CRACKING THE SPEECH CODE IN TONE LANGUAGE SPEAKERS WITH AUTISM SPECTRUM DISORDERS: MECHANISMS AND TREATMENT

FEI CHEN

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Cracking the Speech Code in Tone Language Speakers with Autism Spectrum Disorders: Mechanisms and Treatment

Fei Chen

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

August, 2020

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________(Signed)

______FEI CHEN_____(Name of student)

Abstract

Atypical auditory processing has been regarded as a potential factor related to pathological speech and language processing in individuals with Autism Spectrum Disorders (ASD). Motivated by the linguistic function of pitch in tone languages (such as Mandarin and Cantonese), several recent studies with Chinese people with ASD have provided an updated perspective in this field. Yet, our full understanding of speech processing difficulties and its underlying mechanisms and treatment in tone language speakers with ASD are still far from complete. This dissertation utilized auditory and speech stimuli varying in acoustic cue (spectral or temporal), sound type (speech or nonspeech), and linguistic relevance (native vs. non-native), to reveal the nature of speech processing difficulties in tone language speakers with ASD. Moreover, the training study proposed a music-assisted approach to help improve speech sound acquisition in Mandarin-speaking children with ASD.

The first study compared the categorical perception (CP) of two prominent phonological features in Mandarin Chinese, lexical tones and voice onset time (VOT), which utilize pitch and time changes respectively to convey phonemic contrasts. Results indicated that the basic CP pattern of perceiving both native lexical tones and VOT was largely preserved in high-functioning adolescents with ASD, whereas the degree of CP of lexical tones was much higher than that of VOT. These findings suggest that the unbalanced acoustic processing capacities for pitch and time can be generalized to higher-level phonological processing in ASD. Furthermore, only in the real word condition, the Mandarin-speaking autistic individuals showed a “psychophysical boundary” similar to that observed in non-tonal language speakers who have no tonal language experience.

The second cross-linguistic study evaluated the capacity of imitating complex Cantonese tones and their non-linguistic pitch counterparts in Cantonese-speaking (native) and Mandarin-speaking (non-native) children with and without ASD. Acoustic analyses showed that both native and non-native children with ASD could generally imitate the global tone contours for three level tones and three contour tones in Cantonese, pointing to a preserved acoustic pitch imitation skill in tone-language-speaking children with ASD. However, both Mandarin-speaking and Cantonese-speaking children with ASD exhibited atypical prosodic pitch pattern with increased pitch variations relative to TD children when imitating speech tones, but no group difference when imitating nonspeech sounds. Furthermore, unlike TD children, the non-native Mandarin-speaking children with ASD failed to exploit the phonological knowledge of the familiar segment to compensate for the imitation of syllables with an unfamiliar tonal category. These findings supported the notion that lexical tone imitation was largely intact at the bottom-up acoustic pitch processing level but impaired due to a top-down phonological processing deficit in individuals with ASD.

The third training study evaluated the therapeutic potential of an adapted Melodic Intonation Therapy (MIT) in facilitating speech output for Mandarin-speaking nonverbal and low-verbal children with autism, in comparison with a matched non-MIT-based control treatment. Results indicated that our MIT-based treatment provided a more effective training approach in accelerating the rate of word and speech sound acquisition, especially lexical tone acquisition in the trained items. More importantly, the enhanced training efficacy on Mandarin tone acquisition remained at two weeks post-therapy, and generalized to novel items that were not practiced. These data provide the first empirical evidence for taking advantage of the cognitive strength of pitch processing in music, a ubiquitous nonspeech form, to compensate for the relative weakness of processing in speech sounds, especially lexical tones, in tone-language-speaking children with ASD.

Taken together, the current findings in this dissertation lend support to the notion of speech-specific lexical tone processing difficulties in ASD. Furthermore, the high-functioning tone language speakers with ASD preserved the capacity to perceive or imitate the bottom-up acoustic pitch information of lexical tones, while showed deficits in exploiting the phonological information from the carrying syllables. Thus, speech processing atypicality in ASD is presumably to be not merely domain-specific and cue-specific but also language-dependent. Finally, the efficacy of music-assisted training approach was well proved for improving the spoken language in autistic children from a tonal language background, which adds a new clinical perspective to our understanding of the close relationship between music and speech.

Acknowledgements

First and foremost, my sincere gratitude goes to my supervisor Prof. Gang PENG, who offered me effective guidance, constructive feedback, and enthusiastic encouragement during my Ph.D. study. I learned from him how to conduct novel and rational experiments in many aspects. He has gone far above and beyond the call of duty in offering tremendous help, generous support and friendship since the days I started working with him as a research assistant. My special thanks also go to Prof. William S-Y. WANG, who broadened my view of research and continuously encouraged me to do research in the area of ASD, especially among Chinese populations. Without their inspiration and dedication all along the way, this thesis is a mission impossible. I would also like to thank my committee members: BoE Chair (Dr. Angel CHAN from The Hong Kong Polytechnic University), and External Examiners (Prof. Yang ZHANG from University of Minnesota, and Prof. Suiping WANG from South China Normal University) for their valuable insights and comments.

I am grateful to Dr. Jinting YAN and Dr. Candice Chi-Hang CHEUNG for their kind cooperation with parts of the studies in this dissertation. I would like to offer my gratitude to the faculty and staff in the Research Centre for Language, Cognition, and Neuroscience, & Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University. Besides, the Penglab members have been a great support to me. I would like to thank their company and friendship that made my Ph.D. filled with fond memories. Additional thanks to many students and colleagues, Hanfei LI, Yaru MENG, Hao ZHANG, Yumeng YANG from Penglab, as well as the students from the College of Qiyue Communication, Cangzhou Normal University, for their research assistance. I would like to thank Mr. Lei LIU for his help with developing a smartphone/iPad App in the training study.

I would like to express my special thanks to all the participants and their parents/caregivers for their participation and cooperation. They are from Shenzhen Aixin Zhihui Rehabilitation Centre, Cangzhou Research Centre for Child Language Rehabilitation, Shimeitang Middle School, Tuofu Kindergarten, Jiqing Primary School, Second Foreign Language School of Nanshan, Shenzhen, and several Rehabilitation Centres, Kindergartens and Primary Schools in Hong Kong.

Finally, my gratitude to my family can never be fully expressed. I owe my deepest gratitude to my family for their unfailing love, unwavering understanding, constant support, and encouragement throughout the years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAC: Augmentative and alternative communication

ADI-R: Autism Diagnostic Interview-Revised

ADOS-2: Autism Diagnostic Observation Schedule-2

AMMT: Auditory-Motor Mapping Training

ASD: Autism spectrum disorder

CA: Chronological age

CASD: Cantonese-speaking children with ASD

CTD: Cantonese-speaking TD children

CP: Categorical perception

CT: Cantonese tone

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition

ERP: Event-related potential

F0: Fundamental frequency

GARS-2: Gilliam Autism Rating Scale–Second Edition

GLMM: Generalized linear mixed-effects model

IDS: Infant-directed speech

IPA: International Phonetic Alphabet

IQ: Intelligence quotient

IRN: Iterated rippled noise

LMM: linear mixed-effect model

M: Mean

MASD: Mandarin-speaking children with ASD

MIT: Melodic Intonation Therapy

MMF: Magnetic mismatch field

MMLI: Music-Mediated and Lexicon-Integrated

MMN: Mismatch negativity

MMR: Mismatch Responses

MT: Mandarin tone

MTD: Mandarin-speaking TD children

SD: Standard deviation

SRT: Speech Repetition Therapy

T1: Tone 1

T2: Tone 2

T3: Tone 3

T4: Tone 4

TD: Typically developing

VOT: Voice onset time

WCC: Weak Central Coherence

WISC: Wechsler Intelligence Scale for Children

Chapter 1: Introduction

1.1 Research Background

According to the latest version of the Diagnostic and Statistical Manual of Mental Disorders (DSM–5), autism spectrum disorders (ASD) represents a life-long neurodevelopmental condition affecting how one interacts with others and how they perceive the world around them, which incorporates the previously defined autistic disorder, Asperger's disorder, and pervasive developmental disorder. The individuals with ASD show deficient social communication skills as well as restricted, repetitive, and stereotyped behavior, interests, and activities (American Psychiatric Association, 2013). The delayed/atypical language development (historically linked to a defining feature) has been removed from the diagnostic criteria and is now regarded as a co-occurring characteristic of ASD. However, there is an apparent autism-specific language profile, although with high inter-individual variability. Generally, we could divide the ASD into two categories: high-functioning ASD with an intelligence quotient (IQ) higher than 70, and without severe language delays and social deficits; low-functioning ASD with an overall IQ below 70 accompanied with severe language delays and social deficits. Even for high-functioning individuals with ASD, some atypical language profiles were often observed (Walenski et al., 2006).

Individuals with ASD were reported to show difficulties in several aspects of language, such as aberrant semantic-pragmatic skills largely due to deficits that characterize ASD in the theory of mind (Baron-Cohen, 1995; Volkmar et al., 2005; Walenski et al., 2006). Moreover, children with ASD also exhibited significantly delayed onset and development of speech, with around 10-25% of them remaining nonverbal or with only little spoken language (Klinger et al., 2002; Koegel et al., 2009). The delayed speech development (Boucher, 1976; Rapin et al., 2009; Schoen et al., 2011; Shriberg et al., 2001; Wolk & Brennan, 2013; Wu et al., 2020), as well as atypical or idiosyncratic phonological processes (Cleland et al., 2010; Sheinkopf et al., 2000; Wolk et al., 2016; Wolk & Brennan, 2013; Wolk & Giesen, 2000), were often reported in children with ASD, albeit with varying degrees of severity. Furthermore, the auditory and speech processing atypicalities in ASD have been discussed in a series of comprehensive reviews (Carbajal & Malmierca, 2018; Haesen et al., 2011; Hitoglou et al., 2010; Kujala et al., 2013; Lord et al., 2020; O'connor, 2012; Ouimet et al., 2012; Schwartz et al., 2018).

To date, research in the field of speech sound development/processing in ASD has focused primarily on children's development/processing of Romance and Germanic languages (e.g., English, Spanish, German, French, etc.), which are largely non-tonal languages. For tonal languages such as Mandarin and Cantonese, speech elements contain not only consonants and vowels, but also supra-segmental lexical tones, and each syllable must be attached to one category of lexical tones to carry different lexical meanings (Wang, 1973). Such a language-specific speech feature of lexical tones offers a valuable window to investigate the acoustic vs. phonological processing skills in ASD, since the lower-order acoustic information of pitch was confounded with the higher-order linguistic meaning during the processing of the lexical tones for tone language speakers. Recently, several studies have revealed the delayed native lexical tone development (Wu et al., 2020) as well as lexical tone perception difficulties (Chen et al., 2016; Cheng et al., 2017; Wang et al., 2017; Yu et al., 2015) at both behavioral and neural levels. Yet, our full understanding of the nature of speech processing difficulties and its underlying mechanisms and treatment in tone language speakers with ASD are still far from complete.

1.2 Research Purpose

Firstly, atypical auditory processing patterns might contribute to the speech processing difficulties in autistic individuals. In the auditory modality, the reduced attention to listening to the socially relevant infant-directed speech (IDS), as well as deficits in joint attention, might produce a less sophisticated ability to acquire and process native phonemic units (Kuhl et al., 2005). Furthermore, the detail-focused processing style and enhanced low-level acoustic processing in ASD may cause individuals with ASD to focus on the within-category differences or allophonic variations (O'Riordan & Passetti, 2006; You et al., 2017). Altogether, these may bring difficulties for individuals with ASD in forming a typical categorical perception (CP) mode of native speech sounds, which requires the 'dulled' sensitivity to within-category differences while enhanced sensitivity to cross-category discrimination. Although two studies (Chen et al., 2016; Wang et al., 2017) have offered evidence of impaired CP of lexical tones in young children with ASD, huge within-group differences in the degree of CP were also observed based on the behavioral performance (Chen et al., 2016). It is still unknown whether the impaired CP pattern universally exists among all individuals of the autistic spectrum which may act as one of the biomarkers for early diagnosis of ASD, or the degraded CP performance only exists in a subgroup of ASD. In this dissertation, one of the aims was to investigate whether the high-functioning adolescents with ASD who had better language skills and longer native language experience could perceive lexical tones in a preserved CP manner, and whether the degree of CP in ASD would be indexed by chronological age, language ability as well as phonological working memory.

Furthermore, individuals with ASD showed unbalanced auditory perceptual skills towards stimulus-specific aspects of spectral vs. temporal cues, as represented by hypersensitive processing of acoustic pitch (Bonnel et al., 2003; Foxton et al., 2003; Heaton, 2005; Heaton et al., 1998; Kanner, 1943; Miller, 1989; Mottron et al., 2000; O'Riordan & Passetti, 2006; Rimland & Fein,

1988) and hyposensitive discrimination of sound duration (Brodeur et al., 2014; Falter et al., 2012; Maister & Plaisted-Grant, 2011; Martin et al., 2010; Szelag et al., 2004). It would be meaningful to investigate whether and how the differential auditory and acoustic sensitivity to spectral and temporal cues in ASD extended to the higher-level phonological processing. The formation of CP of speech sounds relies on native language experience, which reflects the higher-level phonological processing. In this dissertation, by comparing the CP competence of linguistic pitch (lexical tone) and linguistic time (VOT) in native Mandarin speakers with ASD at the same time, it can help deepen our understanding of the influence of lower-level acoustic processing on the higher-level phonological processing during speech perception with evidence from a clinical population.

In the research field of auditory and speech processing, several recent studies have reported a speech-specific and language-specific pitch processing difficulties in tonal language speakers with ASD (Jiang et al., 2015; Wang et al., 2017; Yu et al., 2015) at both behavioral and neural levels. For individuals with ASD who have tone language experience, they also exhibited an enhanced pitch processing capacity in the non-linguistic domain (e.g., music and nonspeech domains), but this advantage was conversely changed to be an aberrant perception of lexical tones or intonations in the linguistic speech domain. Such domain specificity of pitch processing was inconsistent with the findings from non-tonal language speakers, which showed a domain-general pitch processing superiority (Haesen et al., 2011; Heaton et al., 2008; Järvinen-Pasley & Heaton, 2007). So far, our full understanding of the speech-specific and language-specific lexical tone processing difficulty and its underlying mechanisms in tone language speakers with ASD are unclear. In this dissertation, the pitch contours were superimposed on various types of speech and nonspeech carriers by teasing apart the influences of stimulus complexity, social relevance, and phonemic/semantic relevance on lexical tone processing in ASD.

Secondly, compared with the auditory and speech perception, the production of speech sounds in autism has been significantly understudied. The previous studies on the production of speech prosody generally showed significantly larger pitch range and/or pitch standard deviation (SD), indicating increased pitch variations of intonation in the ASD group relative to typically developing (TD) controls (Bonneh et al., 2011; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010). The production of intonation in tone-language-speaking individuals with ASD also showed significantly higher SD of F0 than TD children, suggesting that the atypical sentence-level intonation may be a universal characteristic of individuals with ASD regardless of language background (Chan & To, 2016). However, no studies so far have investigated the syllable-level prosodic phonology in tone-language-speaking children with ASD. This dissertation also aimed to address this issue by performing pitch analyses (pitch mean, pitch range, and pitch SD) of imitative syllables produced by tone-language-speaking children with ASD, to explore whether the overall prosodic pitch pattern at the syllable level was more variable in ASD. Moreover, the complex tonal system of Cantonese (three level tones and three contour tones) showed fine-grained pitch differences regarding both pitch height and direction, imitation of Cantonese tones in both native (Cantonese-speaking) and non-native (Mandarin-speaking) autistic children offered a chance to evaluate the competence of imitating the level tones and contour tones in children with ASD, and to further illustrate how such imitative performance was influenced by language experience. Besides, in the current cross-linguistic imitation study of this dissertation, nonspeech analogues were also generated sharing exactly the same pitch trajectories with the lexical tones in speech condition, to test whether the atypical pitch imitation in ASD was speech-specific or domain-general.

Finally, increased speech output is considered a positive prognostic indicator of outcomes for low-functioning children with ASD who were nonverbal or minimally verbal (Lord et al., 2006). Recently, more and more researchers focused on music therapy in special education settings, which is regarded as a promising intervention for individuals with ASD (James et al., 2015; Reschke-Hernández, 2011). Given the behavioral resemblance between singing and speaking (Schlaug et al., 2008), as well as neural overlap in responses to speech and musical stimuli (Peretz et al., 2015), researchers have begun to examine the therapeutic effects of singing/intonation, and how it can potentially ameliorate some of the speech deficits associated with neurological disorders such as ASD (Chenausky et al., 2016, 2017; Hoelzley, 1993; Miller & Toca, 1979; Wan et al., 2010). Especially, a large body of research has implied a two-way transferability of pitch expertise across the domains of music and lexical tones (Bidelman et al., 2013; Chandrasekaran et al., 2009; Lee & Hung, 2008; Peng et al., 2013; Pfordresher & Brown, 2009; Tang et al., 2016; Wong et al., 2007). Until now, there have been no training studies targeted at facilitating lexical tone acquisition in literature for tone-language-speaking children with ASD. This dissertation also tried to integrate recent advances in speech therapy and come up with a language-specific and music-assisted training approach to facilitate speech training in tone-language-speaking children with ASD.

To conclude, although there was some neural and behavioral evidence of the impaired perception of native tonal categories in tone-language-speaking individuals with ASD, the underlying mechanisms responsible for such speech-specific lexical tone processing difficulties are still unclear. The more detailed research was needed to uncover the nature of the speech-specific pitch processing deficits, thus enabling us to come up with more effective treatment approaches. Experiment 1 in this dissertation investigated whether the CP of lexical tones and VOT was atypical in the high-functioning adolescents with ASD, as well as several possible influencing factors including chronological age, language ability, digit span, and nonword repetition. Given there was a lack of behavioral measure for lexical tone/prosody imitation or production, Experiment 2 further evaluated the capacity of imitating complex Cantonese tones and their non-linguistic pitch counterparts in Cantonese-speaking (native) and Mandarin-speaking (non-native) children with ASD. Furthermore, the training study of Experiment 3 utilized the musical elements through a smartphone/iPad App to facilitate lexical tone and other speech elements acquisition in the nonverbal and low-verbal children with ASD.

