proceeding of *the 37th Annual Boston University Conference on Language Development* (pp. 444-455).

Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, 7, 627.

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, 48(5), 1065-1079.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology*, 9, 817.

Yip, M. (2002). *Tone*. Cambridge University Press.

Zamuner, T. S., Kerkhoff, A., & Fikkert, P. (2012). Phonotactics and morphophonology in early child language: Evidence from Dutch. *Applied Psycholinguistics*, 33(3), 481-499.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes*, 9(3), 393-422.

Appendix

Appendix 1. Stimuli adopted, including four conditions: Non-Sandhi Target, Non-Sandhi-Mispronunciation, Sandhi-Target and Sandhi-Mispronunciation.

Type	Stimuli (underlying tones)	Meaning
Non-Sandhi (Tar: **T2**T1; Mis: ***T3**T1)	niu2 shu1	*cow-book*
	niu2 deng1	*cow-lamp*
	e2 che1	*goose-car*
	e2 deng1	*goose-lamp*
	yang2 shu1	*sheep-book*
	yang2 guo1	*sheep-pot*
	yu2 che1	*fish-car*
	yu2 guo1	*fish-pan*
Sandhi (Tar: ***T2**T3; Mis: ***T3**T3)	ma3 gu3	*horse-drum*
	ma3 biao3	*horse-watch*
	gou3 san3	*dog-umbrella*
	gou3 tong3	*dog-barrel*
	hu3 san3	*tiger-umbrella*
	hu3 biao3	*tiger-watch*
	niao3 gu3	*bird-drum*
	niao3 tong3	*bird-barrel*

Appendix 2. Averaged acoustic parameters of both syllables within each stimulus across conditions

Condition	Syllable	Pitch onset (Hz)	Pitch offset (Hz)	Minimum pitch (Hz)	Maximum pitch (Hz)	Mean pitch (Hz)	Duration (ms)
Non-Sandhi-Target	1	223	337	219	337	251	466
	2	333	339	333	339	335	528
Non-Sandhi-Mispronunciation	1	224	165	165	224	191	466
	2	333	339	333	339	335	528
Sandhi-Target	1	223	337	219	337	251	466
	2	228	195	161	228	183	619
Sandhi-Mispronunciation	1	224	165	165	224	191	466
	2	228	195	161	228	183	619

Appendix 3. Results of the linear mixed-effects model on the difference scores of looking to target, with three fixed factors: "Pronunciation" (Target pronunciation and Mispronunciation), "Condition" (Non-Sandhi and Sandhi) and "Group" (3-, 4-, 5-year-olds and Adults). The bold font indicates significant effects.

Factors	**df 1**	**df 2**	**F**	***p***
Pronunciation	**1**	**136**	**31.88**	**<0.001*** **
Condition	**1**	**188**	**8.42**	**<0.01** **
Group	**3**	**113**	**12.97**	**<0.001*** **
Pronunciation × Condition	**1**	**223**	**32.16**	**<0.001*** **
Pronunciation × Group	3	137	2.53	0.06
Condition × Group	**3**	**188**	**4.08**	**<0.01** **
Pronunciation × Condition × Group	3	223	1.64	0.182

Appendix 4. Pairwise comparison of the difference ratio between target pronunciations and mispronunciations across conditions and groups. The bold font indicates significant effects.

Condition	Group	β	SE	df	*t*	*p*
Non-sandhi	**3-year-olds**	**0.15**	**0.05**	**271**	**3.26**	**<0.01****
	4-year-olds	**0.12**	**0.04**	**280**	**2.97**	**<0.01****
	5-year-olds	**0.15**	**0.05**	**271**	**2.96**	**<0.01****
	Adults	**0.31**	**0.04**	**275**	**6.94**	**<0 .001*****
Sandhi	3-year-olds	-0.02	0.05	276	-0.34	0.733
	4-year-olds	-0.01	0.04	271	-0.14	0.893
	5-year-olds	0.03	0.05	277	0.50	0.621
	Adults	0.03	0.04	271	0.57	0.573

Appendix 5. Results of the cluster-based permutation tests.

Condition	Group	Cluster	Direction	SumStatistic	StartTime (ms)	EndTime (ms)	*p*
Non-sandhi	**3-year-olds**	**1**	**Positive**	**31.04**	**5600**	**7600**	**<0.001*** **
	4-year-olds	1	Positive	4.00	1000	1400	0.399
		2	**Positive**	**40.31**	**6000**	**8000**	**<0.001*** **
		3	Negative	-2.64	600	800	0.441
	5-year-olds	1	Positive	2.53	4400	4600	0.526
		2	Positive	10.00	5400	6200	0.069
		3	**Positive**	**31.40**	**6400**	**8000**	**<0.01** **
	Adults	**1**	**Positive**	**103.41**	**5200**	**8000**	**<0.001*** **
Sandhi	3-year-olds	1	Positive	2.08	6600	6800	0.653
		2	Negative	-2.13	6000	6200	0.623
	4-year-olds	-	-	-	-	-	-
	5-year-olds	1	Negative	-2.06	5400	5600	0.756
	Adults	1	Positive	2.21	800	1000	0.595
		2	**Positive**	**12.79**	**6000**	**6800**	**<0.05***

Chapter Eight: Children with Cochlear Implants' Representation of Allophonic Variants of Tone Sandhi

This chapter is based on the following paper:

Tang, P., Xu Rattanasone, N., Yuen, I., Gao, L., Demuth, K. (in preparation). Children with cochlear implants' representation of allophonic variants of tone sandhi. In preparation to submit to the *Preceding of the 19th International Congress of Phonetic Sciences* (*ICPhS 2019*).

All components of this paper, both experimental and written, have been completed by me, with advice from the co-authors (my supervisors and collaborators) when needed.

Abstract

Children with cochlear implants (CIs) face challenges in acquiring tonal languages, as CIs do not efficiently code pitch information. In connected speech, tones undergo modifications, known as tone sandhi processes. In Mandarin, tone sandhi processes change the realization of tone 3 (T3) across tonal contexts, resulting in allophonic variants. Previous studies show that 3-year-olds with normal hearing (NH) correctly produce T3 variants in appropriate sandhi contexts, while children with CIs generally produced unchanged pitch for T3 across contexts unless implanted early (Tang et al., submitted a, b). However, it was unclear whether they have the correct representation of T3 variants in their mental lexicon. The aim of the present study was to explore the phonological representation of allophonic variants of T3 for words that undergo tone sandhi of children with CIs', and whether early implantation helps them build better T3 representations. We tested 46 prelingually deaf 3-7-year-olds implanted from 1 to 5 years and 32 3-year-old NH controls. Using a similar approach of the study of Tang et al. (submitted c) on NH children, we tested children with CIs' sensitivity to allophonic mispronunciations of novel tone sandhi compounds, using the intermodal preferential looking paradigm. The results showed that these children mapped neither the correct pronunciation nor to familiar words or mispronunciation to novel items. This indicates that children with CIs face challenges in recognizing and processing novel words as well as in detecting tonal changes. The implications for future studies are discussed.

Key words: phonological processes; phonological representation; Mandarin Chinese, tone sandhi; cochlear implants; language development

Introduction

Cochlear implant (CI) technology has made speech communication an obtainable goal for children with hearing impairment, as reflected by their improved speech production and perception abilities after implantation (Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000). However, for tone language-learning children with CIs, building a tonal system can be challenging, since CIs do not transmit pitch information effectively (Vandali & van Hoesel, 2012).