1.3 Research Significance

First, as a basic skill combing sound and meaning, speech processing has long been of interest in autism research. It has been shown that early speech processing skills can predict later language development in young children (Junge & Cutler, 2014; Kuhl et al., 2008; Newman et al., 2006; Tsao et al., 2004). Moreover, delayed language development would exert a negative influence on academic and social performance in individuals with ASD (Gillberg, 1991; Gillespie-Lynch et al., 2012). Although we have found some evidence of delayed speech development (Wu et al., 2020), and atypical lexical tone perception (Chen et al., 2016; Cheng et al., 2017; Wang et al., 2017; Yu et al., 2015) in tone-language-speaking individuals with ASD, systematic characterization of lexical tone perception and production problems is still lacking. Given that there was a correlation between speech behavior and language ability (Bartolucci et al., 1976; Wolk & Brennan, 2013; Wolk & Giesen, 2000), some of the speech processing atypicalities might be a side effect of overall language disorders in ASD, which could also be observed in some other neurodevelopmental disorders beyond autism. Thus, in this dissertation, the first two behavioral studies specially targeted at the cognitively able individuals with ASD without severe language delays, in an effort to control potential confounds in consideration of the huge heterogeneity in the ASD population (Happé & Frith, 2020). All in all, this dissertation has the potential to uncover the nature of autism-associated speech perception and production difficulties in tone language speakers with ASD in a cross-linguistic context.

Second, it has important implications for ASD assessment and remediation to examine the mechanisms leading to the speech disorders in ASD. On the one hand, it is necessary to add speech sound investigation to the assessment manual for clinical decision-making and the design of suitable therapeutic programs. On the other hand, speech disorders should be resolved in time through special spoken-language training. Cleland et al. (2010) stressed that speech sound errors might contribute to communication barriers, which supports the inclusion of speech sound components in treatment. The incidence of ASD was close to 1% worldwide (Lai et al., 2014), and we need to improve our understanding of ASD from different language backgrounds and come up with effective treatment strategies. A very recent review estimated that ASD prevalence in China was comparable to the Western world (about 1%) using standardized case identification protocol (Sun et al., 2019). However, the situation of ASD in China lags considerably behind those in the West in terms of public awareness, education opportunities, and life outcomes of autistic people (Liu et al., 2016; Yu et al., 2020). It is compelling to develop and evaluate a tone-language-specific training approach for individuals with ASD who come from a tone language background. Different from non-tonal language speakers, speech therapy in tone-language-speaking children with ASD should not only aim at enhancing consonant and vowel production, but also target at improving lexical tone production additionally. In the third study of this dissertation, a randomized controlled study was performed to evaluate the therapeutic potential of an adapted Melodic Intonation

Therapy (MIT) in facilitating speech output for Mandarin-speaking nonverbal and low-verbal children with autism. Such a language-specific speech training approach could be modified and applied to help some other tone-language-speaking children with autism beyond Mandarin-speaking ones to better acquire the phonological categories of speech sounds.

Third, just as some other research areas, ASD research has mainly been conducted on individuals from Western, educated, industrialized, rich, and democratic countries (Henrich et al., 2010), and most studies on speech development and speech processing in ASD children were conducted primarily with Romance and Germanic languages (e.g., English, Spanish, German, French, etc.) which are largely non-tonal languages. Some estimates suggest that around 60–70% of the world's languages are tonal (Yip, 2002), and more than 50% of the world's population use tonal languages (Fromkin, 1978). Languages of the world exhibit a natural diversity, which is not reflected in the mainstream empirical ASD research. More investigations need to be conducted with regard to auditory and speech processing in speakers from different language backgrounds. It is important to note that the existing cognitive theories explaining the communication disorders in individuals with ASD were mostly put forward based on ASD subjects of non-tonal language speakers and mainly from the visual modality. Actually, not all theories operate universally regardless of the language background of the ASD participants and across different modalities. A natural consequence is that the existing theories explaining communication difficulties in ASD may not be generalizable to all populations. A full understanding of the cognitive explanations of auditory and speech processing in ASD is still needed with more research in participants with ASD from different language backgrounds.

To conclude, it is important to promote diversity (linguistic, ethnic, geographical, lifespan) and intervention in translational research (convert research to training and practice) in terms of

ASD research. Related findings in this dissertation may shed light on this issue to some extent, facilitating a better understanding of this condition from a wide range of theoretical perspectives and clinical practice.

1.4 Research Questions

In this dissertation, three original studies were incorporated and reported.

Study 1 (Chapter 3: Categorical Perception of Pitch Contours and Voice Onset Time in Mandarin-Speaking Adolescents with Autism Spectrum Disorders)

Research question:

- This study focused on two prominent phonological features in Mandarin Chinese, lexical tones and voice onset time (VOT), which utilize pitch and time changes respectively to convey phonological contrasts, aiming to address three main questions: (1) Whether Mandarin-speaking high-functioning adolescents with ASD could perceive two speech continua varying in lexical tone and VOT in a similar categorical manner as neuro-typical peers, (2) Whether the performance in the CP of speech could index chronological age, language ability as well as phonological working memory in individuals with ASD, and (3) Especially, whether different types of speech and nonspeech pitch carriers with varying levels of spectro-temporal complexity or phonemic/semantic relevance could exert an impact on the perception of pitch contours.

Expected outcome:

- For high-functioning Mandarin-speaking adolescents with ASD, they may show a preserved CP pattern due to better cognitive and language skills, as well as longer exposure to the native language. If the reduced CP pattern was also observed in the adolescents with ASD, it would imply the possibility of developing the CP index as one of the biomarkers for early diagnosis of ASD from the auditory modality. In consideration of the unbalanced auditory processing capacities towards pitch and time, the degree of CP of native aspirated vs. unaspirated stops might be lower relative to that of lexical tones. The chronological age, language ability as well as phonological working memory might be correlated with the CP competence in individuals with ASD. Finally, as suggested by the speech-specific mechanism, the atypical patterns of pitch perception might be more apparent during the perception of pitch contours when embedded in speech contexts. If the pitch perception was also atypical in tone speakers with ASD when embedded into the non-speech stimuli with a comparable spectro-temporal complexity as the speech material, we could conclude that the 'Complexity Hypothesis' might be one of the influencing factors.

Study 2 (Chapter 4: Linguistic Tone and Non-Linguistic Pitch Imitation in Children with Autism Spectrum Disorders: A Cross-Linguistic Investigation)

Research question:

- This cross-linguistic study aimed to evaluate the capacity of lexical tone and non-linguistic pitch imitation in Mandarin-speaking and Cantonese-speaking children with ASD. Both native and non-native children with and without ASD were recruited to test the influence of different language experience. Moreover, the nonspeech analogues were also generated sharing exactly the same pitch trajectories with the three level tones and three contour tones in Cantonese, in an effort to test whether the performance of atypical imitation in ASD was speech-specific or domain-general. The primary research questions addressed are the following: (1) Would the syllable-level prosodic pitch produced by tone-language-speaking children with ASD show an increased pitch variation compared to TD children? (2) When imitating the complex pitch contours of Cantonese tones, would native and non-native children with ASD be able to produce normal-like lexical tone productions that are acoustically comparable to those produced by TD peers? (3) How did the top-down phonological knowledge (segmental familiarity: familiar vs. unfamiliar; tonal familiarity: familiar vs. unfamiliar) influence the lexical tone imitation accuracy in children with and without ASD?

Expected outcome:

- Compared to TD controls, tone-language-speaking children with ASD might show higher variations of the prosodic pitch productions at the syllable level, in line with previous findings in autistic children from both tonal and non-tonal language backgrounds at the sentence level. Such atypical prosodic pitch pattern may emerge when imitating speech tones, but not when imitating the nonspeech pitch contours. Similarly, the competence of imitating the complex F0 contours might be atypical or compromised in ASD when imitating lexical tones, especially the three contour tones in Cantonese. However, autistic children might show TD-matched competence in imitating the nonspeech pitch counterparts. Furthermore, for the non-native Mandarin-speaking children with ASD, they may fail to exploit the top-down phonological knowledge to compensate for the imitation of syllables with non-native tonal categories.

Study 3 (Chapter 5: Adapted Melodic Intonation Therapy Facilitates Speech Learning in Tone-Language-Speaking Children with Autism: A Randomized Controlled Study)

Research question:

- This training study evaluated the therapeutic potential of an adapted MIT in facilitating speech output for tone-language-speaking children with ASD. The intervention method was based on a smartphone/iPad app called Music-Mediated and Lexicon-Integrated (MMLI) training, which combines phonology and word learning. By using a randomized controlled trial, this study evaluated the efficacy of MMLI, an MIT-based treatment for facilitating spoken language in Mandarin-speaking nonverbal and low-verbal children with ASD, in comparison to a matched non-MIT-based control treatment, Speech Repetition Therapy (SRT). The current training study aimed to address the following questions: (1) Over the course of intensive training sessions in nonverbal and low-verbal Mandarin-speaking children with ASD, would the MMLI lead to a greater improvement in lexical tone production, word acquisition, and target sounds in initial and final positions? (2) Would the benefits of MMLI be retained after the cessation of daily training sessions and generalize to untrained novel items?

Expected outcome:

- Relative to the control treatment, there may be greater improvements for Mandarin-speaking children with ASD after receiving the intensive MMLI training, in terms of speech sound (lexical tones, initials, and finals), and word acquisition in the trained items. Moreover, such speech training superiority might be generalized to novel items that were not practiced, and may remain after the cessation of daily training sessions. Since the MMLI presented participants with additional information of pitch contours embedded in piano-timbre nonspeech, the training efficacy of MMLI might be higher in the training of lexical tones, compared with other types of speech sounds.

1.5 Structure of the Thesis

This dissertation consists of six chapters. Chapter 1 outlines the overall introduction to this thesis, including the research background, research purpose, research significance, and the research questions in the three main studies. Chapter 2 provides the closely related literature review, including the sensory abnormalities in ASD; auditory and speech processing in both tonal and non-tonal language speakers with ASD; cognitive theories to explain perceptual processing in ASD; speech development and acquisition in children with ASD; and speech therapy for children with ASD. Chapters 3–5 elaborately present the introduction, method, result, and discussion of the three studies (Studies 1–3) conducted in this dissertation. Finally, the summaries of main findings, limitations, and future directions are discussed in Chapter 6.

Chapter 2: Literature Review

2.1 Sensory Abnormalities in ASD

Sensory processing abnormalities constitute one of the core features of ASD (American Psychiatric Association, 2013), but often go unnoticed due to the communication difficulties of these patients (Posar & Visconti, 2018). As precursors to developmental milestones of ASD in social cognition, sensory symptoms may aid in diagnosis and potentially act as early diagnostic markers with the help of neuroimaging and neurobiological techniques (Robertson & Baron-Cohen, 2017). Especially, the autism-related sensory symptoms manifest changes in sensory-dedicated neural circuitry (Dinstein et al., 2012; Robertson et al., 2014), such as anatomical and neuro-molecular alterations among main sensory areas of the brain (McKavanagh et al., 2015; Puts et al., 2017; Robertson et al., 2016). These differences were commonly detected in several humans and genetic mouse models (such as GABAergic signalling) of ASD (Gogolla et al., 2009, 2014; Orefice et al., 2016; Puts et al., 2017; Robertson et al., 2016), which held great promise for translational biomarkers of the ASD condition. Very recently, Orefice (2019) creatively proposed that it is not the aberrant brain function, but the ‘peripheral’ sensory neurons—neurons outside the brain—are the key lesions leading to ASD-related phenotypes.

In visual modality, the detail-focused perceptual style and thus a visual search superiority in autism have been well proved by a series of behavioral studies (Baldassi et al., 2009; Joseph et al., 2009; Keehn et al., 2008; Kéïta et al., 2010; O’Riordan et al., 2001; Plaisted et al., 1998). One hypothesis from these observations would be that individuals with ASD might show superior detection or discrimination for the static visual stimuli (Mottron et al., 2006), whereas they have difficulties with global motion perception especially when the motion signal is weak or the time to integrate is short (Robertson et al., 2012, 2014). Furthermore, the processing of nonsocial visual stimuli tended to be intact, but the processing of social stimuli such as face processing was reported to be impaired in ASD (Celani et al., 1999; Dawson et al., 2005; Swettenham et al., 1998). In terms of the tactile perception, the conclusions were mixed as well. Some studies reported typical tactile processing (Güçlü et al., 2007), and others revealed enhanced (Blakemore et al., 2006) or impaired tactile processing (Puts et al., 2014) relative to TD controls. Also, in the auditory modality, the manifestation of hypersensitivity or hyposensitivity to auditory information in ASD is unbalanced based on the acoustic dimension (spectral or temporal) (Alcántara et al., 2012; Groen et al., 2009; Huang et al., 2018) as well as stimulus complexity (Mottron et al., 2006; Samson et al., 2006). In a short conclusion, the sensory processing capacities in individuals with ASD were not unified as shown in various modalities, but varied along a hyper- to hypo-responsivity continuum.

2.2 Auditory and Speech Processing in ASD

Atypical auditory processing patterns could contribute to the speech processing difficulties in autistic individuals, potentially causing life-long communication difficulties (Kujala et al., 2013; Kujala, 2007). The auditory and speech processing atypicalities in ASD have been discussed in a series of comprehensive reviews (Carbajal & Malmierca, 2018; Haesen et al., 2011; Hitoglou et al., 2010; Kujala et al., 2013; Lord et al., 2020; O'connor, 2012; Ouimet et al., 2012; Schwartz et al., 2018). Previous studies have found substantial evidence for atypical processing of auditory information in ASD, which was reflected by the orientation to different types of auditory stimuli, and by a wide range of auditory processing skills. As mentioned, the basic auditory processing skill in individuals with ASD has been shown as stimulus-dependent, towards the processing of spectral (pitch) vs. temporal (duration), and simple vs. complex auditory stimuli (Alcántara et al., 2012; Groen et al., 2009; Huang et al., 2018; Mottron et al., 2006; Samson et al., 2006; Yu, 2018).

Moreover, evidence for atypical auditory and speech processing in ASD has been observed using different materials of speech vs. nonspeech stimuli. Furthermore, the influence of atypical auditory processing patterns on speech sound processing might vary in individuals with ASD from different language backgrounds.

2.2.1 Orientation to Auditory Stimuli

With evidence from retrospective analyses of home videotapes, individuals who were later diagnosed with ASD showed reduced orientation to their own name being called before one year of age (Osterling et al., 2002; Werner et al., 2000) relative to age-matched TD infants and those with mental retardation. Similar observations have been reported in toddlers and older children who have been diagnosed with ASD. Compared to TD controls, children with ASD were less likely to orientate to, or were slower to orientate to social stimuli (such as child's name, snapping fingers, humming), while the between-group differences were narrowing in response to the nonsocial stimuli (such as phone ringing). Compared to nonspeech sounds, speech signals contain more social information to some extent. It seems that individuals on the autistic spectrum orientate less to the socially relevant auditory stimuli, such as the spoken language.

In line with these findings, several studies have observed reduced orientation to the socially relevant 'infant-directed speech' (IDS) in children with ASD. When addressing infants, people across cultures tend to simplify their output and use the IDS, also termed "parentese" or "motherese", with a variety of linguistic and paralinguistic modifications. Starting from the first month of life, TD infants show increased attention to and preference for hearing IDS over adult-directed speech (Cooper & Aslin, 1990; Schachner & Hannon, 2011; Werker & McLeod, 1989), which plays an important functional role in their socioemotional and language development

(Cristia, 2013; Fernald & Mazzie, 1991; Floccia et al., 2016; Golinkoff et al., 2015; Kubicek et al., 2014; Thiessen et al., 2005; Weisleder & Fernald, 2013; Zhang et al., 2011). However, for young children with ASD, they showed substantial difficulties in the realm of social communication (American Psychiatric Association, 2013; Baron-Cohen et al., 1985). There is reason to hypothesize that the typical preference for the socially relevant linguistic stimuli of IDS might be altered in this clinical population (Filipe et al., 2018). Although the socially relevant IDS is highly attractive to TD children, children with ASD as a whole group do not demonstrate a similar preference (Klin, 1991; Kuhl et al., 2005; Paul et al., 2007). For instance, comparing the percentage of head turns in the direction of the samples of IDS versus nonspeech analogues, most children in the ASD group showed an apparent preference for the nonspeech stimuli, in contrast to TD age-mates who showed a preference for IDS (Kuhl et al., 2005). Moreover, the linguistic measure of phonetic discrimination was assessed with mismatch negativity (MMN). As a whole group, children with ASD differed from controls by a) demonstrating a preference for the nonspeech analogues, and b) failing to show a significant MMN in response to a speech syllable change. When ASD group were further divided into subgroups based on auditory preference, the ASD children who preferred nonspeech stimuli also failed to show an MMN, whereas a small proportion of autistic children who preferred IDS showed typical-like MMN component. These data firmly support the hypothesis of a close association between social capacity and speech processing in individuals with ASD (Kuhl et al., 2005).

One recently published review paper (Filipe et al., 2018) suggested that such atypical responses to IDS might act as a potential early marker of risk for ASD in infants. There was ample behavioral evidence for reduced attention to IDS in children with ASD (e.g., Filipe et al., 2018; Klin, 1991; Kuhl et al., 2005; Paul et al., 2007). More importantly, the behavioral responses to IDS were related to concurrent and later language development (Droucker et al., 2013; Paul et al., 2007), and could predict subsequent communication outcomes in young children with ASD (Watson et al., 2010). The knowledge about the processing atypicality of IDS as a potential early marker of risk for ASD has developed and become more and more promising in recent years (Curtin & Vouloumanos, 2013; Droucker et al., 2013; Filipe et al., 2018; Klimesch et al., 2007; Watson et al., 2010, 2012).

2.2.2 Speech Prosody

In linguistics, prosody refers to the supra-segmental features of speech including changes in pitch (fundamental frequency, or F0), duration, stress, rhythm, loudness, and so on, which are utilized to convey various linguistic, attitudinal, emotional, pragmatic, and idiosyncratic functions (Bolinger, 1972; Cutler & Isard, 1980; Panagos & Prelock, 1997). It was noted that the speech prosody of even highly verbal individuals with ASD could be 'bizarre' (McCann & Peppé, 2003). The atypical speech prosody can be one of the major obstacles to communication competence and social acceptance of the ASD population who evidence prosodic oddities (Shriberg et al. 2001). However, relative to other aspects of auditory and speech processing, the processing of prosodic cues in ASD is an area often neglected.

Some of the earlier descriptions based on subjective impressions include monotonic or exaggerated intonation, the abnormal use of stress, and an inappropriate accent (Fay & Schuler, 1980; Kanner, 1943, 1971; Tager-Flusberg, 1981). Many of the literature on prosody in autism has investigated prosodic expression for pragmatic or affective expressions, indicating that the speech of a child with ASD is often presented by poor inflection and excessive or mis-assigned stress (C. A. M. Baltaxe & Guthrie, 1987; Fosnot & Jun, 1999; McCann & Peppé, 2003). In terms of intonation pattern, in contrast to the traditional stereotype of monotonic intonation in autism, the individuals with ASD generally showed a significantly larger pitch range and/or pitch SD compared to TD children based on acoustic analyses, indicating increased pitch variations of intonation in ASD group (Bonneh et al., 2011; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010). Specially, Chan & To (2016) investigated whether intonation atypicality could also be observed in tone-language-speaking (Cantonese-speaking) adults with high-functioning ASD when compared to the matched neurotypical controls. By analyzing the narrative samples, the acoustic analysis showed a significantly greater pitch SD for the Cantonese-speaking ASD group relative to the TD group, indicating an atypical intonation pattern at the sentence level regardless of language background.

As a test case for 'Theory of Mind' account of autism, most of the previous studies on the prosodic perception mainly focused on the perception of affective prosody by identification or discrimination of different emotional voices (Chevallier et al., 2011; Gebauer et al., 2014; Golan et al., 2007; Heikkinen et al., 2010; Martzoukou et al., 2017; Mazefsky & Oswald, 2007; Peppé et al., 2007; Rutherford et al., 2002; Scheerer et al., 2020). Furthermore, the results on the other aspects of prosodic perception revealed mixed findings. Some studies reported typical receptive prosodic abilities of ASD on the form and function tasks at the single-word level (Järvinen-Pasley, Peppé, et al., 2008), or the turn-end, chunking, and focus tasks (Paul et al., 2005; Peppé et al., 2007), while others studies showed deficits in intonation identification task in non-tone language speakers with ASD (Diehl & Paul, 2013; Hesling et al., 2010). Especially, one study (Jiang et al., 2015) compared the perceptual performance of melodic contour and speech intonation in Mandarin-speaking children with high-functioning ASD. The results revealed impaired identification and discrimination of speech intonation (statement vs. question) but superior/normal melodic contour processing in tone-language-speaking individuals with ASD, implying that tone language experience could not alleviate deficits in speech intonation perception of ASD.

2.2.3 Duration Processing in ASD

The deficits in the auditory processing of sound duration in ASD have been commonly observed with evidence from several behavioral and neuroimaging studies (Brodeur et al., 2014; Falter et al., 2012; Maister & Plaisted-Grant, 2011; Martin et al., 2010; Szelag et al., 2004). The speech sounds are cued by different acoustic features (such as pitch and duration) and different languages employ different acoustic features to form phonological contrasts. For instance, the ‘quantity languages’, such as Finnish and Japanese, make use of vowel length contrast (short vs. long phonemes) to differentiate lexical meanings. One event-related potential (ERP) study (Lepistö et al., 2005) investigated the neural responses to both acoustic and phonological duration changes in nonspeech and speech respectively in Finnish-speaking children with ASD. The findings found that Finnish-speaking children with ASD showed reduced MMNs to the phonological vowel duration contrast as well as acoustic duration change in nonspeech relative to the TD group. Another follow-up study (Lepistö et al., 2006) investigated whether a similar pattern of impaired duration processing would be shown in the Asperger syndrome without accompanying language delay. Results indicated that Finnish-speaking children with Asperger Syndrome also showed diminished MMN-amplitudes and decreased hit rates for both vowel and nonspeech duration changes. Besides, although the ASD and TD groups were not significantly different in magnetic mismatch field (MMF) power, the Japanese-speaking adults with ASD showed a delayed MMF latency in response to the duration change of a pure tone or Japanese vowel /a/ compared to TD controls (Kasai et al., 2005). To conclude, these findings collectively supported domain-general duration perception deficits for both speech and nonspeech stimuli even in high-functioning ASD and in adults with ASD (Kasai et al., 2005; Lepistö et al., 2005, 2006).

In the non-quantity languages such as Mandarin, the vowel duration change does not carry phonemic function, which means that vowel duration contrast in Mandarin reflects within-category allophonic variation. One recent study (Huang et al., 2018) investigated neural sensitivity to duration changes in speech and nonspeech contexts in Mandarin-speaking children with and without ASD, by using a passive oddball paradigm. Consistent with previous findings, the Mandarin-speaking children with ASD had diminished MMN amplitude and delayed latency in response to duration changes of pure tone when compared to the TD controls. However, no group difference was found in response to vowel duration changes in terms of MMN amplitude and latency, which suggested comparable discrimination of allophonic duration changes in children with ASD who came from a language in which duration contrast is not phonemic. Taken together, it appears that the impaired discrimination in speech condition exists in detecting phonemic duration contrast (Finnish vowels, Japanese vowels) but not nonphonemic duration contrast (Mandarin vowels), which emphasized the importance of cross-linguistic comparisons on the speech processing capacity in ASD.

2.2.4 Pitch Processing in ASD

Pitch processing has been one of the widely investigated areas regarding ASD research on auditory and speech processing. One of the prominent trends indicates that although individuals with ASD as a whole group are more proficient than TD controls at processing simple, acoustic pitch stimuli, they tend to show inferior performance when the pitch stimuli become phonologically relevant. Despite various language and cognitive impairments in ASD, some early observations found a small portion of individuals with ASD (autistic savants) showed the phenomenon of 'islands of genius', such as owning absolute or 'perfect' pitch capacity (Kanner, 1943; Miller, 1989; Young & Nettelbeck, 1995). Such pitch processing superiority in ASD has attracted researchers' interest for a long time. Several studies in this area further showed that the estimated prevalence of absolute pitch ability is much higher among musically naïve individuals with ASD relative to musically naïve TD individuals (DePape et al., 2012; Heaton, 2003; Heaton et al., 1998).

Other behavioral studies have investigated various aspects of pitch processing, such as discrimination and identification of pitch from melodic or other nonspeech stimuli. For instance, more and more research suggested that compared to TD controls, ASD as a whole group had a better pitch memory (Heaton et al., 2008; Heaton, 2003) and was generally more accurate at the processing of melodic pitch contours (Foxton et al., 2003; Heaton et al., 2001; Heaton, 2005; Mottron et al., 2000), and better at identification and discrimination of pitch changes in simple pure-tone stimuli (a nonspeech material) (Bonnel et al., 2003, 2010; O'Riordan & Passetti, 2006). Furthermore, the development of neuroimaging techniques enabled us to better understand the neurobiological signatures associated with the processing of pitch. In agreement with the behavioral findings, individuals with ASD showed larger MMN amplitudes (Ferri et al., 2003) and shorter MMN latencies (Gomot et al., 2002, 2011) in response to pure-tone pitch deviants. When the pitch information was superimposed on the spectrally and temporally complex speech stimuli, there is a trend of distinct patterns across different language speakers with ASD (i.e., tone language vs. non-tone language), indicating potentially a language-specific pitch processing pattern in individuals with ASD. The next two sections review studies regarding pitch processing in speech context from the non-tonal language speakers and tonal language speakers with ASD respectively.

2.2.4.1 Domain-General Pitch Processing Superiority in Non-Tonal Language Speakers with ASD

All the above-mentioned studies on melodic and nonspeech pitch processing superiority were all conducted in ASD subjects from the non-tonal language backgrounds. For non-tonal language speakers with ASD, the pitch variations in speech context are unrelated to phonological contrasts, which means that the pitch was perceived non-phonemically. The related perception results showed that non-tonal language speakers with ASD were generally more proficient at discriminating pitch information from both speech and nonspeech conditions relative to TD controls, pointing to a domain-general account of pitch processing superiority in nontonal-language-speaking individuals with ASD (Haesen et al., 2011; Heaton et al., 2008; Järvinen-Pasley & Heaton, 2007). For instance, one study in British English speakers with ASD (Heaton et al., 2008) explored the pitch discrimination skills in stimuli of real word, nonword, and nonspeech pitch contour analogues. The findings revealed superior performance in ASD across all the three conditions, indicating enhanced discrimination of pitch information across both speech and nonspeech auditory materials. Another study (Järvinen-Pasley & Heaton, 2007) asked English-speaking children with ASD and age-matched controls to perform a pitch sequence discrimination task (same/different judgments) with pitch embedded in music and speech stimuli. There was no difference in pitch sensitivity between music and speech conditions in the ASD group, while TD controls exhibited significantly better performance in the music condition, which offered further evidence for reduced domain-specificity in auditory pitch processing in nontonal-language-speaking individuals with ASD. Even when the longer pitch contours were embedded in spoken sentences, the enhanced pitch discrimination skill in speech was also detected (Järvinen-Pasley et al., 2008; Järvinen-Pasley & Heaton, 2007). Furthermore, the larger MMN amplitudes to pitch changes embedded in speech stimuli have also been found in nontonal-language-speaking children and adults with ASD compared to TD controls (Kujala et al., 2010; Lepistö et al., 2005, 2006). All the corroborating evidence pointed to the domain-transferred pitch processing superiority from nonspeech domain to speech domain in ASD subjects from a non-tonal language background.

2.2.4.2 Domain-Specific Pitch Processing Pattern in Tone Language Speakers with ASD

For tone language speakers, pitch processing of lexical tones belongs to one part of phonological processing, which happens at the super-segmental level. It is worth investigating whether the enhancement in pitch processing can generalize to tone language speakers with ASD. Several studies have demonstrated that long-term access to a tonal language environment may shape the neural architectures of pitch processing in native speakers (Gandour et al., 2004; Gu et al., 2013). The specific characteristics of the Chinese language (e.g., lexical tones in Mandarin and Cantonese) can offer us a valuable chance to recheck the obtained findings based on the non-tonal languages in a cross-linguistic framework.

Investigations on speech and/or nonspeech pitch processing in tone language speakers with ASD started very recently from 2015 (Chen et al., 2016; Cheng et al., 2017; Wang et al., 2017; Yu et al., 2015). In terms of lexical tone processing at the syllable level, three studies focused on Mandarin tone perception in Mandarin-speaking children with ASD (Chen et al., 2016; Wang et al., 2017; Yu et al., 2015) and one study on Cantonese tone perception in Cantonese-speaking adults with ASD (Cheng et al., 2017). The core research focus was whether tone language speakers with ASD also exhibited the advantage of superior pitch perception in both lexical tone and nonspeech pitch analogues. Among which, Yu et al. (2015) used the MMN paradigm and explored the pitch perception ability in Mandarin-speaking children with ASD (6–12 years). The pitch contours of two Mandarin typical tone categories (Tone 2 vs. Tone 4) were embedded in different types of speech and nonspeech carriers. The ERP data showed an enhanced neural sensitivity to pitch changes in ASD under two nonspeech conditions (pure tone and hum), but reduced sensitivity to lexical tone changes in speech stimuli (real word and nonword), which pointed to speech-specific lexical tone processing difficulties in tone language speakers with ASD.

To further investigate a more fine-grained perception of pitch contours in Mandarin-speaking children with ASD, both one behavioral study (Chen et al., 2016) and one ERP study (Wang et al., 2017) utilized the CP paradigm. Using a classic paradigm of behavioral CP, Chen et al. (2016) investigated the identification and discrimination of Mandarin tones (Tone 1 vs. Tone 2) in six- to eight-year-old children with ASD, and compared it with age-matched TD children. In stark contrast to the TD controls, the Mandarin-speaking children with ASD exhibited no enhanced discrimination accuracy for the between-category pairs and showed a wider perceptual width around the boundary. Moreover, the identification slope of lexical tones in ASD was strongly correlated with the developmental age of language ability. The behavioral results showed an impaired lexical tone processing in Mandarin-speaking young children with ASD, especially among those with severe language delay or impairment. The notion of impaired CP of lexical tones in speech condition was further supported by another ERP study (Wang et al., 2017), which examined the Mandarin-speaking children's (9–13 years of age) Mismatch Responses (MMR) to within-category and between-category pitch deviants in both speech and nonspeech carriers. The results of both MMR data and intertrial phase coherence offered the electrophysiological evidence for lack of CP of lexical tones in speech condition, whereas a typical-like CP pattern in the nonspeech condition (harmonic sound).

The perception of Cantonese tones in ASD was investigated by (Cheng et al., 2017), with the Cantonese tone system being more complex relative to the Mandarin tones. In the study (Cheng et al., 2017), Cantonese-speaking adults with ASD and neurotypical individuals were asked to discriminate the pairs of real syllables, pseudo-syllable, and nonspeech stimuli (with high-frequency components removed) contrasting pitch levels. Perceptual results showed, when with one semitone difference, the higher discrimination ability was found for the level-pitched nonspeech stimuli compared with the pseudo-syllables. Moreover, further analysis revealed that increased pitch perception skill in ASD was found in a subgroup without speech delay. To conclude, irrespective of the tone inventory (Mandarin or Cantonese), the enhanced or at least preserved non-linguistic pitch perception skill in ASD generalized to those from a tone language background. Nevertheless, such an enhanced acoustic pitch perception skill was conversely changed to the compromised perception of lexical tones when the pitch changes are phonologically relevant for the tone language speakers with ASD.

2.3 Speech Development and Acquisition in Children with ASD

Some earlier conducted studies indicated that, unlike other speech and language behaviors, articulatory skills were relatively intact in children with ASD (Bartak et al., 1975; Kjelgaard & Tager-Flusberg, 2001; Rapin & Dunn, 2003). For instance, Kjelgaard and Tager-Flusberg (2001) tested the phonological, lexical, and higher-order semantic and grammatical skills of 89 ASD children, and found that they had relatively intact articulation, but their non-articulatory abilities showed great heterogeneity. This conclusion may have underestimated the extent to which these children experienced difficulties in speech sound production as single-word articulation was judged as simply either correct or incorrect. This binary classification may overlook many atypical production patterns presented by individuals with ASD. A good score does not necessarily indicate normal speech production. A detailed analysis of these error patterns is necessary. More comprehensive analyses of phonological capacities have reported articulatory impairments in individuals with ASD (Boucher, 1976; Shriberg et al., 2001). Speech sound impairments might involve problems regarding the place and manner of articulation and include specific difficulties in maintaining the syllabic structure of words. In examining articulation distortion errors in a structured conversational interaction, Shriberg et al. (2001) found a high prevalence of vocalization problems among 30 male participants diagnosed with ASD. In another study (Rapin et al., 2009), around 25% of 62 school-age ASD children showed significant speech production disorders in a cluster analysis of speech sound errors.

Recently, research methods of investigating speech sound characteristics of children with ASD have become more comprehensive, which have examined speech production patterns in greater details (Cleland et al., 2010; Eigsti et al., 2011; McCleery et al., 2006; Schoen et al., 2011; Wolk & Brennan, 2013; Wu et al., 2020). McCleery et al. (2006) investigated the consonant production of 14 ASD children with severe language delays (2;1–6;11) via an imitation task. In comparison with speech productions of ten TD children (1;1–1;2), similar production patterns were found between TD children and nonverbal or minimally verbal ASD peers. Wolk and Brennan (2013) documented the phonetic inventories, typical vs. atypical error patterns, and the correlation between language delay and the number of error patterns of eight children with ASD (5;3–15;1). They asked the child participants to spontaneously produce speech in an object-naming task. The results showed that all the ASD participants exhibited certain typical error patterns (e.g., fronting, stopping, and gliding) reflecting delayed language development, while some also showed atypical error patterns (e.g., deaffrication, migration, and palatalization). Wolk and Giesen (2000) observed a "chronological mismatch" in ASD children's speech sound development such that early-developing sounds (e.g., /s/ and /ʃ/) were absent while later-developing sounds (e.g., /z/ and /ʒ/) were present. Schoen et al. (2011) found that significantly more atypical nonspeech vocalizations (e.g., high-pitched squeals) occurred in the ASD group than among age- and language-matched controls. It thus appears that the major difference separating the vocal development of the ASD population and TD individuals is the amount of atypical nonspeech vocalizations being produced. Furthermore, one recently conducted study investigated speech acquisition and development in Mandarin-speaking children with ASD (Wu et al., 2020). The judgement scores on Mandarin initials, finals, and lexical tones in Mandarin-speaking children with ASD (3–6 years) were significantly lower relative to age-matched TD children, whereas no score differences were found between Mandarin-speaking children with ASD (mean age = 10.82 years) and language-matched TD children (mean age = 4.20 years). What's more, Mandarin-speaking children with ASD showed atypical development sequences in both Mandarin initials and finals.

To conclude, while some children with ASD showed a level of phonological development close to that of TD children (Bartak et al., 1975; Kjelgaard & Tager-Flusberg, 2001; Rapin & Dunn, 2003), others showed varying degrees of delayed speech sound development (Boucher, 1976; Rapin et al., 2009; Schoen et al., 2011; Shriberg et al., 2001; Wolk & Brennan, 2013; Wu et al., 2020) and exhibited atypical phonological processes (Cleland et al., 2010; Sheinkopf et al., 2000; Wolk et al., 2016; Wolk & Brennan, 2013; Wolk & Giesen, 2000; Wu et al., 2020). Several factors may account for the inconsistent findings. First, the different methods used to elicit speech samples may have led to different results. One commonly used method is to collect spontaneous speech via picture/object-naming tasks or recording daily communications between children and parents or clinicians. However, it could be difficult to use the same methods to elicit spontaneous speech in ASD children, especially younger ones, since the lack of social motivation can curb speech output (Schoen et al., 2011). Another method is imitation, which may overestimate ASD participants' actual production abilities since speech models are provided for them (Wolk et al., 2016). This method is nevertheless more suitable for children with ASD who lack the desire to communicate. Second, the ASD group shows great heterogeneity due to individual differences. The importance of subgrouping children with ASD should be highlighted for distinct subtypes. Some may have language skills within the normal range, while others present similar language impairments to those reported for children with special language impairment (Kjelgaard & Tager-Flusberg, 2001). It is important to note that there was a strong correlation between speech development and the degree of overall language development in ASD as reported in various studies (Bartolucci et al., 1976; Schoen et al., 2011; Wolk & Brennan, 2013; Wolk & Giesen, 2000).