Mandarin is a tone language using four lexical tones in addition of consonants and vowels to contrast word meanings (Yip, 2002). These four tones are primarily contrastive in pitch contours, i.e., level tone 1 (T1), rising tone 2 (T2), dipping tone 3 (T3) and falling tone 4 (T4; Figure 1). Previous studies have shown early acquisition of these four lexical tones in children with normal hearing (NH), i.e., before age 3, with correct pitch implementations (Tang et al., submitted a).

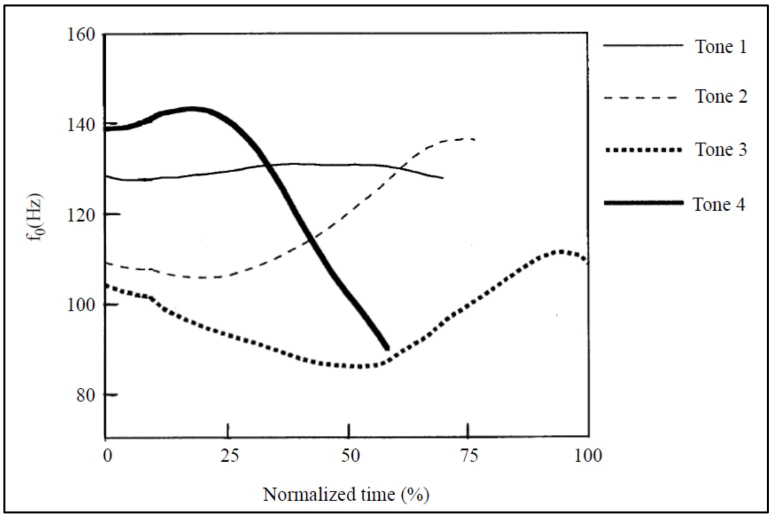


Figure 1. Mandarin lexical tone pitch contours, from Xu (1997).

For children with CIs, acquiring typical lexical tones is challenging. It is generally agreed that, relative to NH children, the tone perception and production accuracy of children with CIs is lower, but **early implantation** typically predicts better outcomes in tone acquisition (Han et al., 2007; Zhou et al., 2013; Tan, Rowell & Vogel, 2016; Chen & Wong, 2017; Tang et al., submitted b). In perception, for instance, Zhou et al. (2013) reported that the tone identification accuracy was 67% (SD: 13.51%) for 2-16-year-old children with CIs (implanted between 1 to 12 years; Mean: 3.96 years; SD: 2.70 years), while it was 99% (SD: 2.67%) for NH children (3-10 years). Among the four lexical tones, **T2 and T3** were the most difficult tones for these children to perceive (Chen et al., 2014; Mao & Xu, 2017). In addition, Zhou et al. (2013) reported that children with CIs' tone perception performance was significantly correlated with age at implantation, suggesting that early implantation is critical for them to perceive lexical tones accurately. In production, Tang et al. (submitted b) compared the acoustic features of lexical tone productions for 3-7-year-old prelingually deaf children with CIs (implanted between 1 to 5 years; Mean: 3 years; SD: 11.7 years) and 3-year-old NH children. It was found that only those implanted before 2 years produced target-like pitch contours for lexical tones, with other children generally producing much flatter pitch contours, including those with more than 4 years of CI experience, again suggesting that early implantation is critical for children with CIs to acquire typical lexical tone productions.

In daily conversation, however, people talk to each other using connected speech instead of single words, where lexical tones undergo tonal modifications known as tone **sandhi** processes (Chen, 2000). The most well-studied tone sandhi process in Mandarin Chinese is T3 sandhi (hereafter "tone sandhi"). Depending on the following lexical tones, tone sandhi processes modify the surface realization of T3 syllables, generating two allophonic variants (Chen, 2000). In a T3+T3 sequence, where **full sandhi** occurs, the initial T3 is realized as the *rising* tone similar to T2 (referred to as *T2 to differentiate it from the underlying lexical T2): T3+T3→ *T2+T3. In a T3+T1/T2/T4 sequence, where **half sandhi** occurs, T3 is realized with a *low-falling* contour (referred to as half-T3): T3+T1/2/4 → half-T3+T1/2/4.

It has been shown that Mandarin-speaking NH children have mastered the tone sandhi process before age 3 (Tang et al., submitted a). Using a novel word formation task, Tang et al. (submitted a) examined NH 3-5-year-olds' productions of novel tone sandhi compounds such as "ma3 gu3" *horse-drum*. The results showed that all child groups correctly produced both allophonic variants of T3 in appropriate sandhi contexts, i.e., *T2 before T3 and half-T3 before T1/2/4, suggesting that 3-year-olds have already developed a good understanding of the allophonic variants of T3 and can use these productively in tone sandhi processes.

For children with CIs, correctly producing the allophonic variants of T3 in appropriate sandhi contexts is challenging; even those with a long CI experience (more than 4 years) generally produced invariant T3 with level pitch contours irrespective of the tonal context (Tang et al., submitted b). However, it was unclear whether they represent both allophonic variants of T3 in their mental lexicon. Moreover, it has been found that children implanted before age 2 were able to correctly produce both variants of T3 in appropriate contexts (Tang et al., submitted b). This raises the question of whether T3 is represented differently for children with different implantation ages. To further explore these issues, it is therefore necessary to extend the investigation of tone sandhi acquisition from speech production to speech perception, and examine the lexical representation of allophonic variants of T3 in children with CIs.

A recent study tested the representation of allophonic variants of T3 resulting from tone sandhi in NH 3-5-year-olds, using intermodal preferential looking (IPL) task (Golinkoff, et al., 1987; Tang et al., submitted c). In that study, NH 3-5-year-olds were presented with a pair of pictures, illustrating a tone sandhi object (novel T3T3 compound), i.e., "ma3 gu3" *horse-drum*, and a novel distractor. An audio stimulus was played with either *T2T3 correct sandhi tones, or half-T3T3 mispronunciations. The results showed that NH children, as well as adult controls, accepted both the *T2T3 correct pronunciation and the half-T3T3 mispronunciation as possible surface realizations for underlying T3T3 items, indicating that both allophonic variants of T3 were accessed in recognizing T3 syllables in novel tone sandhi compounds. Thus, both variants (i.e., *T2 and half-T3) are represented in NH 3-5-year-olds' mental lexicon, together under the same abstract T3 category. This raises the question of whether children with CIs represent allophonic variants of T3 in the same way as NH children.

The aim of the present study was therefore to examine the perceptual representation of the allophonic variants of T3 resulting from tone sandhi processes in children with CIs. Using a procedure similar to that used in Tang et al. (submitted c) with NH children, we asked if children with CIs would accept the half-T3T3 'mispronunciation' for the target *T2T3 forms for novel T3T3 compounds, as NH children did. Based on previous findings that children with early implantation used both variants of T3 in production (Tang et al., submitted b), we predicted that perhaps they would also accept both pronunciations in perception, while those with later implantation will accept *T2T3 only.

Given previous findings that children with CIs have problems in perceiving the difference between T2 and T3, it is also possible that they might not detect the acoustic changes between correct pronunciations and mispronunciations. Therefore, we also included non-sandhi T2T1 compounds (with half-T3T1 mispronunciations) as control items, where the target and mispronounced syllable 1 was acoustically identical to the sandhi items, but where no sandhi process was involved. We predicted that children implanted later would show difficulties in detecting the tonal changes in the control condition.