2.4 Speech Therapy for Children with ASD

Timely Intervention and support for ASD should be individually tailed and, if appropriate, multimodal, and multidisciplinary. There are various behavioral approaches mainly targeted at improving a broad range of skills (cognitive, language, sensorimotor, and adaptive behaviors) for individuals with ASD, including, amongst others, the widely used Applied Behavioral Analysis (Peters-Scheffer et al., 2011; Smith & Eikeseth, 2011), the Treatment and Education of Autistic and related Communication-handicapped Children (Callahan et al., 2010; Virues-Ortega et al., 2013), peer- and parent-mediated programs (Chang & Locke, 2016; Stadnick et al., 2015), augmentative and alternative communication (AAC) systems (Beukelman & Mirenda, 1998; Flippin et al., 2010), computer-based or other technology-aided (such as virtual reality) instruction (Bosseler & Massaro, 2003; Mesa-Gresa et al., 2018), as well as such as music- or animal-assisted therapy (O'Haire, 2013; Reschke-Hernández, 2011). Specifically, delayed speech and language development constitute one of the earliest symptoms of many young children who are later diagnosed with ASD (Tager-Flusberg & Finch, 2020). However, there have been limited empirical studies on intervention methods specifically aiming at improving speech acquisition in autistic individuals so far.

2.4.1 Speech Training Methods for Nonverbal or Minimally Verbal Children with ASD

Given the autistic features especially among low-functioning individuals who begin treatment with limited or even no spoken words, most of the speech training approach in nonverbal or minimally verbal children with ASD often utilized orienting cues or motor activities to attract attention (Koegel et al., 2009; Paul et al., 2013; Rogers et al., 2006; Tsiouri & Greer, 2003), added AAC modes of communication (Romski et al., 2010; Schlosser & Wendt, 2008; Sulzer-Azaroff et al., 2009; Tager-Flusberg & Kasari, 2013), and adopted the computer-assisted or smartphone/iPad-based pronunciation training methods (Chen et al., 2019; King et al., 2014; Tager-Flusberg & Kasari, 2013). For instance, one recent study (Chen et al., 2019) developed and evaluated a computer-assisted pronunciation tutor for Mandarin-speaking preschoolers with ASD. The findings indicated that low-functioning children with ASD who are struggling with speech sound acquisition could benefit more from the 3-D virtual pronunciation tutor, compared to the real human face tutor. By demonstrating additional visual information of internal articulators and airflow changes, the 3-D virtual pronunciation tutor attracted their attention, and provided an efficient pronunciation training approach to enhance consonant and vowel production skills among the Mandarin-speaking children with ASD. However, the visual speech presented in the virtual tutor did not incorporate the lexical tone information; other speech training approaches should be developed to help tone-language-speaking children with ASD alleviate their difficulties in lexical tone acquisition.

2.4.2 Music-Assisted Speech Training

In one literature review (James et al., 2015), the authors pointed to an emerging evidence base on the effects of music therapy for individuals with ASD. There is conclusive evidence reporting positive outcomes to classify music therapy as a promising intervention for individuals with ASD (James et al., 2015; Reschke-Hernández, 2011). Although music and speech belong to different domains with different representations, an increasing number of neuroimaging studies point to a large neural overlap in responses to speech and musical stimuli (see Peretz et al. 2015 for a review), which implies a close relationship between musicality and speech-processing capacity. Researchers have begun to examine the therapeutic effects of Melodic Intonation Therapy (MIT; Albert et al., 1973; Sparks et al., 1974), initially designed for improving spoken language in left-hemisphere stroke patients with severe non-fluent aphasia, on speech therapy in ASD. The MIT approach combines the singing/intonation with motor activities such as finger tapping/clapping, and such "music and movement" combination has been regarded as a powerful clinical tool for ASD (Srinivasan & Bhat, 2013). To conclude, the music-assisted speech therapy can potentially ameliorate some of the speech deficits associated with neurological disorders such as ASD (Chenausky et al., 2016; Chenausky et al., 2017; Hoelzley, 1993; Miller & Toca, 1979; Sandiford et al., 2013; Wan et al., 2010).

Furthermore, many studies have implied a two-way transferability of pitch expertise across the domains of music and lexical tones (Bidelman et al., 2013; Chandrasekaran et al., 2009; Lee & Hung, 2008; Peng et al., 2013; Pfordresher & Brown, 2009; Tang et al., 2016; Wong et al., 2007). For instance, on the one hand, a tone language background is often associated with better performance in musical pitch processing (Bidelman et al., 2013; Peng et al., 2013; Pfordresher & Brown, 2009). On the other hand, musicians are likely to be more adept at detecting non-native lexical tone changes compared with non-musicians (Chandrasekaran et al., 2009; Lee & Hung, 2008; Wong et al., 2007) and even more sensitive to native lexical tone categories (Tang et al., 2016). Since both music and lexical tones share the same psycho-acoustical attribute of pitch, and most children with ASD showed an intrinsic interest in musical notes, and exhibited a nonspeech pitch processing superiority, it would be reasonable to make use of music therapy for tone-language-speaking children with ASD. However, until now, there have been no training studies in literature, especially targeted at utilizing music-assisted therapy to facilitate lexical tone acquisition in ASD.

2.4.3 Auditory-Motor Mapping Training

The multi-modal training methodology, called Auditory-Motor Mapping Training (AMMT), was initially proposed by Wan et al. (2011) to facilitate speech output for English-speaking nonverbal children with autism. After therapy, all the nonverbal children with ASD started to articulate some word approximations and phrases. Later, the higher efficacy of AMMT has been further observed when compared with a non-AMMT-based treatment in minimally verbal children with ASD (Chenausky et al., 2016) and one more-verbal child with ASD (Chenausky et al., 2017).

There are several neurophysiological mechanisms underlying the efficiency of auditory-motor mapping training in facilitating speech output. First, the arcuate fasciculus (AF), a fiber bundle that connects the auditory perceptual regions in the temporal lobe with the motor-related regions in frontal lobe (Catani et al., 2005), could be developed or reconstructed through auditory-motor mapping activities (Wan et al., 2010, 2011). The AF might be responsible for the bidirectional mapping between speech articulation and acoustics (Leclercq et al., 2010), as well as facilitating new word learning especially in the left bundle (López-Barroso et al., 2013). Second, another neural substrate likely to be engaged during auditory-motor mapping is the putative mirror neuron system (MNS). It has been suggested that dysfunctional MNS underlies some of the speech and language deficits in individuals with ASD (Iacoboni & Dapretto, 2006). The AMMT may activate brain regions that overlap with MNS, thus highlighting the potential benefits of such sensorimotor training to facilitate expressive language in developmental disorders such as autism (Overy & Molnar-Szakacs, 2009). Third, it has been suggested that auditory-motor link may act as a unique manner for engaging the inferior frontal gyrus in two of its potential functions—the mapping of sounds to articulatory actions and their sequential execution (Wan et al., 2010, 2011), thus providing its potential facilitative effects on speech perception and production.

2.5 Theories of Altered Sensory and Cognitive Functioning in ASD

There were three primary cognitive theories to explain the atypical perceptual performance in ASD: a) 'Weak Central Coherence' (WCC) theory which indicated a reduced trend to process information into a 'global' whole, while showing an increased focus on the detail (Frith, 1989), b) 'Social Theory' which pointed out an impairment in social cognition and social perception (O'connor, 2012), and c) 'Complexity Hypothesis' which showed a perceptual deficit in the more complex stimuli (Bertone et al., 2005).

Among the first category, the WCC theory (Frith, 1989, 2003) emphasized that the enhanced perception of detail may be accompanied by reduced attention to global. This theory indicated that perceivers with ASD might show a cognitive style focusing on details when processing information and show difficulty in changing the absolute attention from the local area to the global context (local vs. global). The enhanced pitch processing skill in the musical and nonspeech stimuli for both tone language and non-tone language individuals with ASD can be well explained by the WCC theory, with a well-proved acoustic pitch processing skill of local superiority in ASD (Happé & Frith, 2006). Consistent with WCC theory, an alternative, more moderate view of local vs. global processing in ASD, was suggested by the 'Enhanced Perceptual Functioning' theory (Mottron & Burack, 2001), which also supports enhanced local processing in ASD, similar to the WCC theory, but does not necessarily expect a global processing deficiency especially when the contextual information was relatively simple to process.

Second, the 'Social Theory' was based on the idea that individuals with ASD showed a deficit in understanding human emotions and were unable to detect the mental states (such as beliefs, intentions) in other persons. They were less attracted by social information processing in both auditory and visual modalities (Baron-Cohen, 1989; Dawson et al., 2005; Kuhl et al., 2005; Rutherford et al., 2002). Specifically, one important study (Kuhl et al., 2005) has examined the speech and social processing in individuals with ASD. Results indicated that children with ASD showed a reduced preference for the IDS, and did not exhibit a significant MMR component when reacting to the changes of speech stimuli during the pre-attentive processing stage. Thus, the socially relevant child-directed speech signals contain social information to some extent. Orientation to speech plays a vital role in processing and comprehending oral language and social communication (Kuhl et al., 2005). It is likely that individuals with ASD, who lack social interest, may show compromised pitch processing when embedded in the speech environment.

Third, the role of stimulus complexity is often not explicitly considered while investigating sensory processing, but could exert a potential influence on research outcomes. The review paper by Minshew & Goldstein (1998) suggested ASD as a disorder of complex information processing. According to their theory, the challenges individuals with ASD experience are not dependent solely on the social nature of the stimuli, and the local/global level of processing, but preferably on the complexity level of the information processed. One similar theory, specifically used to explain auditory processing in ASD, is the ‘Neural Complexity Hypothesis’ (Samson et al., 2006), which attributed the auditory processing atypicality in ASD to altered neural hierarchy. This theory proposes that individuals with ASD tend to perform better with spectro-temporally simple sounds relative to TD controls, but have difficulties in processing spectrally and temporally complex auditory information. The abnormal auditory and speech processing in autism are often intertwined with stimulus complexity. Thus far, the impact of stimulus complexity on auditory processing in ASD has not been systematically investigated. The speech stimuli are more complex in terms of the spectro-temporal components, compared with different types of nonspeech stimuli (e.g., pure tones, harmonics, hum, or filtered sounds) used in previous studies. Thus, as suggested by the ‘Neural Complexity Hypothesis’, the pitch processing skill of nonspeech stimuli (simple stimuli) might be preserved or even enhanced in ASD, while the pitch processing ability of speech stimuli (complex stimuli) might be reduced to a great extent.

Chapter 3: Categorical Perception of Pitch Contours and Voice Onset Time in Mandarin-Speaking Adolescents with Autism Spectrum Disorders

3.1 Introduction

The speech and language difficulties are central to ASD, and are one of the critical signs and symptoms for establishing a diagnosis with ASD (Landa, 2008; You et al., 2017). However, the mechanisms that underlie speech and language delay/impairment in ASD remain poorly understood. Atypical perceptual processing patterns might contribute to the speech processing difficulties in autistic individuals, who have a unique way to perceive the surrounding world. In the auditory modality, it was widely reported that TD children tended to show increased attention to and preference for hearing the socially relevant infant-directed speech sounds (Cooper & Aslin, 1990; Schachner & Hannon, 2011; Werker & McLeod, 1989). However, children with ASD as a whole group did not demonstrate a similar preference (Klin, 1991; Kuhl et al., 2005; Paul et al., 2007), and instead, most of them showed a preference for the corresponding nonspeech/non-linguistic analogs. Importantly, when dividing the autistic individuals into different subgroups who preferred speech or nonspeech (Kuhl et al., 2005), the subgroup who showed an apparent preference for the nonspeech stimuli (20 out of 27) did not exhibit a significant mismatch responses (MMRs) when reacting to the changes of speech stimuli (/ba/ vs. /wa/), while the minority of them (7 out of 27) who showed a preference for IDS did. Thus, the performance of speech processing was closely related to social capacity in children with ASD (Constantino et al., 2004, 2007). Since the early acquisition of speech phonemes greatly depended on a socially interacted environment (Kuhl et al., 2003), there is reason to hypothesize that in autism, the lack of motivation to engage in social learning and reduced attention to listening to speech, might produce a less sophisticated ability to process native phonemic units.

Previous studies on the development of speech perception in TD infants showed a language-dependent perceptual reorganization between 6 and 12 months, by perceptual narrowing for the native phonological contrasts and by perceptually ‘tuning out’ irrelevant acoustic information (Best & McRoberts, 2003; Kuhl, 2000, 2004). Such Perceptual Magnet Effect (Kuhl, 1991; Kuhl et al., 2008) around prototypes laid a foundation for the categorical perception (CP) mode of native phonological categories in TD children, with a much higher sensitivity to the auditory stimuli across the category boundary than of equivalently separated stimuli within the same phonetic category (Liberman et al., 1957). The early social deficits in ASD might lead to poorer specialization and categorization for the native speech sounds. As mentioned above, the perceptual reorganization process may be atypical in children with ASD, and they might show compromised performance in discriminating different speech categories in native language compared to TD children. An accumulating of neural evidence showed that MMRs to the changes of native vowel contrast (Čeponienė et al., 2003; Lepistö et al., 2006, 2008), consonant contrast (Jansson-Verkasalo et al., 2003; Kuhl et al., 2005; Kujala et al., 2010), or lexical tone contrast (Yu et al., 2015) tended to be weaker in children with ASD than in TD children. On the other hand, as suggested by the ‘Weak Central Coherence’ theory (Frith, 1989; Happé & Frith, 2006) and enhanced perceptual functioning in autism (Mottron & Burack, 2001), such detail-focused processing style and enhanced low-level acoustic processing may cause autistic individuals to focus on the intrinsic acoustic differences between speech sounds of the same category (M. O’Riordan & Passetti, 2006; You et al., 2017). That is to say, individuals with ASD were likely to outperform TD controls in detecting within-category acoustic changes following the ‘allophonic perception’ theory for autism (Huang et al., 2018; You et al., 2017). Altogether, these may bring difficulties for young children with ASD in forming a typical CP pattern, which requires the ‘dulled’ sensitivity to within-category differences as well as enhanced between-category discrimination. Thus, we hypothesized that individuals with ASD might show an impaired CP pattern, or at least a reduced degree of CP of different native phonemic units compared to age-matched TD controls.

Furthermore, at the basic acoustic processing level, individuals with ASD showed unbalanced auditory perceptual skills towards different aspects of spectral vs. temporal cues, as represented by a hypersensitive processing of pitch and hyposensitive discrimination of sound duration (Haesen et al., 2011). Specifically, clinical observations have reported that a small portion of ASD owned ‘islands of genius’, such as case descriptions of musical savants with autism owning absolute or ‘perfect’ pitch (Heaton et al., 1998; Kanner, 1943; Miller, 1989; Rimland & Fein, 1988). In addition, more and more research suggested that compared to TD ones, ASD as a whole group had a better pitch memory (Heaton et al., 2008; Heaton, 2003) and was generally more accurate at the processing of melodic pitch contours (Foxton et al., 2003; Heaton et al., 2001; Heaton, 2005; Mottron et al., 2000), and better at identification and discrimination of pitch changes in pure-tone stimuli (a nonspeech material) (Bonnel et al., 2003; O’Riordan & Passetti, 2006). In stark contrast, autistic individuals tended to show poorer performance in the basic auditory processing of sound duration from the evidence of both behavioral and neuroimaging studies (Brodeur et al., 2014; Falter et al., 2012; Maister & Plaisted-Grant, 2011; Martin et al., 2010; Szelag et al., 2004), indicating that timing impairments may underpin core features of ASD. As suggested by the functional hypothesis (Lancker, 1980), the accurate and complete perception of speech sound in native language speakers involves the processing of acoustic information as well as the phonological information. It would be meaningful to investigate whether and how the differential auditory and acoustic sensitivity to spectral and temporal cues in ASD extended to the higher-level phonological processing.

Besides the supra-segmental lexical tones, Mandarin phonology is also well-known for its varieties of aspirated vs. unaspirated consonants. The lexical tone uses the spectral cue of fundamental frequency (F0) to differentiate lexical meanings. For example, the Mandarin syllable *ba*/pa/ (*Pinyin*, an alphabetic phonological coding system used in Mainland China, and the corresponding International Phonetic Alphabet enclosed by backslashes) with high-level pitch (Tone 1) means "eight", and the same syllable means "to pull" when it is pronounced with high-rising pitch contour (Tone 2). Furthermore, the distinctive feature of unaspirated vs. aspirated contrast was acoustically realized as the temporal cue of voice onset time (VOT, defined as the time interval between the beginning of release burst and the onset of glottal pulsing) in Mandarin stops (Abramson & Whalen, 2017; Lisker & Abramson, 1964). For instance, the aspirated Mandarin stop *p*/p^h/ carries a much longer VOT compared to the corresponding unaspirated *b*/p/. A series of behavioral and neurophysiological studies have proved that Mandarin-speaking TD adults and children perceived both the linguistic pitch (lexical tone: Chen et al., 2017; Wang, 1976; Xi et al., 2010; Xu et al., 2006) and linguistic time (VOT; Cheung et al., 2009; Feng, 2018; Xi et al., 2009) in a highly categorical manner, with a sharp identification boundary and enhanced sensitivity to between-category contrasts relative to within-category ones. Moreover, developmental studies showed that Mandarin-speaking TD children from six-year-olds started to show an adult-like competence in the CP of tones (Chen et al., 2017; Xi et al., 2009), and 10-year-old children generally reached an adult-like CP of VOT (Feng, 2018). Two studies investigated the CP of lexical tones in Mandarin-speaking children with ASD (Chen et al., 2016; Wang et al., 2017), but none of the previous studies have simultaneously explored the processing of lexical tones and VOT in ASD within one single study. To fill the research gap, in the current study, we compared the competence of CP of lexical tones and VOT in Mandarin-speaking individuals with

ASD, in an effort to investigate how different auditory weighting system towards spectral and temporal cues in ASD influenced their linguistic processing of native phonological categories of lexical tones (Tone 1 vs. Tone 2) and VOT (*b*/p/ vs. *p*/p^{h}/) as indexed by CP measures. Answers to this question would deepen our understanding of the influence of lower-level acoustic processing on the higher-level phonological processing during speech perception with evidence from clinical populations.