Methods

Participants

Participants included 40 3-7-year-old (mean age: 4;11; SD: 10.4 months) prelingually deafened Mandarin-speaking children with severe to profound hearing loss and 32 3-year-old NH controls. Children with CIs were recruited from speech rehabilitation centres in Beijing, China. They were implanted at between 13 to 58 months of age (see Figure 2; Appendix 1). All were unilaterally implanted, and used a hearing aid (HA) for the contralateral ear, wearing both CI and HA devices during the experiment. According to reports from their respective institutions, none of these children had intellectual difficulties. Additional six children with CIs were excluded given that they might not understand the task (see the following analysis section for more details of exclusion). Based on the age at implantation, the 40 participants were grouped into four groups: 1-2 years (CI_Imp 1-2: 3 children); 2-3 years (CI_Imp 2-3: 18 children); 3-4 years (CI_Imp 3-4: 13 children); and 4-5 years (CI_Imp 4-5: 6 children). The NH controls (referred as: NH = 3; mean age: 3;8; SD: 3 months) were recruited from the affiliated kindergarten of Beijing Language and Culture University.

The study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Panel.

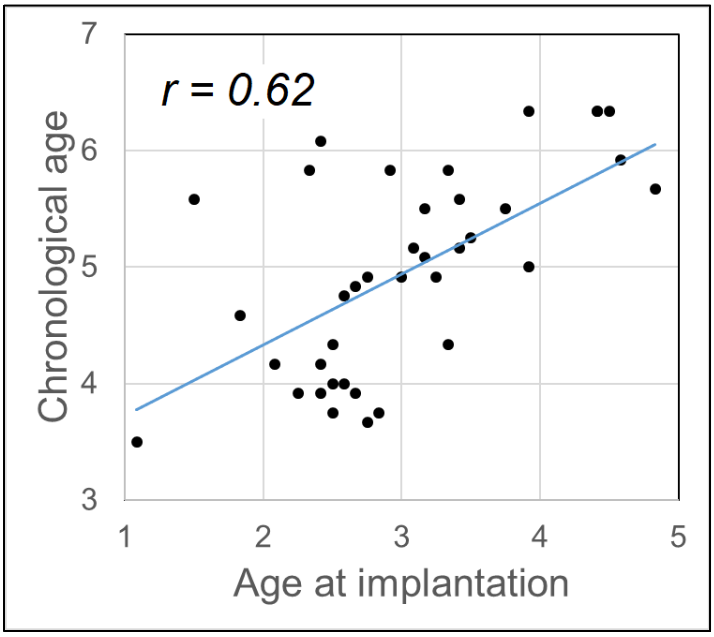


Figure 2. Ages at implantation and chronological ages of children with CIs.

Stimuli

Two types of disyllabic "animal + object" novel compounds were used as stimuli, including eight *non-sandhi* T2T1 control items, e.g., "niu2 deng1" *cow-bulb*, and eight *tone sandhi* (underlying T3T3) target items, e.g., "ma3 gu3" *horse-drum*, with the surface sandhi realization ***T2**T3. Following the methods used in the study of Tang et al., (submitted c), tonal mispronunciations were obtained by changing the target syllable (underlying T2 and surface *T2) to half-T3, resulting in eight *non-sandhi* and eight *sandhi* mispronunciations (see Table 2; also see Tang et al. submitted c, for more details).

Table 2. Novel compound tonal stimuli used in this study.

	Non-sandhi (underlying T2T1)	**Sandhi** (underlying T3T3)
Target surface tones	T2T1	*T2T3
Mispronounced surface tones	half-T3T1	half-T3T3

Audio stimuli

All the target speech stimuli were recorded by a female Beijing-Mandarin speaker in a child-directed speech style. Using the same method of Tang et al. (submitted c), mispronunciations were made via replacing the pitch of the target T2/*T2 syllables with that of half-T3 syllables, using the Praat software (Boersma & Weenink, 2016; Figure 3). The Syllable 1 non-sandhi control target and mispronounced tones are acoustically identical to the target and mispronounced sandhi forms. Thus, any differences between these two conditions will be due to phonological, not phonetic, differences.

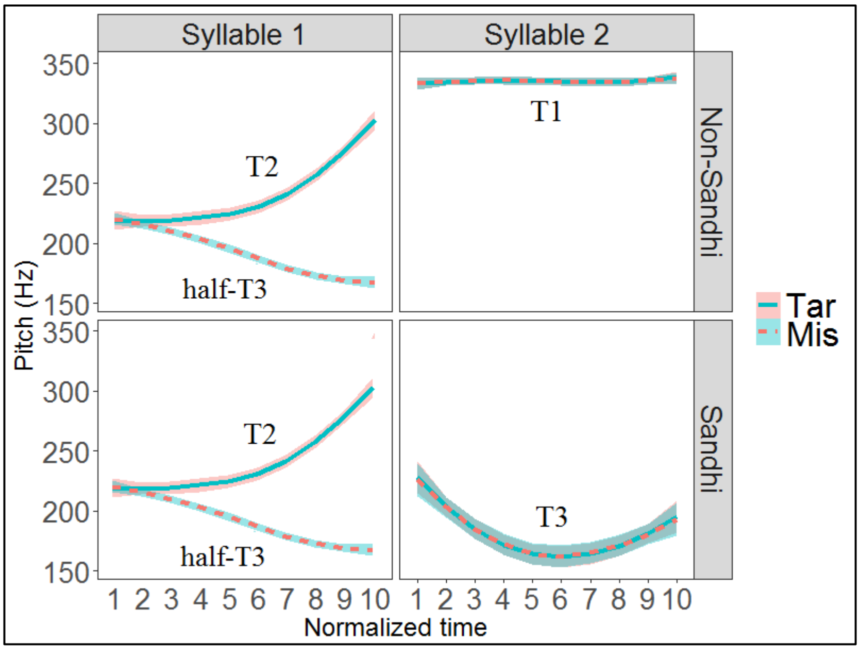


Figure 3. Pitch contours of syllables 1 and 2 within each stimulus for target pronunciations (Tar) and mispronunciations (Mis) in Non-Sandhi and Sandhi conditions. The tonal mispronunciation was made on the first syllable, changing it from *T2/T2 to half-T3.

Visual stimuli

Each audio stimulus was matched with a pair of yoked familiar vs. novel pictures, e.g., the familiar picture for “ma3 gu3” *horse-drum* illustrates a drum with a horse on it, was yoked to a novel picture illustrating a drum with a novel animal on it. During the experiment, the yoked familiar-novel pictures were displayed side-by-side horizontally on a 15.6 inch (38.7 cm * 25.9 cm) white screen.

Apparatus

Audio stimuli were played via speakers (Logitech Z120, for children with CIs) or a headphone (AKG K612 Pro, for NH children) at approximately 70 dB. The SMI

Red-250 portable eye-tracker was used to record participants' visual fixations with a 9-point calibration procedure prior to testing.

Procedure

All participants were tested individually in quiet rooms in their own rehabilitation centre (children with CIs) or kindergarten (NH children). Prior to testing, a picture naming task had been conducted to ensure that all participants knew the underlying tone of each word used to form the novel compounds in the perception experiment (cf. Tang et al., submitted c).

The procedure of the perception experiment was identical as that used in Tang et al. (submitted c; also see Figure 4), i.e., three practice trials followed by 19 test trials. Within the 19 test trials, there were three *word*-mispronunciation trials (where Syllable 1 was replaced with a syllable that differed from the target in segments and tone), used as a screener to exclude any participant who did not understand the task (see the exclusion section below for more details).

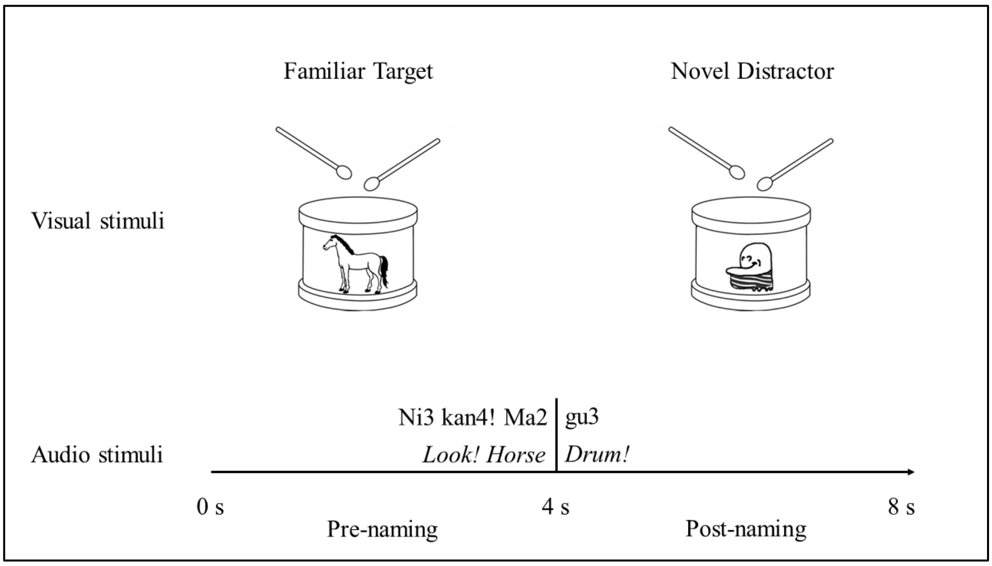


Figure 4. Sequence of events in each trial.

Data analysis

Data analysis was identical as that used in Tang et al. (submitted c), including (1) the difference ratio of looks to the familiar picture and (2) the time course analysis. The analysis of difference ratio of looks provides information about the overall effect of the auditory stimuli on participants' preference between the familiar and novel pictures. The time course analysis provides additional fine-grained information regarding fixation patterns that may be obscured by the aggregate difference measures.

A total of 1248 trials were originally obtained. Participants were screened using the three word-mispronunciation trials, for which we expected negative difference ratios of looking to the familiar picture in at least one trial. Six children with CIs were initially excluded from further analysis for not reaching this criterion. For the remaining participants, 87 trials with greater than 50% track-loss data were then also excluded. This resulted in a total of 40 children with CIs (553 trials) and 32 NH 3-year-olds (512 trials) included in the final analysis.

Results

Difference ratio

A series of one-sample *t*-tests (with the Bonferroni adjustment) were computed to compare the difference ratios to baseline (difference ratio = 0) across conditions and groups (Figure 5). These comparisons explored whether children were able to associate the audio stimuli with the visual objects, i.e., familiar or novel pictures, across conditions. The results showed that NH 3-year-olds looked to target pictures in three conditions: non-sandhi, sandhi correct and sandhi mispronunciations.

However, children with CIs generally did not look to either familiar or novel picture upon hearing the audio stimuli, except for the CI_Imp 2-3 and CI_Imp 3-4 groups that looked to target pictures upon hearing non-sandhi correct pronunciations and mispronunciations. This result suggests that NH children accepted non-sandhi and sandhi correct pronunciations and sandhi mispronunciations, but most CI groups (even the few with early implantation) did not associate the audio stimuli with the visual objects. This indicates that children with CIs could not identify correct production of lexical tone and tone sandhi words.

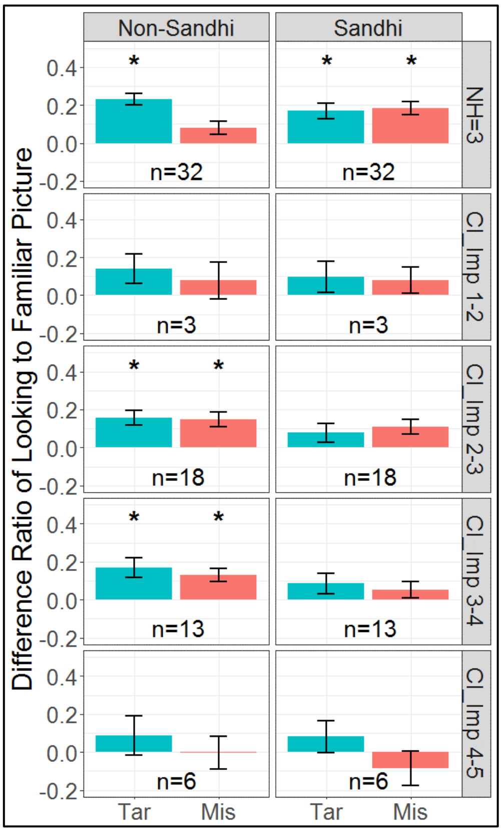


Figure 5. Difference ratios of looking to the familiar picture for the target pronunciations (Tar) and mispronunciations (Mis) across conditions (Non-Sandhi and Sandhi) and groups (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5). Asterisks indicate significant difference compared to the baseline (i.e., difference ratio = 0, indicating no preference on either familiar or novel picture upon hearing the audio stimuli). N indicates the number of participants in each group.

A linear-mixed effects model was also performed to compare the difference ratio of looks to the familiar picture between the correct and mispronunciation trials across conditions and groups, with three fixed factors “Groups” (NH=3, CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5), “Condition” (Non-Sandhi and Sandhi) and “Pronunciation” (Target and Mispronunciation) and a random factor “Participant”. The results are presented in Table 2, showing that there was a significant interaction of “Group × Condition × Pronunciation”. A post-hoc test was then performed on this interaction to compare the difference ratio between target and mispronunciation trials across groups and conditions. The results showed that, in the non-sandhi control condition, 3-year-olds NH looked significantly less to the target picture when hearing the mispronunciations ($\beta = 0.15$, SE = 0.07, $t(536)=1.97$, $p<0.05$). However, CI groups did not show any difference of looking between target pronunciation and mispronunciation trials. This result indicates that children with CIs, including those with early implantation, did not detect the tonal changes between correct and mispronunciations even in the control condition.

Table 2. Results of the linear mixed-effect model of the difference ratio across Implantation age (CI_Imp 1-2, CI_Imp 2-3, CI_Imp 3-4 and CI_Imp 4-5), Condition (Non-Sandhi and Sandhi) and Pronunciation (Target and Mispronunciation).