In the existing literature, one ERP study (Wang et al., 2017) investigated the neural responses to the equivalent pitch deviations representing within-category and between-category differences in speech (lexical tone) and nonspeech (harmonic sound) conditions in Mandarin-speaking children with ASD. The MMRs and neural oscillatory activities showed that, in speech condition, the TD children showed typical CP of lexical tones with enhanced neural sensitivity to the between-category deviant relative to the within-category one as expected, whereas children with ASD showed a lack of CP of lexical tones with equivalent neural responses to the two types of deviants. In nonspeech condition, however, both ASD and TD groups showed a preserved CP pattern with cross-boundary benefits, pointing to a speech-specific CP deficit in autism in the pre-attentive neural processing stage. Another behavioral study (Chen et al., 2016) also suggested that the low-verbal children with ASD exhibited no enhanced discrimination accuracy for the between-category pairs, and showed a much wider perceptual width around the boundary relative to age-matched TD children, pointing to an impaired CP of lexical tones. Moreover, a strong correlation was found between the boundary width and developmental age of language ability among the child participants with ASD, with some autistic participants with higher developmental age showing a much sharper identification curve even close to that in TD controls. Although both studies (Chen et al., 2016; Wang et al., 2017) have offered empirical evidence of impaired CP of lexical tones in autistic children as a whole group, a huge within-group heterogeneity in the degree of CP was also observed based on the behavioral performance (Chen et al., 2016). It remained unclear whether the CP of lexical tones was universally impaired among all the native individuals of the autistic spectrum regardless of age and language/cognitive capacity. If this were the case, it would imply the possibility of developing the CP index as one of the biomarkers for early diagnosis of ASD from the auditory modality. Alternatively, if the degree of CP was altered among different subgroups of ASD and different ages, it would be necessary to uncover the possible influencing factors. The degree of CP of native phonemes in TD children increased with age, due to an accumulation of perceptual development from the tonal information of ambient sound input (Chen et al., 2017). Also, some sub-tests of phonological working memory, such as digit span and nonword repetition, were considered to contribute to the behavioral performance of speech perception (Millman & Mattys, 2017). In this study, we also aimed to investigate whether the high-functioning adolescents with ASD who had longer native language experience could perceive native speech sounds in a preserved CP manner, and whether the performance in CP of speech would be indexed by chronological age, language ability as well as phonological working memory in ASD.

In the research field of auditory and speech processing, several recent studies have reported a speech-specific pitch processing atypicality in tonal language speakers with ASD (Jiang et al., 2015; Wang et al., 2017; Yu et al., 2015). In terms of syllable-level pitch processing, Mandarin-speaking children with ASD showed an atypical or impaired processing of lexical tones (Wang et al., 2017; Yu et al., 2015), whereas they showed normal or even enhanced processing of the same pitch information in the nonspeech materials (pure tone or harmonic sound). Such domain specificity of pitch processing was inconsistent with the relevant findings based on non-tonal language speakers. For instance, one study in British English speakers with ASD (Heaton et al., 2008) explored the pitch discrimination skills in stimuli of real word, nonword, and nonspeech conditions. The perceptual results showed that English-speaking children with ASD were generally more proficient at discriminating pitch contours from both speech (real word and nonword) and nonspeech conditions relative to TD controls, pointing to a domain-general account of pitch processing superiority in non-tonal language speakers with ASD (Haesen et al., 2011; Heaton et al., 2008; Järvinen-Pasley & Heaton, 2007). The discrepancy between different language backgrounds suggested that the speech-specific lexical tone processing difficulties in autism were likely to be related to the unique phonological role of lexical tones. Given that lexical tones are superimposed on the syllabic segments, the semantic status of the pitch carriers in speech (real word vs. nonword) might contribute to the performance in CP of lexical tones. Furthermore, as suggested by the Neural Complexity Hypothesis (Samson et al., 2006) in depicting auditory processing in ASD, which proposes that individuals with ASD have difficulties in processing spectrally and temporally complex auditory information. The speech stimuli are more complex regarding the spectro-temporal components, compared with different types of nonspeech stimuli (e.g., pure tones, harmonics, or filtered sounds) adopted in previous studies. Thus, in order to test whether the acoustic complexity also plays a role, it is necessary to introduce another type of nonspeech material in this study, such as iterated rippled noise (IRN) which is comparable to the speech materials in terms of the spectro-temporal complexity (see *Methods* for more details). All in all, this study tried to uncover the nature of speech-specific lexical tone processing difficulties in tone-language-speaking individuals with ASD, by comparing the CP of pitch contours embedded in various types of pitch carriers with varying levels of spectro-temporal complexity or phonemic/semantic relevance.

To conclude, we focus on two prominent phonological features in Mandarin Chinese, lexical tones and voice onset time (VOT), which utilize pitch and time changes respectively to convey phonological contrasts, aiming to address three main questions: (1) Whether Mandarin-speaking high-functioning adolescents with ASD could perceive two speech continua varying in lexical tone and VOT in a similar categorical manner as neuro-typical peers, (2) whether the performance in the CP of speech could index chronological age, language ability as well as phonological working memory in individuals with ASD, and (3) Especially, whether different types of speech and nonspeech pitch carriers with varying levels of spectro-temporal complexity or phonemic/semantic relevance could exert an impact on the perception of pitch contours.

3.2 Methods

3.2.1 Participants

We have initially recruited 22 high-functioning Mandarin-speaking adolescents with ASD and 20 age-matched TD controls to participate in this study. Assessed by the local administrant hospital, all the participants with ASD had nonverbal IQ above 70 using the Raven’s Standard Progressive Matrices Test (Raven & Court, 1998) and without moderate to severe language delays (i.e., capable of using full, complex sentences). They were screened for hearing loss using pure tone audiometry and met the criteria for normal hearing. The clinical diagnosis of ASD was established according to the DSM-5 criteria for ASD (American Psychiatric Association, 2013), and further confirmed using the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012), or Gilliam Autism Rating Scale–Second Edition (GARS-2; Gilliam, 2006) by pediatricians and child psychiatrists with expertise in diagnosing ASD in local hospitals. Approval of the research was granted by the local institutional review board of the Hong Kong Polytechnic University, and a written consent form was obtained from each participant.

To avoid the inclusion of individuals who had problems in perceiving synthetic sounds, a minimum accuracy score of 80% in the identification of two ending stimuli (i.e., prototypical tonal/aspiration category) in each continuum was required for the analyses of CP data. All the 20 TD controls (males = 14) met the accuracy criterion for the identification accuracy of both pitch contours and VOT. Of the 22 adolescents with ASD, 20 (males = 16) and 15 (males = 11) subjects with ASD met the criterion for pitch contour and VOT condition, respectively. The overview of participant characteristics is presented in Table 3.1. The two ASD subgroups did not differ from each other in terms of chronological age ($t(33) = -0.19$, $p = .848$), language ability ($t(33) = -0.09$, $p = .926$), forward digit span ($t(33) = -0.03$, $p = .979$), as well as nonword repetition ($t(33) = -0.15$, $p = .881$). Moreover, both ASD subgroups had similar chronological age and forward digit span as the TD controls (*p*s > .05), but significantly lagged behind the TD controls in terms of language ability (both *p*s < .01) as well as nonword repetition (both *p*s < .001).

Table 3.1 *Means (and standard deviations) of chronological age, language ability, forward digit span, and nonword repetition for two ASD subgroups in the perception of pitch contours and VOT respectively, and TD controls.*

Group	Number (male)	Chronological Age	Language Ability	Forward Digit Span	Nonword Repetition
ASD (Pitch)	20 (16)	13.87 (2.88)	87.40 (10.38)	6.45 (1.99)	75.03% (7.83%)
ASD (VOT)	15 (11)	14.06 (2.87)	87.73 (10.50)	6.47 (1.65)	75.42% (7.00%)
TD controls	20 (14)	13.46 (0.79)	98.60 (2.06)	7.40 (0.77)	84.96% (2.72%)

3.2.2 Stimuli

Table 3.2 *The stimulus features of four types of pitch carriers.*

Type	*Pitch contrast*	*Spectro-temporal complexity*	*Phonemic contrast*	*Semantic contrast*
Real word (speech)	+	+	+	+
Nonword (speech)	+	+	+	-
IRN (nonspeech)	+	+	-	-
Pure Tone (nonspeech)	+	-	-	-

The pitch contours ranging from Mandarin Tone 2 (high-rising tone) to Tone 1 (high-level tone) were embedded in four types of sound materials: real word (speech), nonword (speech), IRN (nonspeech), and pure tone (nonspeech). The stimulus features are shown in Table 3.2 in terms of pitch contrast, spectro-temporal complexity, phonemic contrast, and semantic contrast among four types of pitch carriers. The Mandarin monosyllabic words *ba*/pa/ with the Tone 1 and Tone 2 were recorded by a native female speaker (44100 Hz sampling rate, 16-bit resolution). On the basis of the natural pitch templates with Tone 2 (stimulus #1, meaning "to pull") and Tone 1 (stimulus #7, meaning "eight"), the seven stimuli along lexical tone continuum (Figure 3.1a) were synthesized using TANDEM-STRAIGHT software (Kawahara et al., 2009). Then, the seven pitch tiers were extracted and superimposed on the other three pitch carriers, including nonword (Figure 3.1b), IRN (Figure 3.1c), and pure tone (Figure 3.1d) using the Pitch-Synchronous Overlap Add implanted in Praat (Boersma & Weenink, 2016). Specially, the nonword *bü*/py/ was chosen since it does not exist in Mandarin, but the constituent units of consonant /p/ and vowel /y/ belong to native phonemes for Mandarin speakers. That is to say, the nonsense syllable of *bü*/py/ contained the phonemic contrast but not semantic contrast. Additionally, two types of nonspeech materials with different levels of spectro-temporal complexity were adopted. The nonspeech material of pure tone is acoustically much simpler than the speech carriers, while the other nonspeech of iterated ripple noise (IRN) using 64 iteration steps (Swaminathan et al., 2008) has a comparable level of spectro-temporal complexity as the speech sounds. All the stimuli were normalized to be 300 ms.

To further match the loudness level, the intensity level of pure tone was set to 85 dB, 15 dB higher than that of the other three types of pitch carriers.

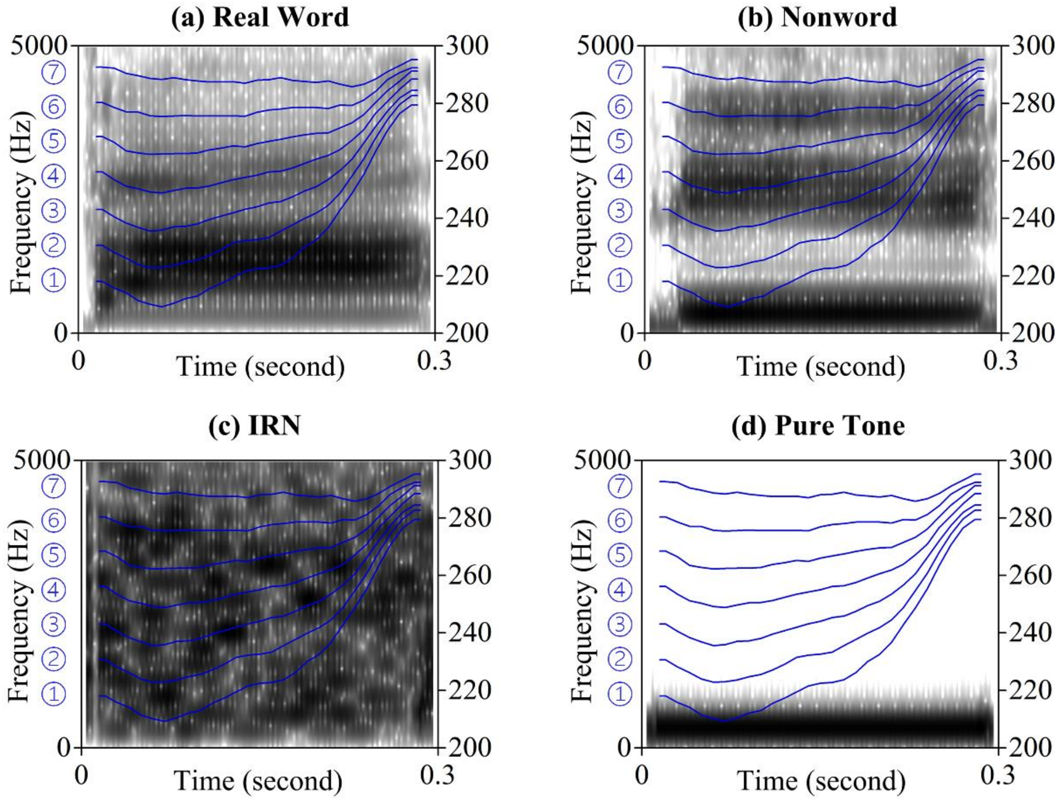


Figure 3.1 Schematic diagram of pitch contours embedded in (a) real word, (b) nonword, (c) IRN, and (d) pure tone. The right-side y-axis indicates the F0 in Hz. The blue curves indicate the pitch contours along each continuum.

A schematic diagram of the seven stimuli along the VOT continuum is shown in Figure 3.6(a). The VOT continuum was synthesized with the following procedures. First, the monosyllabic word *pa*/p^{h}a55/ (with Tone 1, meaning “lying down”) was produced from the same native female speaker, which was used as the basis for manipulation. Then, the syllable *pa*/p^{h}a55/ was normalized to be 300 ms, and it was divided into three parts: the burst release (~5 ms), aspiration (~48 ms), and vowel /a55/ (~247ms). The burst release referred to the abrupt burst in the waveform caused by the sudden release of the oral closure when producing stops; the aspiration part contained the frication noise along with the expiratory airflow right after the release of closure and before the vowel portion. During manipulation, the burst release was kept constant, while the aspiration part and vowel part were shortened and lengthened respectively in seven steps ($\Delta = 8$ ms). Lastly, the three parts were concatenated in Praat, generating a continuum of seven equally distanced stimuli ranging from the unaspirated *ba*/pa55/ (stimulus #1, meaning “eight”) to the aspirated *pa*/p^{h}a55/ (stimulus #7, meaning “lying down”). All the seven stimuli along the VOT continuum were set to be 300 ms in duration, and 70 dB in mean intensity.

3.2.3 Tasks and procedure

Before performing experimental CP tasks (identification test and discrimination test), three additional tasks evaluating language ability, digit span, and nonword repetition were performed respectively for each subject.

Language ability: The overall language ability (Chen et al., 2017; Ning, 2013) was evaluated for each Mandarin-speaking participant, which consists of five subtests (including Test of Mandarin Grammar, Word Definition Test, Rapid Automatized Naming, Narrative Test, and Sentence Comprehension Test). These subtests evaluated both language comprehension and language expression, and aimed to assess different aspects of language abilities such as phonology, lexicons, grammar, and semantics. The administration time is around 30 minutes.

Digit span: In order to evaluate the short-term phonological working memory, we administered a digit span task, which included both the forward digit span and backward digit span. However, during data collection, some of the individuals with ASD could not fully understand and follow the instructions of backward digit span with a requirement on both storage information and manipulation by the executive control (Hamann, 2017). Consequently, only the results from the simpler task of forward digit span were reported to ensure the reliability of performance. In the forward digit span task, a series of numbers were played to participants auditorily and they were asked to repeat them immediately. For each digit length (two to nine digits), there were two separate items (see Appendix A). The response for each item was regarded as correct and awarded 0.5 point only when the participants could correctly repeat every digit in the right order. The full score for the test of forward digit span is 8.

Nonword repetition: The nonword repetition task was comprised of 60 items divided equally into 20 monosyllabic (e.g., rai4), 20 disyllabic (e.g., bong1nua2), and 20 trisyllabic (e.g., sua3piong4buai1) nonwords, which is compatible with the number of syllables in previous studies (Gathercole et al., 1991); see the Appendix B for the stimuli adopted in this task. Some of the syllables carried a nasal coda, and each syllable carried diphthongs or triphthongs together with one of the four Mandarin lexical tones, in an effort to increase the task demand and phonological complexity. All the onsets, finals, and tones existed in Mandarin phonology, but the combined syllables within each nonword did not exist in Mandarin. The items, prerecorded by a female native speaker of Mandarin, were played back to participants using E-Prime 2.0 (Psychology Software Tools Inc., USA) on a Windows-based laptop. Monosyllabic, disyllabic, and trisyllabic nonwords were presented in three separate blocks with a pause between the blocks; the items within each block were randomized. Three practice trials were presented within each block before the experimental trials to familiarize subjects with the task. All the TD and ASD participants were asked to repeat the nonwords as accurately as possible. The repeated productions from each subject were recorded, and the recordings were transcribed and scored by a native phonetician afterwards.

For each item, the average percentage of phonemes (including consonant, vowel, and lexical tone) correctly repeated per nonword was calculated.

Experimental CP task 1: Identification test. Two classical tests for the CP of speech sounds, the identification test and the discrimination test, were both conducted via E-Prime 2.0. First, in the identification test, participants were asked to perform a two-alternative forced-choice (2AFC) paradigm. In the identification training of pitch contours among the four continua, participants were trained to point to one picture depicting a car driving on a level road when heard a ‘level tone (Tone 1)’, and point to the other picture depicting a car driving on a rising road when heard a ‘rising tone (Tone 2)’. In the identification training of VOT continuum, participants were trained to point to one picture with a blue circle when heard the sound of ‘*ba*/pa55/’, and point to the other picture with a blue square when heard ‘*pa*/p^{h}a55/’. After participants have acquired the matching between the sounds and their corresponding pictures, a practice block was offered before formal testing. In the practice block, the two ending stimuli of each continuum (#1 and #7) were repeated four times, and minimum accuracy of 80% was required before moving on to the formal block. During the practice blocks, the feedback was offered to the participant, but not in the formal blocks. In the formal block, each stimulus was repeated five times and played randomly. There were totally five different blocks of five continua, including four continua for the pitch condition and one continuum for the VOT condition. Subjects were asked to identify 175 sounds in total (7 stimuli × 5 repetitions × 5 continua) among five formal blocks. The four identification blocks containing four types of pitch contours were randomly presented, and the order of pitch and VOT conditions was also presented in a random order among participants. The participants’ responses were logged by the experimenter via pressing the corresponding keys on the keyboard. Both TD and ASD adolescents were free to have a rest whenever they wanted. The whole identification test lasted around 25 minutes for each participant.