Factor	**df 1**	**df 2**	**F**	***p***
Group	4	51	1.62	0.183
Condition	1	538	1.70	0.193
Pronunciation	1	538	3.19	**<0.05***
Group × Condition	4	538	0.60	0.664
Group × Pronunciation	4	538	0.79	0.532
Condition × Pronunciation	1	537	0.37	0.544
Group × Condition × Pronunciation	**4**	**538**	**2.48**	**<0.05***

Time course analysis

We now move to the time course analysis, trying to explore more fine-grained differences of children's gaze patterns between correct and mispronunciations (Figure 6). Cluster-based permutation tests were performed to find the divergence period between the correct and mispronunciation trials across conditions and groups. However, the results only produced a divergence window of NH 3-year-olds in the non-sandhi condition (Figure). This result further revealed that, consistent with the analysis of difference ratio, NH 3-year-olds looked less to the target picture upon hearing the non-sandhi mispronunciations, and children with CIs did not detect any difference between target pronunciations and mispronunciations in any condition.

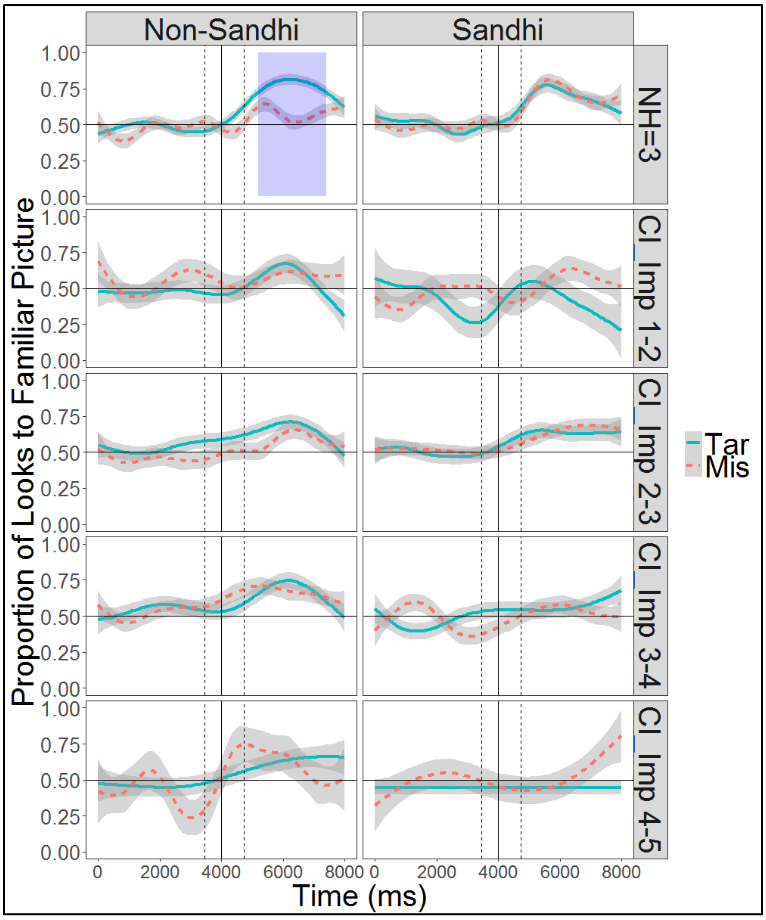


Figure 6. Time-courses of looking proportion toward target of correct and mispronunciations for Non-Sandhi and Sandhi conditions across groups. The horizontal solid lines indicate the change level (proportion = 0.5) of looking toward target, and the vertical solid lines indicate the onset of the second syllable within each stimulus, marked as the boundary between pre- and post-naming stages. Dash lines mark the averaged onset and offset points of stimuli. Blue shadows indicate the significant divergence between two curves based on cluster-based permutation tests.

Discussion

The present study examined children with CIs' representation of the allophonic variants of T3 using IPL paradigm (Golinkoff, et al., 1987). The results showed that NH children accepted both *T2T3 target pronunciations and half-T3T3 mispronciations for underlying T3T3 compounds, indicating that they had accessed both *T2 and half-T3 in processing novel T3 words in the tone sandhi context, suggesting that they have represented both allophonic variants of T3 in their mental lexicon, together under the same T3 category. However, the result of children with CIs differed from that of NH children in two aspects. First, most CI groups, including those with early implantation (i.e., before age 2), were not able to associate the audio stimuli with the visual objects; they did not show any preference to either familiar or novel picture upon hearing the audio stimuli. Second, detecting the tonal mispronunciation between T2/*T2 and half-T3 is challenging for these children; none of the CI groups showed a looking difference between the target pronunciation and mispronunciation trials. Thus, it remains unclear when and how allophonic variants of T3 are lexically represented in children with CIs.

The poor performance of associating the audio stimuli to the target picture by these children might be related to the task used in this study, i.e., novel word recognition. It has been well-documented that children with CIs exhibit difficulties in learning new words, performing more poorly than age-matched NH children in building word-object associations during word learning processes (Houston et al., 2005, 2012; Davidson, Geers & Nicholas, 2014). This is because, on the one hand, the auditory information provided by cochlear implants is impoverished and highly degraded and, on the other hand, the period of early sensory deprivation prior to implantation may lead to a delayed and/or disordered course of language development

(Houston et al., 2005). In the present study, our task required children to recognize novel compounds that they have never heard before, e.g., *'horse-drum'*, and link these novel compounds to novel pictures, e.g., a drum with a horse on it, which they had not seen before. This novel word-picture matching task might require a high word-learning and phonological processing ability, which is limited for children with CIs, resulting in their confusion/inability to perform the task.

Apart from the novel word recognition task, another challenge that children with CIs faced in this experiment was to detect the tonal mispronunciation between T2/*T2 and half-T3, which might be subtle and hard to perceive for these children. It is well documented that children with CIs face great challenges in correctly differentiating between T2 and T3 (e.g., Chen et al., 2014; Mao & Xu, 2016). This is because the pitch contours of T2 and T3 are very similar, and correctly differentiating between them requires children to work out the subtle contour difference using the fine-grained pitch information, which is severely degraded in CI devices (Vandali & van Hoesel, 2012). This also indicates that children with CIs face challenges in building robust representations of T2 and T3. This is supported by the previous production experiment which found that children with CIs could not produce typical T2 and T3 productions (Tang et al., submitted b). In the current experiment, our task requires these children to detect tonal mispronunciations of underlying T2 (in the non-sandhi control condition) and underlying T3 words (in the sandhi condition). Our results suggest that they do not have robust phonological representations of either T2 or T3, thus it is hard for them to detect these tonal mispronunciations.

One of the central questions we asked is whether early implantation would benefit children with CIs to develop typical phonological representation of allophonic variants of T3. Thus, we expected that early implanted group would perform in the same way as NH children, i.e., reject the tonal mispronunciation in the control condition but accept it in the sandhi condition (Tang et al., submitted c). However, the results did not show this pattern, with no preference to either target picture or picture upon hearing the audio stimuli. This might be related to the insufficient number of participants in the early implantation group, i.e., only five early implanted children were initially tested and only three of them were included in the analysis after exclusion based on performance of the three *word*-mispronunciation trials. Therefore, based on the current study, it is hard to draw a strong conclusion regarding the effect of early implantation of the development of T3 representations.

Although it is still unclear how allophonic variants of T3 are lexically represented in children with CIs, our study has implications for future research to further explore this issue. First, given these children's poor ability in recognizing and processing novel compounds, it will be helpful in future studies to test their performance on familiar/lexicalized tone sandhi words, i.e., 'yu3 san3' *umbrella*, like what has been done in the study of Wewalaarachchi and Singh (2016) on NH children, which might require less effort for children with CIs. Second, given the insufficient number of participants in the critical early implantation group, it is necessary in the future studies to test more children with early implantation, i.e., implanted before age 2 or even age 1, to gain a deeper understanding of the effect of early implantation on the development of tonal representations.

In summary, the current study examined the phonological representation of allophonic variants of T3 from tone sandhi in lexical representations of children with CIs, using the IPL paradigm. The results showed that these children mapped neither the correct nor the mispronunciations to either familiar or novel picture. This indicates that children with CIs face challenges in recognizing and processing new words and detecting tonal changes involving T2 and T3. This suggests that future research could use familiar words, testing greater numbers of early implanted children, to further explore this issue.

References

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. [Computer program]. Version 6.0. 19.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press.

Chen, Y., Wong, L. L., Chen, F., & Xi, X. (2014). Tone and sentence perception in young Mandarin-speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 78(11), 1923-1930.

Chen, Y., & Wong, L. L. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology*, 56(sup2), S7-S16.

Davidson, L. S., Geers, A. E., & Nicholas, J. G. (2014). The effects of audibility and novel word learning ability on vocabulary level in children with cochlear implants. *Cochlear Implants International*, 15(4), 211-221.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23-45.

Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., & Xu, L. (2007). Tone production of Mandarin Chinese speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 71(6), 875-880.

Houston, D. M., Carter, A. K., Pisoni, D. B., Kirk, K. I., & Ying, E. A. (2005). Word learning in children following cochlear implantation. *The Volta Review*, 105(1), 41.

Houston, D. M., Stewart, J., Moberly, A., Hollich, G., & Miyamoto, R. T. (2012). Word learning in deaf children with cochlear implants: Effects of early auditory experience. *Developmental Science*, 15(3), 448-461.

Mao, Y., & Xu, L. (2017). Lexical tone recognition in noise in normal-hearing children and prelingually deafened children with cochlear implants. *International Journal of Audiolog*y, 56(sup2), S23-S30.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science*, 11(2), 153-158.

Tan, J., Dowell, R., & Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American Journal of Audiology*, 25(3), 246-256.

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (submitted a). The acquisition of phonological alternations: the case of the Mandarin tone sandhi process.

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (submitted b). The acquisition of Mandarin tonal processes by children with cochlear implants.

Tang, P., Xu Rattanasone, Yuen, I., N., Gao, L., & Demuth, K. (submitted c). The representation of allophonic variants of tone sandhi in the developing lexicon.

Vandali, A. E., & van Hoesel, R. J. (2012). Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking. *The Journal of the Acoustical Society of America*, 132(1), 392-402.

Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, 7, 627.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83.

Yip, M. (2002). *Tone*. Cambridge University Press.

Zhou, N., Huang, J., Chen, X., & Xu, L. (2013). Relationship between tone perception and production in prelingually-deafened children with cochlear implants. *Otol Neurotol*, 34(3), 499-506.

Appendix

Appendix 1. Demographic information of children with CIs

Participant number	Gender	Residual hearing (L/R in dB)	CI Brand	Age at implantation (in months)	Amount of CI experience (in months)	Chronological age (in months)
1	F	100/100	MEDEL	53	22	76
2	F	100/100	MEDEL	40	11	52
3	M	98/110	MEDEL	36	22	59
4	M	100/110	Cochlear	28	42	70
5	M	100/105	MEDEL	54	21	76
6	M	100/110	Cochlear	47	29	76
7	F	100/110	AB	29	43	73
8	M	90/110	Cochlear	55	16	71
9	M	100/100	MEDEL	29	21	50
10	F	97/97	MEDEL	34	10	45
11	M	97/97	MEDEL	30	14	45

12	F	100/90	MEDEL	52	14	66
13	M	80/100	MEDEL	22	33	55
14	F	96/96	MEDEL	31	16	48
15	M	97/97	MEDEL	26	10	37
16	M	100/105	MEDEL	35	35	70
17	M	97/97	Cochlear	40	29	70
18	M	100/90	MEDEL	38	23	61
19	F	80/100	Cochlear	58	10	68
20	M	96/96	Cochlear	18	48	67
21	M	97/97	Cochlear	38	28	66
22	M	110/110	Cochlear	41	20	62
23	M	95/95	MEDEL	13	29	42
24	F	110/110	MEDEL	30	18	48
25	M	110/110	Cochlear	27	20	47

26	M	80/110	MEDEL	32	14	47
27	M	110/100	Cochlear	30	22	52
28	F	110/100	MEDEL	30	34	64
29	F	100/90	Cochlear	40	16	56
30	M	97/97	MEDEL	32	26	58
31	M	110/100	MEDEL	53	34	87
32	F	100/100	Cochlear	39	19	59
33	F	80/110	MEDEL	41	25	67
34	F	110/100	Cochlear	30	28	59
35	M	95/95	Cochlear	47	13	60
36	M	115/115	MEDEL	33	26	59
37	F	100/100	Cochlear	45	20	66
38	M	100/70	Cochlear	31	25	57
39	M	110/110	Cochlear	25	25	50

40	F	100/100	MEDEL	42	20	63

Chapter Nine: Thesis Discussion

Discussion

In this thesis, we explored the acquisition of Mandarin contextual tones, i.e., neutral tone and tone sandhi, by both typically developing children and children with CIs. Previous studies suggested that contextual tones are later acquired by typically developing children (Li & Thompson, 1977; Hua & Dodd, 2000; Wang, 2011; Xu Rattanasone, et al., 2018), with the reasons remaining unclear. We therefore examined the potential reasons from the perspectives of early language input (Chapters 2 & 3) and the variable surface realization of contextual tones (Chapters 4 & 5). Furthermore, previous studies suggested that children with CIs have problems in acquiring lexical tones unless implanted early or have long CI experience, since CIs do not transmit pitch information effectively (Tan, Dowell & Vogel, 2016; Chen & Wong, 2017). We therefore examined the contextual tone productions from children with CIs and explored the effect of early implantation and long CI experience in their tonal development (Chapter 6). Finally, we extended the tonal acquisition investigation from speech production to speech perception, examining how tone sandhi forms are represented in children's mental lexicon (Chapters 7 & 8).

We firstly explored the potential reason leading to the later acquisition of contextual tones from the perspective of input, examining the acoustic realization of contextual tones in IDS and Lombard speech (Chapters 2 & 3). The results showed that, despite both registers exhibiting tonal hyperarticulation and slower speaking rate, the key features of contextual tones are still preserved. This result therefore indicates that children's later acquisition of contextual tones, as suggested by previous studies (Li & Thompson, 1977; Hua & Dodd, 2000; Wang, 2011), is unlikely driven by insufficient input.

Our studies also extend previous research on the tonal realization in IDS and Lombard speech from lexical tones to contextual tones (Liu, Tsao & Kuhl, 2007; Zhao & Jurafsky, 2009), furthering our understanding of speech modification of these registers from single word to connected speech. The results therefore have implications for the role of early language input in children's language development. Previous studies have suggested that speech modifications in IDS might potentially facilitate children's language learning, such as phoneme perception, syntactic parsing, word-segmentation and boundary detection (see Singh et al., 2009, for a review). Our results thus provide further evidence that, in Mandarin Chinese, speech modification in IDS may enhance the acquisition of phonological processes such as tone sandhi, which might potentially facilitate children's phonological development.