Experimental CP task 2: Discrimination test. For the discrimination test, the AX paradigm was adopted to instruct subjects to discriminate the two sounds of each pitch/VOT pair as the ‘same’ or ‘different’. Also, during the training stage, participants were trained to point to one picture (a happy face with two identical eyes) representing the same pairs, and point to the other picture (a sad face with two different eyes) representing the different pairs. Each practice block contained four pairs along the pitch/VOT continuum (i.e., 1-1, 7-7, 1-7, and 7-1), with each practice pair repeating twice. The minimum discrimination accuracy of 80% was required before moving to the formal block. Feedback was provided to the participants in the practice blocks, but not in the formal blocks. There were 10 testing pairs for each pitch/VOT continuum in the 2-step discrimination task, including six pairs (different pairs) consisting of two different stimuli separated by 2 steps in either forward (1-3, 3-5, 5-7) or reverse order (3-1, 5-3, 7-5), as well as four pairs (same pairs) each paired with itself (1-1, 3-3, 5-5, 7-7). Each pair was repeated four times randomly within one formal block, with a 500 ms inter-stimulus interval. There were totally 200 pairs (10 pairs × 4 repetitions × 5 continua) distributed among five formal blocks (four pitch blocks and one VOT block). The four blocks of pitch discrimination were randomly presented, and the order of pitch and VOT conditions was presented in a random order as well. All participants were free to have a rest whenever they wanted, and the whole discrimination test lasted around 40 minutes for each subject.

3.2.4 Scoring and Data Analyses

For both identification and discrimination tests, only the data in formal blocks were involved in further analyses. First, identification curve was analyzed in term of two key parameters using

Probit analyses (Finney, 1971): boundary position, which is defined as the corresponding 50% crossover point in a continuum, and the boundary width, defined as the linear distance along the stimulus step between the 25th and 75th percentiles (Peng et al., 2010). The boundary position refers to the identification midpoint dividing the two tonal/aspiration categories, and the boundary width indicates the steepness of the response shift around the categorical boundary. Importantly, the boundary width was used to measure the degree of CP in the identification test (Chen et al., 2017). The narrower the boundary width, the steeper the boundary shift, and vice versa.

Second, the discrimination pairs were divided into three comparison units (units 1-3, 3-5, and 5-7), each containing four types of discrimination pairs: the same pairs (AA and BB) and different pairs (AB and BA). Then, the discrimination accuracy (%) was transferred into the sensitivity index *d′* for each comparison unit (Macmillan & Creelman, 2005), which takes response bias into consideration. Specifically, for each comparison unit, the d-prime (*d′*) score was computed as the difference between standard normal deviate (z-score) of hit rate ("different" responses to different pairs: AB and BA) and that of false alarm rate ("different" responses to the same pairs: AA and BB). In reference to the boundary position, the comparison units were further classified as between-category type and within-category type for each individual subject. For instance, if one participant showed a boundary position of 3.94, then for this subject, the between-category sensitivity referred to the *d′* of the comparison unit of 3-5, while the within-category sensitivity was calculated as the averaged *d′* for the comparison units of 1-3 and 5-7. Finally, the *d′* score of between-category type minus that of within-category type was referred to as the "peakedness score" (Jiang et al., 2012), which represents the benefit magnitude in the discrimination test.

Statistical analyses were conducted using linear mixed-effect models (LMMs) in R (R Core Team, 2014), by using the package of lme4 (Bates et al., 2014) to create the LMMs. Data points with standardized residuals over 2.5 standard deviations were removed that did not follow a normal distribution. For the condition of pitch perception, the models were built with *group* (ASD vs. TD), *pitch carrier* (real word, nonword, IRN, and pure tone), and their two-way interaction acting as fixed factors to analyze the boundary width and boundary position for identification analysis, as well as the peakedness score for discrimination analysis. Another LMM was constructed for the pitch discrimination performance using *category type* (within-category vs. between-category), *group* (ASD vs. TD), *pitch carrier* (real word, nonword, IRN, and pure tone), and all possible interactions acting as fixed factors. Furthermore, for each LMM in VOT condition, the boundary width/boundary position/peakedness score was entered as the dependent measure, with *group* (ASD vs. TD) acting as the fixed effect. In addition, to compare the within- vs. between-category sensitivities to VOT changes, the LMM was built with *category type* (within-category vs. between-category), *group* (ASD vs. TD), and their interaction acting as fixed factors. When fitting all the LMMs in the analyses of identification and discrimination data, the factors of *language ability*, *digit span*, and *nonword repetition* were regarded as controlled covariates, which were centered to reduce multicollinearity; *participant* was included as a random effect. By-participant random intercepts and slopes for all possible fixed factors were included in the initial model (Barr et al., 2013), which was compared with a simplified model that excluded a specific fixed factor using the ANOVA function in lmerTest package (Kuznetsova et al., 2017). Post-hoc pairwise comparisons were performed using the lsmeans package (Lenth, 2016) with Tukey adjustment.

Furthermore, linear regression models were constructed in R to examine the potential variables contributing to the ASD participants' CP performance. The approach of linear regression models is considered superior to traditional methods of correlation analyses such as Pearson/Spearman’s correlation (Koerner & Zhang, 2017), since the linear regression models consider the mutual influence of different predictors. Hypothesized predictors for the CP of pitch/VOT included chronological age, language ability, forward digit span, and nonword repetition. Evaluation of the degree of CP in the current report included (1) boundary width across four types of pitch carriers, (2) peakedness score across four types of pitch carriers, (3) boundary width in the VOT condition, (4) peakedness score in the VOT condition. Separate models were created for each estimate of CP performance, with all the four predictors added as fixed effects. Parameter estimates, standard errors, *t* values, and *p* values for the fixed effects were assessed and reported.

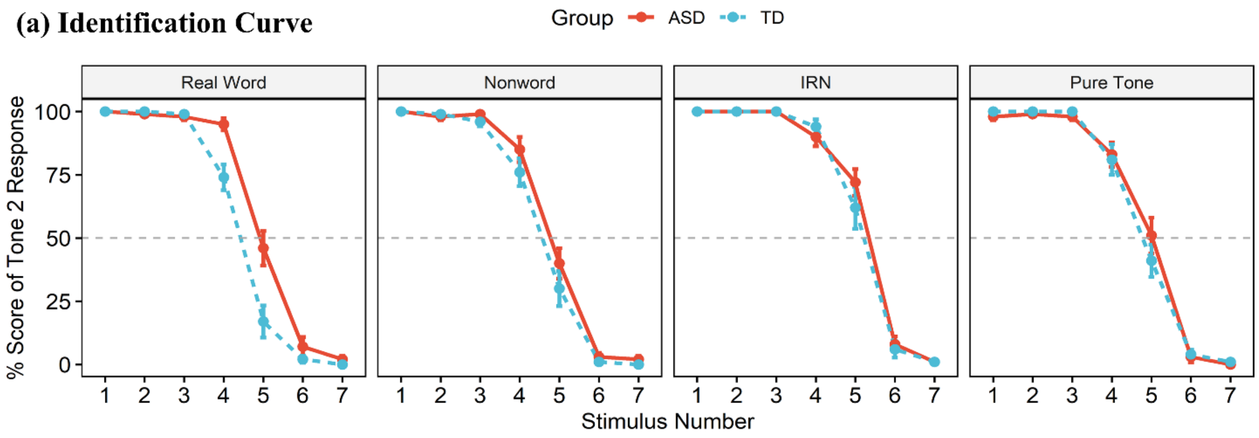
3.3 Results

3.3.1 Categorical Perception of Pitch Contours

3.3.1.1 Identification Result

Figure 3.2(a) shows the overall identification curves, and Figure 3.2(b) displays the boundary positions for the ASD and the TD groups among four types of pitch carriers. The LMM on boundary position showed a significant two-way interaction of *group* × *pitch carrier* (χ^2 (3) = 13.07, $p < .01$), which was further analyzed under different types of pitch carriers respectively. In the real word condition, compared to the TD group (M = 4.42, SD = 0.43), the ASD group (M = 4.97, SD = 0.41) exhibited a much larger boundary position which was closer to the level end (β = 0.52, SE = 0.19, t = 2.73, $p < .01$). However, as shown in Figure 3.2(b), ASD group and TD group showed a similar boundary position in nonword condition (β = 0.25, SE = 0.19, t = 1.29, p = .202), in IRN condition (β = 0.04, SE = 0.19, *t* = 0.20, *p* = .839), as well as in pure tone condition (β = -0.01, SE = 0.19, *t* = -0.06, *p* = .954).

(a) Identification Curve



(b) Boundary Position

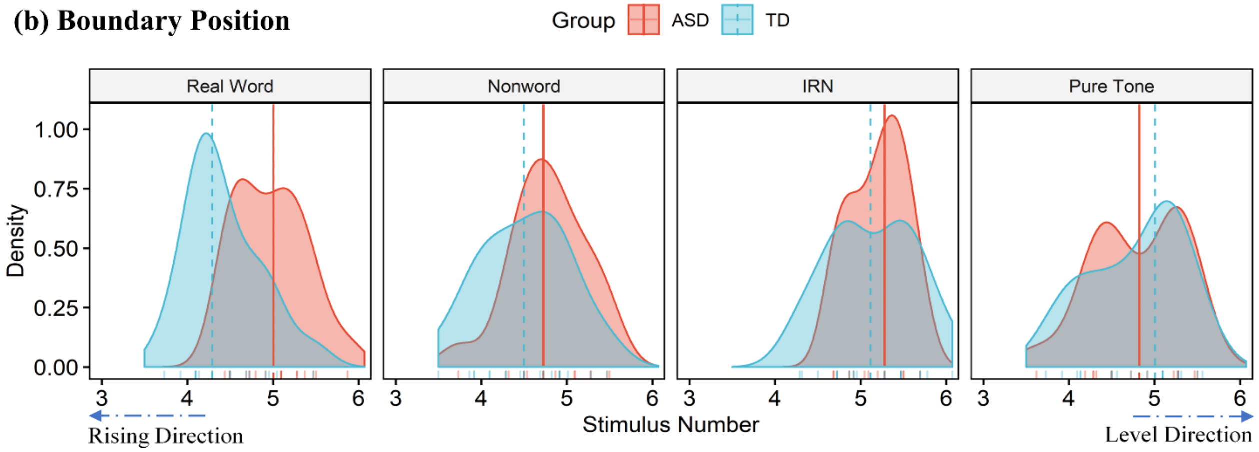


Figure 3.2 (a) The identification curves of Tone 2 responses, and (b) boundary positions in ASD and TD groups among four types of pitch contours (real word, nonword, IRN, and pure tone). The vertical lines indicate the median values of the boundary position.

Moreover, the obtained boundary widths for the ASD and the TD groups among four types of pitch carriers were shown in Figure 3.3. The mean boundary widths (*SD*) for adolescents with ASD and TD controls were 0.84 (0.55) and 0.63 (0.33) respectively in the real word condition, 0.87 (0.68) and 0.78 (0.40) in nonword condition, 0.84 (0.43) and 0.66 (0.20) in IRN condition, 0.93 (0.57) and 0.71 (0.29) in pure tone condition. The LMM on boundary width revealed neither significant main effects of *group* (χ^2 (1) = 0.97, *p* = .325), *pitch carrier* (χ^2 (3) = 2.01, *p* = .569), nor significant interaction of *group* × *pitch carrier* (χ^2 (3) = 0.82, *p* = .845). The obtained boundary width in the identification of pitch contours did not differ between ASD and TD groups, and among different types of pitch carriers (Figure 3.3).

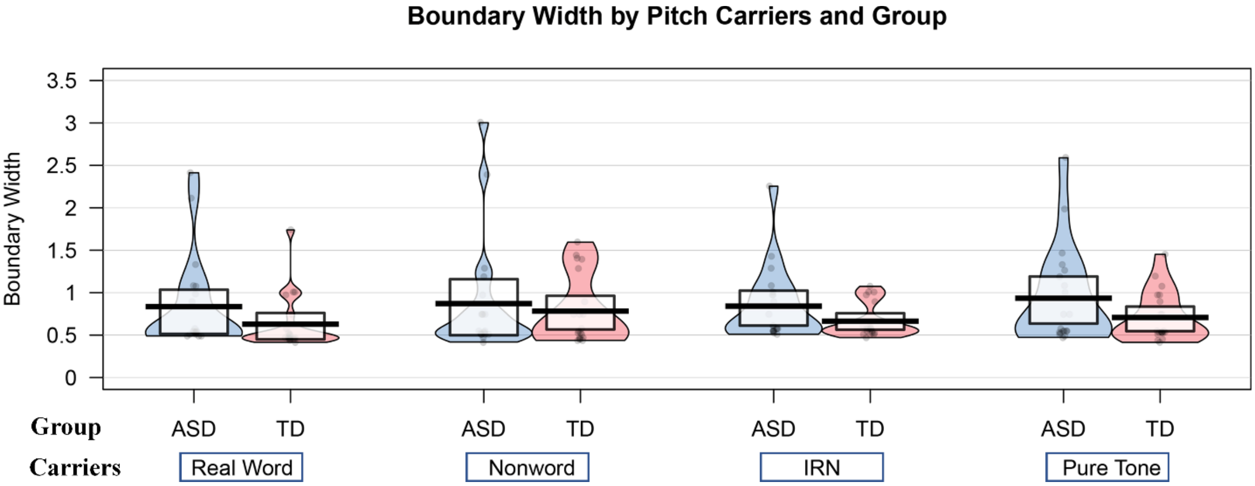


Figure 3.3 The boundary widths in ASD and TD groups among four types of pitch contours (real word, nonword, IRN, and pure tone).

3.3.1.2 Discrimination Result

The *d'* values of the between-category and within-category comparison units in ASD and TD groups are displayed in Figure 3.4 across different pitch carriers. Statistical analysis revealed a significant three-way interaction of *group* × *category type* × *pitch carrier* (χ^2 (3) = 8.47, *p* < .05), which was further analyzed under different pitch carriers respectively. First, in the real word condition, LMM on *d'* values revealed significant main effects of *category type* (χ^2 (1) = 19.16, *p* < .001), and *group* (χ^2 (1) = 5.28, *p* < .05), while the interaction of *group* × *category type* did not reach significance (χ^2 (1) = 0.31, p = .575). Second, in the nonword condition, there were significant main effects of *category type* (χ^2 (1) = 19.81, p < .001), and *group* (χ^2 (1) = 6.16, p < .05) on the *d'* values, while there was no significant interaction effect of *group* × *category type* (χ^2 (1) = 2.32, p = .127). Third, the LMM on *d'* values in the IRN condition showed significant main effect of *category type* (χ^2 (1) = 23.46, p < .001), while the interaction effect of *group* × *category type* was not significant (χ^2 (1) = 0.06, p = .813). These findings above indicated that both ASD and TD groups showed much higher *d'* values in response to between-category unit compared to the within-category unit across pitch carriers of real word, nonword, and IRN. Moreover, as shown in Figure 3.4, relative to TD controls, the ASD group showed relatively smaller *d'* values in discriminating both within-category and between-category comparison units in pitch carriers of real word and nonword. Fourth, in the pure tone condition, LMM on *d'* values exhibited significant main effect of *category type* (χ^2 (1) = 35.30, p < .001), as well as interaction effect of *group* × *category type* (χ^2 (1) = 10.47, p = .001). Post-hoc pairwise comparisons showed that relative to TD group, the ASD group had a lower *d'* in response to between-category unit (β = -0.89, SE = 0.33, t = -2.70, p < .01), while the two groups had similar *d'* values in response to within-category unit (β = 0.21, SE = 0.33, t = 0.62, p = .535). Moreover, the between-category unit generated a much higher *d'* value than the within-category type in pure tone condition (Figure 3.4), for both ASD group (β = 0.64, SE = 0.23, t = 2.80, p < .01), and TD group (β = 1.74, SE = 0.23, t = 7.57, p < .001).

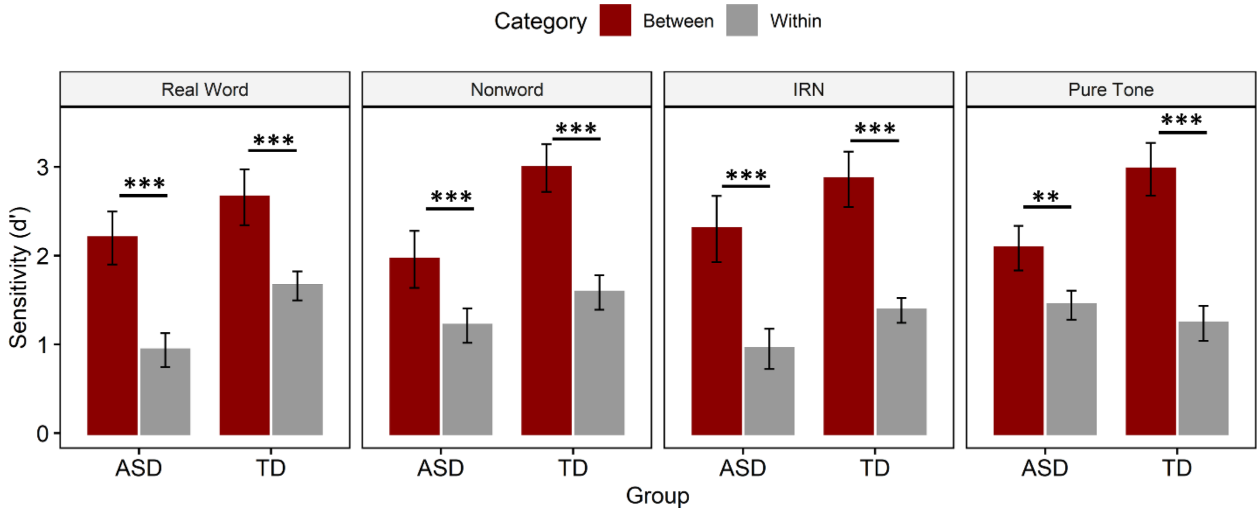


Figure 3.4 The *d'* of the between- and within-category units for the ASD and TD participants under different types of pitch carriers (real word, nonword, IRN, and pure tone). Error bars: +/- 1 standard error.

The peakedness scores (the *d'* score of between-category unit minus that of within-category unit) in TD and ASD groups are shown in Figure 3.5 across different pitch carriers. The mean peakedness scores (*SD*) for ASD group and TD group were 1.26 (1.50) and 1.00 (1.59) respectively in the real word condition, 0.75 (1.54) and 1.40 (1.19) in nonword condition, 1.35 (2.00) and 1.48 (1.76) in IRN condition, 0.67 (1.09) and 1.74 (1.00) in pure tone condition. For the LMM on peakedness score in the discrimination of pitch contours, neither main effects of *group* (χ^2 (1) = 0.88, p = .348), *pitch carrier* (χ^2 (3) = 1.47, p = .688), nor interaction effect of *group* × *pitch carrier* (χ^2 (3) = 5.63, p = .131) reached significance. The peakedness scores were comparable between ASD and TD groups, and among different types of pitch carriers (Figure 3.5).

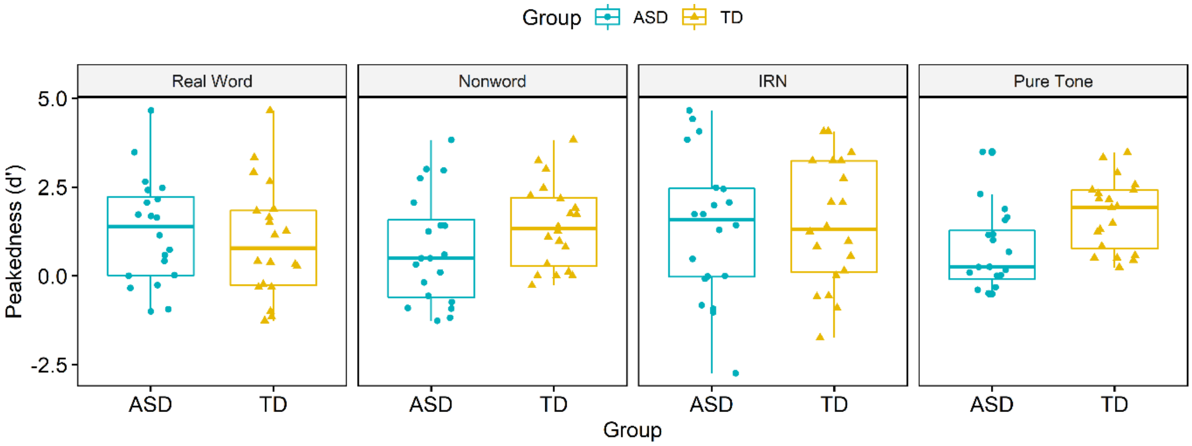


Figure 3.5 Box plots of peakedness scores (*d'*) for the ASD and TD participants under different types of pitch carriers (real word, nonword, IRN, and pure tone).

3.3.1.3 Linear Regression Result

Table 3.3 shows the regression coefficients indicating the relationships between predictors (chronological age, language ability, digit span, nonword repetition) and the degree of CP of pitch contours in individuals with ASD. In the real word condition, language ability (β = -0.03, SE = 0.01, t = -3.01, $p < .01$) and digit span (β = -0.11, SE = 0.05, t = -2.17, $p < .05$) were significant predictors for the boundary width (Table 3.3). The negative regression coefficients (β) indicated that the better language ability or digit span in ASD led to a narrower boundary width (i.e., a steeper identification slope) in the identification of lexical tones in real word. In the nonword condition, only nonword repetition in ASD was significantly associated with the boundary width in the identification of pitch contours embedded in nonword material (β = -5.30, SE = 1.66, t = -3.20, $p < .01$), with higher accuracy of nonword repetition contributing to a steeper slope. In the nonspeech pitch carriers of IRN and pure tone, however, none of the significant correlations between predictors and the degree of CP were detected as shown in Table 3.3.

Table 3.3 *The regression coefficients indicating the relationships between chronological age/language ability/digit span/nonword repetition and the degree of CP of pitch contours (boundary width/peakedness score) in individuals with ASD.*

Pitch Carriers	Predictors	Boundary Width				Peakedness Score			
		β	SE	*t*	*p* value	β	SE	*t*	*p* value
Real Word	Chronological Age	0.018	0.028	0.615	0.548	0.130	0.115	1.132	0.275
	Language Ability	**-0.032**	**0.011**	**-3.009**	**0.009****	-0.021	0.043	-0.482	0.637
	Digit Span	**-0.109**	**0.050**	**-2.170**	**0.046***	0.361	0.203	1.784	0.094
	Nonword Repetition	1.128	1.186	0.951	0.357	-7.081	4.774	-1.483	0.159
Nonword	Chronological Age	0.057	0.047	1.209	0.246	0.012	0.145	0.080	0.937
	Language Ability	0.013	0.016	0.801	0.435	0.046	0.049	0.918	0.373
	Digit Span	-0.084	0.077	-1.089	0.293	-0.296	0.236	-1.254	0.229
	Nonword Repetition	**-5.302**	**1.656**	**-3.201**	**0.005****	-3.043	5.079	-0.599	0.558
IRN	Chronological Age	0.005	0.037	0.129	0.899	-0.034	0.171	-0.197	0.846
	Language Ability	-0.020	0.013	-1.575	0.136	0.053	0.059	0.900	0.382
	Digit Span	-0.040	0.060	-0.675	0.510	0.376	0.279	1.346	0.198
	Nonword Repetition	1.234	1.286	0.959	0.353	-2.696	6.015	-0.448	0.660
Pure Tone	Chronological Age	0.056	0.045	1.232	0.237	0.012	0.105	0.113	0.911
	Language Ability	-0.022	0.016	-1.417	0.177	-0.033	0.036	-0.931	0.367
	Digit Span	-0.098	0.074	-1.333	0.202	0.026	0.171	0.149	0.883
	Nonword Repetition	0.542	1.589	0.341	0.738	3.545	3.678	0.964	0.350

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

3.3.2 Categorical Perception of VOT

Figure 3.6(b) showed the identification curves of VOT perception in ASD and TD groups. The analysis on boundary position of VOT continuum showed that two groups had a similar boundary position (β = 0.33, SE = 0.38, t = 0.89, p = .383). However, compared to ASD group, the TD controls showed a much narrower boundary width in the identification of VOT as shown in Figure 3.6c (β = -0.88, SE = 0.30, t = -2.92, p < .01), as well as a higher peakedness score in the discrimination of VOT as shown in Figure 3.6e (β = 1.65, SE = 0.62, t = 2.68, p < .05). Furthermore, the LMM on *d'* values of different category types in VOT condition exhibited a significant interaction effect of *group* × *category type* (χ^2 (1) = 8.60, p < .01). Post-hoc analyses demonstrated that both ASD group (β = 1.23, SE = 0.43, t = 2.89, p < .01) and TD group (β = 2.88, SE = 0.37, t = 7.82, p < .001) showed a higher between-category *d'* value relative to the within-category one

(Figure 3.6d). In addition, the ASD participants showed a much lower *d'* in response to between-category unit (β = -2.07, SE = 0.40, *t* = -5.20, *p* < .001), while the two groups had similar *d'* values in response to within-category unit during VOT perception (β = -0.42, SE = 0.40, *t* = -1.05, *p* = .297).

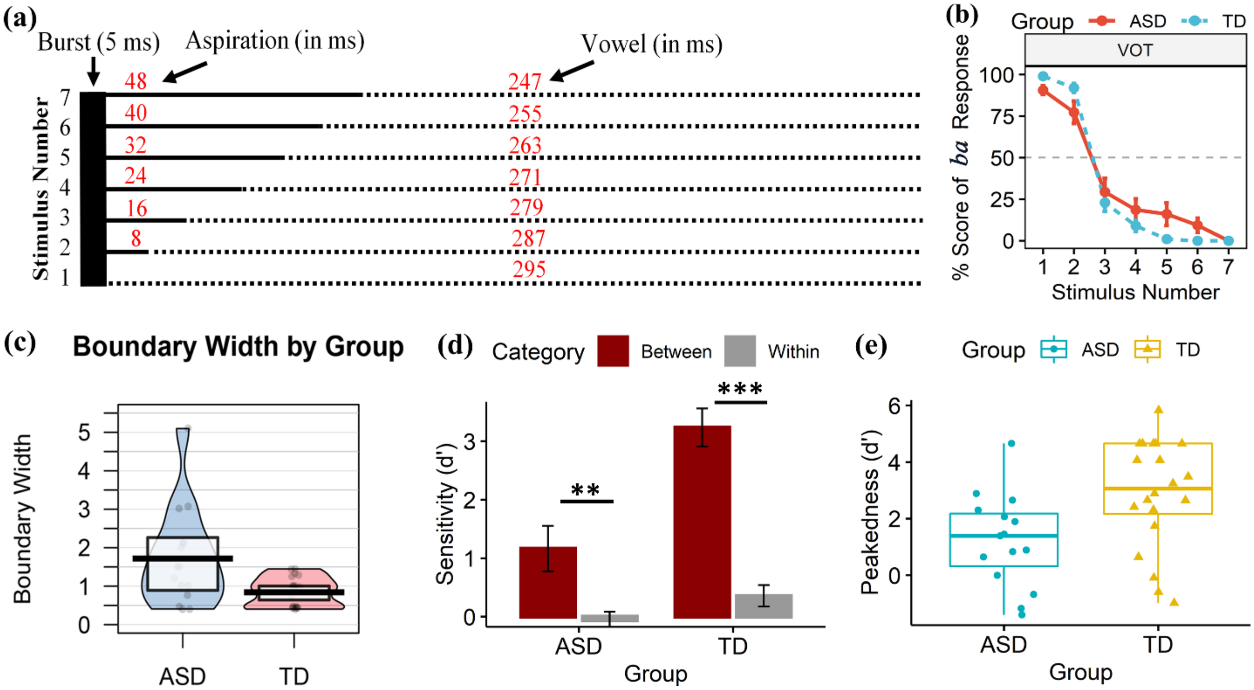


Figure 3.6 (a) Schematic diagram of VOT continuum; (b) The identification curves of unaspirated *ba*/pa/ responses; (c) The boundary widths for ASD and TD groups in the identification of VOT; (d) The *d'* values of the between-category and within-category units for the ASD and TD participants in the discrimination of VOT; (e) Box plots of peakedness scores (*d'*) for the ASD and TD participants in discrimination of VOT.

The regression coefficients (estimates, standard errors, *t* values, and *p* values) are presented in Table 3.4, indicating the relationships between predictors and the degree of CP of VOT among

ASD participants. None of the four predictors in ASD participants revealed a significant relationship with the boundary width in VOT identification test (all *p*s > .05). Similarly, there were no significant correlations between the ASD participants' performance on chronological age/language ability/digit span/nonword repetition and the peakedness score in VOT discrimination test (all *p*s > .05).

Table 3.4 *The regression coefficients indicating the relationships between chronological age/language ability/digit span/nonword repetition and the degree of CP of VOT (boundary width/peakedness score) among participants with ASD.*

	Predictors	Boundary Width				Peakedness Score			
		β	SE	*t*	*p* value	β	SE	*t*	*p* value
VOT	Chronological Age	-0.001	0.141	-0.010	0.992	-0.134	0.175	-0.772	0.458
	Language Ability	0.068	0.078	0.874	0.403	-0.106	0.096	-1.094	0.299
	Digit Span	-0.354	0.490	-0.721	0.487	0.843	0.604	1.395	0.193
	Nonword Repetition	-5.537	5.480	-1.010	0.336	3.182	6.745	0.472	0.647

3.4 Discussion

3.4.1 Preserved CP Pattern in High-functioning Adolescents with ASD

CP provides an account for how human symbolic thinking is grounded in perception and action (Zhang, 2016), which refers to a tendency for native listeners of a particular language to classify the speech signals with infinite variability as discrete, finite, and linguistic representations. To enhance our perceptual efficiency, we would suppress the ability to distinguish irrelevant acoustic information by showing 'dulled' sensitivity to within-category differences or allophonic variations, whereas showing enhanced sensitivity to the between-category stimulus pair that spans the boundary between two categories. There is no support for claiming that CP pattern occurs if there is no such benefit for between-category compared to within-category discrimination (Liberman et al., 1957; Massaro, 1987). Thus, the "cross-boundary benefit" is likely to be the defining feature of CP pattern, which was well observed in Mandarin-speaking TD children and adults during the

CP of Mandarin tones (Chen et al., 2017; Wang, 1976; Xi et al., 2010; Xu et al., 2006) and VOT (Cheung et al., 2009; Feng, 2018; Xi et al., 2009) from both behavioral and neuroimaging studies. For individuals with ASD, however, given their lack of early interest in social interaction and preference to speech sounds (Constantino et al., 2004, 2007; Dawson et al., 1998; Klin, 1991; Kuhl et al., 2005), as well as their atypical processing bias towards local details and low-level acoustic features (Frith, 1989; Happé & Frith, 2006; Mottron & Burack, 2001), they might show difficulties in extracting relevant invariant phonetic features and forming proper phonological categories, as reflected by a reduced or even impaired CP of speech sounds in ASD.

In line with our prediction, by using the mismatch negativity (MMN) paradigm, Wang et al. (2017) revealed a speech-specific deficit in CP of lexical tones in Mandarin-speaking children with autism (mean age = 10.4 years), with similar MMR amplitudes elicited from between-category tonal deviants relative to within-category deviants. Another behavioral study (Chen et al., 2016) in low-functioning young children with autism (mean age = 7.6 years; developmental age = 3.7 years) also failed to reveal a cross-boundary benefit by observing similar discrimination accuracies for the between- and within-category tonal pairs. These corroborating findings seemed to imply an impaired CP of lexical tones in tone language speakers with ASD. However, the notion of an impaired CP pattern (i.e., continuous/non-categorical perception pattern) among all the autistic individuals must be interpreted with caution. Firstly, the MMR component (Wang et al., 2017) is elicited without behavioral requirements and in the absence of focal attention, it is highly likely that the attention of ASD perceivers per se might exert an influence on phonological processing of speech sounds (Whitehouse & Bishop, 2008). Secondly, the autistic spectrum showed huge variability in terms of speech and language development. The high-functioning ASD with better language and cognitive capacity may not necessarily show an impaired CP pattern.

Thirdly, even in TD children, they usually display a less precise categorization of speech sounds compared to healthy adults (Hoonhorst et al., 2011; Medina et al., 2010), which was related to the less sufficient experience to native speech sounds compared to adults. Similarly, children with ASD might merely exhibit "weaker or delayed" category formation compared to age-matched TD controls (Soulières et al., 2007), but rather an "impaired" CP pattern throughout their lifespan. To test these hypotheses, by using a behavioral CP paradigm, we investigated the performance of CP of native speech sounds (lexical tones and VOT) in high-functioning adolescents with ASD without severe language/cognitive delays.

As seen in Figure 3.4 and Figure 3.6(d), the high-functioning adolescents with ASD in the current study did perceive the lexical tones and VOT in a preserved CP pattern, as indicated by a much higher *d'* for between-category pairs than for within-category pairs in both types of continua. Furthermore, the preserved CP pattern of cross-boundary benefit in the speech context was transferred to nonspeech counterparts for both ASD and TD groups (Figure 3.4), reflecting a carry-over influence of long-term phonological processing from the speech to nonspeech domain. The preserved CP pattern in ASD was also detected in the perception of other types of speech sounds, such as the CP of vowels (/i-y /continuum) and consonants by place of articulation (/d-b/continuum) in both high-functioning children with autism and Asperger syndrome (You et al., 2017), as well as the CP of VOT (/g-k/ continuum) in high-functioning and cognitively able adults with ASD (Stewart et al., 2018). Such perceptual pattern was quite different from non-native speakers who showed a continuous or non-categorical perception pattern with similar sensitivity to within- and between-category pairs (Xu et al., 2006). Thus, following the above evidence from the perception of various types of speech sounds, we would be confident to infer that the impaired CP pattern might not be applicable to all the autistic individuals, but rather tend to be part of a shared vulnerability of language or cognitive delay/impairment in a subgroup of ASD.

3.4.2 The Influence of Low-level Acoustic Processing on the High-level Phonological Processing

Auditory processing abnormality has been suggested as one of the key factors underlying the pathological speech and language processing in ASD (Alcántara et al., 2004; Čeponienė et al., 2003; O'connor, 2012). In the research field of general auditory processing, a plethora of studies have implied that individuals with ASD were reported to show atypical and unbalanced auditory processing depending on the acoustic dimensions (spectral vs. temporal) (Alcántara et al., 2012; Groen et al., 2009; Yu, 2018). Specifically, individuals with ASD as a whole group tended to exhibit an enhanced acoustic processing skill of pitch (Bonnel et al., 2003; Foxton et al., 2003; Heaton et al., 2008; Heaton, 2003, 2005; Mottron et al., 2000; O'Riordan & Passetti, 2006), while compromised acoustic processing of sound duration (Brodeur et al., 2014; Falter et al., 2012; Maister & Plaisted-Grant, 2011; Martin et al., 2010; Szelag et al., 2004). As CP of speech sounds reflects the higher-level phonological processing mode, by comparing the CP competence of linguistic pitch (lexical tone) and linguistic time (VOT) in native speakers with ASD at the same time, it can help uncover whether and how lower-level acoustic processing could influence higher-level phonological processing.