After eliminating the potential effect of lack of appropriate input, we then examined another potential reason for the reported later acquisition of tones in context: the surface variation of contextual tones, exploring when children are able to develop the correct neutral tone *category* and extract the tone sandhi *process* from their variable surface realizations (Chapters 4 & 5). We firstly examined 3-5-year-olds' **neutral tone** productions across contexts in both familiar and novel items (Chapter 4). The results showed that all three age groups produced correct tonal variation of neutral tone syllables across contexts for both familiar and novel words, with 3- and 4-year-olds lengthened the neutral tone duration to a certain degree relative to adults. These results therefore suggest that children at age 3 have already developed the correct neutral tone category and are able to productively implement it on novel items, while the adult-like acoustic implementation of neutral tone is not fully established before age 5. This result is not fully consistent with previous findings that neutral tone is acquired after 4;6 (e.g., Hua & Dodd, 2000). This difference is probably driven by the different methods used, i.e., *perceptual coding* versus *acoustic analysis*. The acoustic analysis used in the present study allows us to investigate the fine-grained acoustic realization of children's neutral tone productions across contexts, which could not be observed using perceptual coding. Our results thus have implications for the acquisition of weak syllables more generally. It has been suggested that children learning English (and Dutch) tended to omit the weak syllable of a weak–strong word like *giraffe* (Gerken, 1994; Demuth, 1996). In Mandarin, the weak syllable is carried by a toneless category, i.e., neutral tone. The stress pattern of neutral tone words is always 'strong-weak', i.e., full tone + neutral tone, and therefore children do not omit the neutral tone syllable. However, they face problems in reducing the duration of neutral tone syllables, similar to findings in English (Yuen, Demuth & Johnson, 2011). Taken together, these studies therefore suggest that mastering adult-like acoustic realizations of weak syllables is a protracted process, sometimes deleting and sometimes lengthening, depending on the specific linguistic contexts of the language.

We then examined 3-5-year-olds' acquisition of **tone sandhi** processes in *novel* disyllabic and trisyllabic compounds, testing their productive knowledge of the tone sandhi *process* and their ability to apply the process in response to the trisyllabic *prosodic structures*, i.e., left-branching ((T3T3) T3) and right-branching (T3 (T3T3)) novel words. The results showed that all three age groups applied tone sandhi on disyllabic compounds, while they produced the same surface output for trisyllabic items irrespective of the prosodic structure. These results therefore suggest that children at age 3 have already mastered productive knowledge of the tone sandhi process, while mastering the application of tone sandhi in response to the prosodic context appears to be a protracted process not fully completed by the age of 5. This study therefore extends previous investigation on the acquisition of tone sandhi (Li & Thompson, 1977; Hua & Dodd, 2000; Wang, 2011; Xu Rattanasone, et al., 2018) from lexical words to novel compounds, tapping into children's knowledge of the *tone sandhi process* rather than their memory of the surface realization of lexicalized *tone sandhi words* in the input, revealing early productive knowledge of tone sandhi.

Our results thus reveal a different picture on the acquisition of phonological processes. It has been found that in many non-tonal languages, the acquisition of phonological processes is a protracted process, with children at age 3 or 4 still failing to generalize phonological processes to novel items (e.g., Kazazis, 1969; Zamuner, Kerkhoff & Fikkert, 2006; Kerkhoff, 2007; Albright & Hayes, 2011; van de Vijver & Baer-Henney, 2013). In contrast, our results show that Mandarin-learning 3-year-olds have already acquired the tone sandhi processes, with productive application on novel items they have never heard before.

In Chapter 6, we moved our investigation from typically developing children to children with CIs, examining their acquisition of contextual tones. The results showed that only early implanted children (before age 2) acquired typical lexical tone productions and correctly produced the tonal variation for *familiar* neutral tone and *disyllabic* tone sandhi items but not for *novel* neutral tone and *trisyllabic* tone sandhi items. In contrast, later implanted children with a longer CI experience (i.e., more than 4 years of CI experience) still produced unchanged tonal realization for both types of contextual tones across contexts. These results therefore confirm previous findings that early implantation is critical for these children to acquire correct lexical tones, highlighting the importance of early implantation in the tonal development (Tan, Dowell & Vogel, 2016; Chen & Wong, 2017). More importantly, our results imply that building correct lexical tone categories is an essential prerequisite before children with CIs can acquire contextual tones, since they need to build correct tonal context to trigger the appropriate contextual tones. Finally, even early implanted children (before age 2) did not fully master neutral tone in novel items and tone sandhi in larger units (i.e., trisyllabic items), indicating that to fully master contextual tones, children with CIs may need to receive implantation even earlier, perhaps before age 1.

In the final two chapters, we extended our investigation from speech production to speech perception, exploring the representation of tone sandhi forms in children's mental lexicon. In chapter 7, we examined the tone sandhi representation of typically developing children from 3 to 5 years, using the tonal mispronunciation task. The results showed that all child groups (and adults) accepted both ***T2**T3 and half-**T3**T3 as the possible surface realization for novel **T3**T3 words, suggesting that they represented both allophonic variants *T2 and half-T3 for the underlying T3 category. This result is not consistent with previous studies that used lexicalized tone sandhi words (e.g., "shou3 zhi3" *finger*) as test items and found that 3-5-year-olds accepted ***T2**T3 only (Wewalaarachchi & Singh, 2016). The difference might be driven by the stimuli being used, i.e., familiar words might only tap into children's memory/representation of the surface realization of these tone sandhi words in the input they hear, and therefore children rejected the mispronunciations that were acoustically different. In contrast, the novel items in our study tap into children's representation of the allophonic variants of T3 and their knowledge of the tone sandhi processes in the novel word formation process. Our results therefore imply that children show flexibility in accepting the allophonic mispronunciations involving phonological processes in learning new words, suggesting that the learning of phonological processes may enhance the establishment of more robust phonological representations.

In chapter 8, we carried out the same task with children with CIs, examining their perception of tone sandhi mispronunciations. The results showed that these children were not able to associate the audio stimuli with the visual picture, suggesting that they face challenges in recognizing/processing novel words. Moreover, these children were not sensitive to tonal mispronunciations in either the control condition or target sandhi condition, suggesting that detecting tone changes involving T2 and T3 is challenging for them. Our results therefore confirm previous and extend findings from previous studies showing children with CIs exhibit difficulties in learning new words, (Houston et al., 2005, 2012; Davidson, Geers & Nicholas, 2014), and that these children face great challenges in correctly differentiating between T2 and T3 that sharing similar pitch contours (Chen et al., 2014; Mao & Xu, 2016). Moreover, even the early implanted children (i.e., before age 2) could not detect these tonal mispronunciations. This might be related to the insufficient number of children with early implantation tested in this study, i.e., only 3 children with early implantation (before age 2) were tested. These results suggest that follow-up studies could use familiar items as stimuli and test more children with early implantation, further exploring the development of the representation of tone sandhi or phonological processes by children with CIs.

Overall, this thesis provides acoustic evidence showing that, 1) despite the pitch exaggeration and slower speaking rate of IDS and Lombard speech, contextual tones are well-realized in both registers, suggesting that children can receive good input of contextual tones; 2) typically developing 3-5-year-olds show a good understanding of the neutral tone category and the tone sandhi process, productively applying them on novel items, exhibiting adult-like tone sandhi representations; 3) children with CIs face challenges in acquiring both lexical tones and contextual tones, while early implantation benefits them in tonal acquisition. Taken together these findings suggest that the acquisition of neutral tone and tone sandhi might not be as late acquired as suggested by previous studies, i.e., after 4;6 (Li and Thompson, 1977; Hua & Dodd, 2000; Wang, 2011; Xu Rattanasone et al., 2018). Children at age 3 have already acquired the productive knowledge of the neutral tone category and the tone sandhi process, demonstrating an early acquisition of these tones. For children with CIs, developing a typical tonal system is possible, as long as they are implanted early.

Study Limitations and Future Directions

One limitation of this thesis relates to the age span of our child participants, i.e., 3 to 5 years. We did not find any developmental trend in the acquisition of the tone sandhi process of typically developing 3- to 5-year-olds, because even the youngest age group had already acquired this process. This suggests that the acquisition of tone sandhi might occur earlier than we expected. Therefore, it would be helpful for future studies to test younger children, i.e., 1- or 2-year-olds, to better capture the developmental trend in the acquisition of tone sandhi processes.

Due to the limited attention span of young children, the amount of stimuli in some studies is small, i.e., there is only 1 token per condition in Chapters 4 (neutral tone) and 5 (trisyllabic tone sandhi). This limits the generalizability of the current results to a certain degree. It thus raises further questions for future studies: can 3-5-year-olds correctly produce the tonal variation of neutral tone syllables in all types of neutral tone words? Do adults consistently show variability in their tone sandhi application on right-branching items (T3 (T3T3))? These appeal future studies to include more test items to replicate the current results.

Regarding the acquisition of contextual tones by children with CIs, a potential limitation lies in our investigation of the realization of contextual tones in their language input, i.e., Lombard speech, which is a proxy of the input these children might receive, rather than the real input they would hear. Therefore, it is still unclear how tones are realized in the real speech input from parents and teachers in their daily conversation with these children, and this appeals future studies to address this issue by testing the real input. Furthermore, future studies could also investigate the training method of teachers in their daily speaking training, i.e., whether they train these children to speak syllable by syllable to improve speech clarity. If this is the case, children with CIs would not receive enough input of connected speech, and thus might lead to the later acquisition of certain phonological processes in connected speech.

Another limitation regarding the tonal acquisition of children with CIs lies in the small number of early implanted children we have tested. These early implanted children showed greater advantages in the acquisition of contextual tones than their later implanted peers, while the small participant number (7 children) of this group suggests that future studies should test more early implanted children to confirm the results of the current results. Moreover, our results also showed that even these early implanted children are still left behind their hearing peers in both tone production and perception. This also suggests that future investigations could test children implanted within the first year of life, to future explore the effect of early implantation on their tonal development. Furthermore, given that most stimuli in the present study are novel items, posing challenges to those later implanted children in performing the task, future studies could include easier, lexicalized items, to better capture the development trend of tonal acquisition by children with CIs.

Apart from the aforementioned future directions derived from the limitation of the present study, more research could also be conducted to further explore the acquisition of Mandarin tones. First, tonal language such as Mandarin is unique in that both tones and intonation are primarily coded via pitch information. Thus an interesting question that future studies could explore is how children learn to decouple the pitch information conveying tone and intonation. Although a great number of studies have investigated how adults encode and decode lexical tones with different intonation in Mandarin Chinese (e.g., Xu, 1997), research exploring how children acquire this ability is still lacking, suggesting future studies could conduce production and perception experiments to explore this issue. Secondly, CI devices do not transmit pitch information effectively, and it has been suggested that children with CIs sometimes rely on secondary cues such as in duration and amplitude to perceive lexical tones (Tan, Dowell & Vogel, 2016). This raises the question of whether these children rely on secondary cues to perceive or represent contextual tones. If this is the case, they might develop different phonological representations for contextual tones compared to children with normal hearing. Exploring this issue will help us to better understand how auditory input shapes phonological representations, with both theoretical and clinical implications.

To conclude, this thesis brings together seven studies exploring children's acquisition of Mandarin tones in context (i.e., neutral tone and tone sandhi) by both typically developing children and children with CIs. The overall results show the good realization of contextual tones in the early language input and the early acquisition of these tones by typically developing children, suggesting that learning tones in context may not be as challenging as often assumed. The results of children with CIs also suggest that, despite the deficiency of CIs in transmitting pitch information, early implantation makes acquiring the tonal system possible.

References

Albright, A., & Hayes, B. (2011). Learning and learnability in phonology. *The Handbook of Phonological Theory* (Second Edition), 661-690.

Chen, Y., & Wong, L. L. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology*, 56(sup2), S7-S16.

Ching, T. Y., Dillon, H., Leigh, G., & Cupples, L. (2018). Learning from the Longitudinal Outcomes of Children with Hearing Impairment (LOCHI) study: Summary of 5-year findings and implications. *International Journal of Audiology*, 57(sup2), 105-111.

Colletti, V., Carner, M., Miorelli, V., Guida, M., Colletti, L., & Fiorino, F. G. (2005). Cochlear implantation at under 12 months: report on 10 patients. *The Laryngoscope*, 115(3), 445-449.

Demuth, K. 1996. The prosodic structure of early words. In J. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates. pp. 171-184.

Dettman, S. J., Pinder, D., Briggs, R. J., Dowell, R. C., & Leigh, J. R. (2007). Communication development in children who receive the cochlear implant younger than 12 months: risks versus benefits. *Ear and Hearing*, 28(2), 11S-18S.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, 21(3), 565-584.

Hua, Z. (2002). *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children*. Clevedon: Multilingual Matters Ltd.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3-42.

Kazazis, K. (1969). Possible evidence for (near-) underlying forms in the speech of a child. In *Proceedings of the Chicago Linguistic Society* (Vol. 5, pp. 382-388). Chicago: Chicago Linguistic Society.

Kerkhoff, A. O. (2007). *Acquisition of Morpho-phonology: The Dutch Voicing Alternation* (Doctoral dissertation, LOT).

Leigh, J., Dettman, S., Dowell, R., & Briggs, R. (2013). Communication development in children who receive a cochlear implant by 12 months of age. *Otology & Neurotology*, 34(3), 443-450.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43, 912-917.

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, 14(6), 654-666.

Tan, J., Dowell, R., & Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American Journal of Audiology*, 25(3), 246-256.

Van de Vijver, R. & Baer-Henney, D. (2013). On the development of the productivity of plural suffixes in German. In *The 37th Annual Boston University Conference on Language Development* (pp. 444-455).

Waltzman, S. B., & Roland, J. T. (2005). Cochlear implantation in children younger than 12 months. *Pediatrics*, 116(4), e487-e493.

Wang, C. Y. (2011). *Children's Acquisition of Tone 3 Sandhi in Mandarin*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, 7, 627.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61-83.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' Acoustic Realization of Mandarin Tone Sandhi and Lexical Tones in Context are Not Yet Fully Adult-like. *Frontiers in Psychology*, 9, 817.

Yuen, I., Demuth, K., & Johnson, M. (2011). Prosodic structure in child speech planning and production. In *Proceedings from the 17th International Congress of Phonetic Sciences* (pp. 2248-2251).

Zamuner, T. S., Kerkhoff, A. & Fikkert, J. P. M. (2006). Acquisition of voicing neutralization and alternations in Dutch. In *The 32th Annual Boston University Conference on Language Development*, Boston.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37(2), 231-247.