Based on the current findings, although both the perception of lexical tones and VOT showed a typical CP pattern in Mandarin-speaking adolescents with ASD, the degree of CP between the two types of continua varied greatly. As for the behavioral indexes of CP competence (Chen et al., 2019), the boundary width was used to measure the degree of CP in the identification function, with a narrower boundary width indicating a steeper slope, and vice versa; the peakedness score reflected the magnitude of cross-boundary benefit in a discrimination test. During the perception of lexical tone continuum, the ASD participants showed similar boundary width (Figure 3.3) and peakedness score (Figure 3.5) relative to the neuro-typical peers. In stark contrast, during the perception of VOT continuum, the ASD group showed a much wider boundary width (Figure 3.6c) and lower peakedness score (Figure 3.6e). Taken together, for the high-functioning Mandarin-speaking adolescents with ASD, the CP of native lexical tones was largely intact, meanwhile the degree of CP of VOT was greatly reduced. These findings suggest that the unbalanced acoustic processing capacities for pitch and time can be generalized to higher-level linguistic processing form the evidence in ASD. There is a concern that the inferior performance on the CP of VOT in ASD could also be attributed to the relative difficulty levels since the aspirated vs. unaspirated contrast tended to be acquired later relative to lexical tones in Mandarin-speaking TD children (Hua & Dodd, 2000). However, even for the high-functioning adults with ASD, they also showed a less categorical fashion in the perception of VOT continuum when compared with IQ-matched typically developed adults (Stewart et al., 2018). Thus, the auditory processing difficulties of sound duration in autism are manifested profoundly and further persist into the higher-level phonological processing that involves the basic CP competence of VOT, and the processing of vowel length contrast phonemically to mark semantic distinction such as in Finnish-speaking (Lepistö et al., 2005, 2006) and Japanese-speaking (Kasai et al., 2005) individuals with ASD.

It has been suggested that lower-level acoustic processing and higher-level phonological processing are represented differently in our human brain. Some evidence reveals that speech is processed hierarchically along the auditory pathways, with the upstream areas (e.g., the dorsal STG areas) responsible for acoustic processing and the downstream regions (e.g., the ventral superior temporal sulcus, STS, and middle temporal gyrus, MTG regions) performing phonological processing (Okada et al., 2010; Wessinger et al., 2001; Zhang et al., 2011). Although the different hierarchy levels (dorsal STG vs. ventral STS/MTG) implied a dissociation existed between lower-level acoustic and higher-level phonological processing, the way these two levels of processing interact with each other functionally during speech sound processing is not well understood. The 'feed-forward mechanisms' (Binder, 2000; Scott & Wise, 2004) suggested that speech processing begins from the core auditory areas (STG) to downstream brain areas and then to more lateral and anterior regions, indicating that initial bottom-up acoustic processing might lay the foundation of phonological processing. The current findings provide direct evidence on the feed-forward mechanisms by showing that lower-level acoustics underlie higher-level phonological processing in speech perception since the unbalanced acoustic processing skill (pitch vs. time) in ASD extends to the different degrees of CP of speech sounds (lexical tones vs. VOT) in native perceivers from the clinical population.

3.4.3 Factors Related to the Level of CP Competence in ASD

As discussed earlier, the autistic ones showed large individual variability in terms of CP of speech sounds, with some individuals showing a profoundly impaired CP pattern (Chen et al., 2016; Wang et al., 2017) while others exhibiting the preserved CP pattern albeit with varying levels of competence (this study; Stewart et al., 2018; White et al., 2006; You et al., 2017). In the current study, we further investigated whether and how chronological age, language ability, and phonological working memory (digit span and nonword repetition) were related to the level of CP competence among individuals with ASD. As shown in Table 3.3, the overall language ability in ASD was a significant predictor for the boundary width ($p < .01$) in real word condition. The autistic participants with better language ability tended to elicit a narrower boundary width (i.e., a steeper slope) in the identification of lexical tones. The higher degree of CP of lexical tones correlated with better language ability in Mandarin-speaking high-functioning adolescents with ASD. Such correlation was observed as well in the low-functioning younger autistic children (Chen et al., 2016). Furthermore, the degree of CP was correlated with the verbal ability of reading, lexical decision, and verbal IQ in adults with ASD (Stewart et al., 2018). Collectively, the close relationship between auditory perception skills and language functions was consistently observed in individuals with ASD of various cognitive abilities and different age ranges (Bishop et al., 2004; Chen et al., 2016; Constantino et al., 2007; Stewart et al., 2018). This is consistent with the notion that early speech processing can predict later language development in TD children (Kuhl et al., 2008; Tsao et al., 2004). The implication is that some aspects of language difficulties found in individuals with ASD may be related to the reduced CP competence of speech sounds. Our current findings suggested the necessity of further examining the potential links among social competence, speech processing and language functioning among autistic individuals in further prospective longitudinal work. If the CP competence of native speech turned out to be a reliable predictor of certain language-related abilities in ASD, it would call for an inclusion of CP-related testing and training in the evaluation and intervention of ASD at an early stage.

Furthermore, the regression analyses showed that the capacity of digit span in ASD could be a contributing factor for the identification acuity of lexical tones; the nonword repetition in the autism group was a significant predictor for the identification acuity of pitch contours in nonword condition (Table 3.3). There findings were not surprising given that in the behavioral CP tests, three forms of memory—sensory memory and the short- and long-term forms of categorical memory—are involved (Xu et al., 2006). Besides, the discrimination task in AX pattern required the recruitment of short-term working memory to store one stimulus and then to compare it with the subsequent one (Mitterer & Mattys, 2017). Both digit span and nonword repetition were used to evaluate phonological working memory (Hamann, 2017; Rispens & Baker, 2012); the nonword repetition task further draws on sub-lexical knowledge to access and maintain new phonological codes, which is thought to measure the representations of "chunks' of phonemes in long-term memory (Shao et al., 2020; Szewczyk et al., 2018). The close correlations between digit span/nonword repetition and the degree of CP found in this study called for the controlling of such confounding factors of the cognitive capacities such as phonological working memory of ASD (Boets et al., 2015). Contrary to our prediction, this study failed to reveal a relationship between chronological age and the CP competence across all the stimulus conditions. However, the lack of age effect must be interpreted with caution as it may be attributed to the relatively matured perceptual development in our samples of high-functioning adolescents with ASD, and the lack of power.

3.4.4 Lexical Tone Perception Difficulties in ASD and the Underlying Mechanisms

For tone language speakers with ASD, several studies (Chen et al., 2016; Cheng et al., 2017; Wang et al., 2017; Wu et al., 2020; Yu et al., 2015) pointed to the native lexical tone perception difficulties at both behavioral and neural levels. Yet, our full understanding of lexical tone perception difficulties and its underlying mechanisms in tone language speakers with ASD are still far from complete. Some scholars have proposed a speech-specific mechanism to explain the pitch perception difficulties only in the speech context (Wang et al., 2017; Yu et al., 2015). Others tried to explain the deficits with the 'allophonic perception' theory for autism (Huang et al., 2018; M. O'Riordan & Passetti, 2006; You et al., 2017), due to the detail-oriented processing style and enhanced acoustic pitch discrimination skills in autism. By using a fine-grained CP approach, this study investigated the identification, as well as within-category and between-category discrimination of pitch contours embedded in various types of speech and nonspeech contexts. The four different types of pitch carriers (real word, nonword, IRN, pure tone) differ in the levels of spectro-temporal complexity or phonemic/semantic relevance.

The degree of CP, as assessed by both boundary width (Figure 3.3) and peakedness score (Figure 3.5), did not differ between ASD and TD groups among all the four types of pitch carriers, indicating the well-developed CP of lexical tones in high-functioning adolescents with ASD, and its carry-over influence of long-term phonological processing from the speech to nonspeech domain regardless of the word status and spectro-temporal complexity. Interestingly, the boundary position differed between ASD and TD groups only in the real word condition with semantic information. More specifically, as shown in Figure 3.2, Mandarin-speaking participants with ASD showed a much higher boundary position (i.e., closer to the level end) relative to TD controls in the real word condition, with a similar pattern called "psychophysical boundary" observed in the non-tonal language speakers who had no tonal language experience (Wang, 1976). In other words, the relative perceptual space for the level tone (Tone 1) in Mandarin-speaking individuals with ASD was compressed compared to the TD controls, with ASD participants displaying less tolerance for the ambiguous rising contours to be judged as Mandarin level tone. Compared to the other three types of pitch carriers, the real word condition additionally carried the semantic contrast with the level-ending stimuli (#7) meaning "eight" and the rising-ending stimuli (#1) meaning "to pull". As suggested by the 'Ganong effect' (Ganong, 1980; Stewart & Ota, 2008), which proposed that the boundary of phonetic categorization shifted as a function of lexical-semantic influence from real words, the enlarged perceptual space for the high-level Tone 1 in TD group might be attributed to the stronger influence from semantic effect in the high-frequency numeric word "eight". Therefore, it was possible that the autistic individuals were less susceptible to higher-level semantic capture when performing a pitch identification task, in line with the ‘Weak Central Coherence’ theory (Frith, 1989; Happé & Frith, 2006). These findings implied that one of the potential reasons responsible for the speech-specific lexical tone perception difficulties in ASD might be caused by a weaker feedback loop from the lexicon to phonemic activation (McClelland & Elman, 1986). This hypothesis should be tested in future studies with the paradigm of ‘Ganong effect’ (Ganong, 1980; Stewart & Ota, 2008), which directly investigates whether the extent to which tone categorization biases the judgment toward a known word is weakened in the ASD group relative to neurotypicals.

In the discrimination test, the sensitivity to within-category pitch discrimination was not elevated for the ASD group compared to the TD group in the speech conditions, which seemed to contradict with the ‘allophonic perception’ theory for autism (Huang et al., 2018; M. O’Riordan & Passetti, 2006; You et al., 2017). But this phenomenon should be explained with caution. On the one hand, the within-category discrimination does not merely reflect the acoustic pitch processing for the native speakers, since the “dulled” within-category sensitivity was gradually formed with native language experience by perceptually ‘tuning out’ irrelevant acoustic information. Following this line, the high-functioning adolescents with ASD in this study who had an intact CP pattern might not show the ‘allophonic perception’ feature, which was corroborated with the findings in the high-functioning adults with ASD who did not show more accurate within-category discrimination in comparison with TD adults (Stewart et al., 2018). We would speculate the ‘allophonic perception’ pattern to emerge in low-functioning young children with ASD who lack of inhibitory mechanism for suppressing the detection of irrelevant within-category pitch differences, and thus to cause an impaired CP pattern. On the other hand, the behavioral AX discrimination task taps into attentional, and working memory processes, and is not assessing discrimination, which has been noted to be unsuitable for the ASD population since a large proportion of them were accompanied with attention and working memory deficits (Heaton et al., 2008). This might be one of the reasons to explain why the autistic subjects performed inferiorly compared to controls across the board regardless of within- and between-category pitch discriminations in the conditions of real word, nonword, as well as IRN. Although with such profound attentional and memory disadvantage, the ASD group nevertheless showed a comparable *d'* values (even higher values but not statistically significant) in response to within-category pairs relative to TD controls in the pitch carrier of pure tone. Compared with other carriers of speech sounds and the nonspeech IRN, the pure tone is much simpler in terms of spectro-temporal complexity (see Figure 3.1). Our current observations on pure tone were consistent with the previous behavioral and MMN findings (Bonnel et al., 2003; Ferri et al., 2003; Gomot et al., 2002; M. O'Riordan & Passetti, 2006), which highly supports the Neural Complexity Hypothesis (Samson et al., 2006). That is, the individuals with autism may display enhancement in pitch discrimination where spectro-temporally simple but not complex stimuli yield superior performances.

3.4.5 Limitations

This study has several limitations. First, it is important to note that the current conclusions were limited to the high-functioning adolescents with ASD, but not necessarily extended to younger children or low-functioning individuals with ASD. Given the huge heterogeneity within the autistic spectrum, in order to obtain a more robust statistical power, a larger sample size with a broader range of demographic characteristics is needed. Second, one of the big challenges for performing behavioral studies in autism is the reliability and validity of the data, given the serious attention deficits in most of the individuals with ASD. To address this concern, before formal CP tests we have performed the training stage and practice blocks to familiarize the subjects with the experimental procedures, and a minimum accuracy of 80% was required before moving to the formal blocks. However, in the formal testing blocks, no control items were included to monitor the performance of the participants in this study. For instance, in the discrimination task, only the two-step discrimination pairs were incorporated as the testing stimuli, which was hard for us to judge the reliability of participants' responses. In future studies, we could also involve the easily discriminable pairs (such as the discrimination of the two ending stimuli) with much larger acoustic distance in the testing blocks in an effort to monitor the behavioral responses during the testing stage. Furthermore, only behavioral protocols were adopted in the current study. A body of evidence has revealed the underlying neural correlates of the CP of lexical tones in native tone language speakers at both the pre-attentive stage (Xi et al., 2010; Yu et al., 2014, 2017) and the attentive stage (Zheng et al., 2012) by exploiting MMN and P300 paradigms respectively. Further investigations with electrophysiological approaches are warranted to uncover the underlying neural mechanisms of the CP of lexical tones during the pre-attentive and attentive processing stages for tone language speakers with ASD, and their correlations with behavioral measurements. Finally, future longitudinal research with younger children with ASD is necessary to chart and compare the developmental trajectory of CP of lexical tones and VOT, to further investigate the age effect on the perception of linguistic pitch and linguistic time in ASD.

3.5 Conclusion

Despite large individual variability, findings of the current study revealed a preserved CP pattern when perceiving the native lexical tones and VOT in Mandarin-speaking high-functioning adolescents with ASD, with a much higher sensitivity to the between-category pairs compared to the within-category pairs. However, the degree of CP (assessed with boundary width and peakedness score) was much higher for the perception of lexical tones than the perception of VOT, reflecting the influence from lower-level acoustic processing of pitch and time. The language ability, digit span, and nonword repetition of ASD participants were found to be significant predictors for the levels of CP competence in some speech conditions of pitch perception. Furthermore, individuals with ASD showed a "psychophysical boundary" similar to the non-tonal language speakers, potentially due to the reduced access to the semantic information of real word. These findings deepened our understanding of phonological processing of different speech elements in the subgroup of high-functioning ASD without severe language/cognitive delay.

Chapter 4: Linguistic Tone and Non-Linguistic Pitch Imitation in Children with Autism Spectrum Disorders: A Cross-Linguistic Investigation

4.1 Introduction

Prosody is a broad term including suprasegmental properties such as intonation, tone, rhythm, and stress, which is used to convey various linguistic, attitudinal, emotional, pragmatic, and idiosyncratic functions (Bolinger, 1972; Cutler & Isard, 1980). The acoustic correlates of speech prosody involve pitch (fundamental frequency, F0), duration, intensity, and their coordination. The two gold-standard clinical assessments of ASD, Autism Diagnostic Observation Schedule, second version (ADOS-2; Lord et al., 2012) and Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003), have included prosodic atypicality as one of the diagnostic characteristics. Starting from the earliest delineation of the autistic syndrome with peculiar use of the tone of voice (Kanner, 1943), unusual prosody has been frequently identified as a central feature of speech and communication in autism (Baltaxe & D'Angiola, 1992; McCann & Peppé, 2003; Pronovost et al., 1966; Rutter et al., 1967; Tager-Flusberg, 1981). Furthermore, prosodic differences have been found to be closely associated with the general ratings of social and communicative competence in autism (Paul et al., 2005).

To investigate the prosodic pitch features in ASD, several studies focused on the production of intonation, which is expressed as the variation in spoken pitch at the sentence level. The earlier reports based on observation or subjective ratings revealed dull, monotonic, or machine-like intonation in speech produced by some children with autism (Fay & Schuler, 1980; Kanner, 1943, 1971; Tager-Flusberg, 1981). However, contrary to the common impression of monotonic speech in autism, most of the studies adopting acoustic analyses consistently showed a wider pitch range and/or greater pitch standard deviation (SD), indicating an increased pitch variability of intonation produced by individuals with ASD, when recorded during natural speech (Diehl et al., 2009; Green & Tobin, 2009; Nadig & Shaw, 2012; Sharda et al., 2010), during reading as well as imitating/repeating sentences produced by others (Fosnot & Jun, 1999; Hubbard & Trauner, 2007), and in the picture-naming task (Bonneh et al., 2011; Filipe et al., 2014). The only exception was reported by Nakai et al. (2014), which showed a smaller pitch range of intonation in Japanese-speaking children with ASD compared to typically-developing (TD) peers at school age. In a short conclusion, the subjective impression of "flat" intonation in ASD has not been confirmed by accumulating evidence from the acoustic analyses in most of the studies mentioned above in non-tonal language speakers with ASD, including English-speaking children (Diehl et al., 2009; Fosnot & Jun, 1999; Hubbard & Trauner, 2007; Nadig & Shaw, 2012), Hebrew-speaking children (Bonneh et al., 2011; Green & Tobin, 2009), Portuguese-speaking children (Filipe et al., 2014), and English-Hindi bilinguals with ASD (Sharda et al., 2010).

The world's languages displayed a great diversity, among which, the tonal languages such as Mandarin and Cantonese make use of F0 changes at the syllable level to mark phonological contrasts (Wang, 1973). Thus, in tonal languages, the F0-based prosodic changes could be realized not only in the larger prosodic unit of the whole sentence (i.e., intonation), but also in a smaller unit of the syllable (i.e., lexical tone). The production of intonation in tone-language-speaking individuals with ASD has been investigated by Chan and To (2016). The results showed that, in consistent with previous findings in non-tonal languages, Cantonese-speaking adults with high-functioning autism also demonstrated significantly higher SD of F0 than TD children, suggesting that the atypical sentence-level intonation may be a universal characteristic of individuals with ASD. However, to the best of our knowledge, no studies so far have depicted the syllable-level prosodic phonology in tone-language-speaking children with ASD. There has been an analogized relationship between lexical tone and intonation as “small ripples riding on large waves” (Chao, 1968:39), implying that the dynamic changes in intonation at sentence level might not transform or modify the lexical tones at the syllable level. Some other scholars, however, proposed a close interaction between lexical tone and intonation (Liu & Xu, 2005; Ma et al., 2006; Yuan, 2011). The first aim of this study was therefore to address this issue by performing pitch analyses of imitative syllables in tone-language-speaking children, to explore whether the overall prosodic pitch pattern at the syllable level was more variable in the clinical population of ASD.

More importantly, going beyond the basic prosodic features, lexical tone constitutes one of speech elements to distinguish lexical meanings, with the same kind of effect as vowel or consonant. For example, in Mandarin, the syllable /ma/ with high-level tone means 妈 “mother”, and the same syllable with dipping tone means 马 “horse”. The mispronounced pitch contours of lexical tones could lead to comprehension and communication barriers. The Cantonese and Mandarin are two widely-spoken and well-studied tonal languages, with Cantonese tonal system being more complex. In Cantonese tone (CT), there are six citation tones in open syllables contrasting in both pitch height and slope (Jack Gandour, 1981): Three level tones (high-level CT55 vs. mid-level CT33 vs. low-level CT22), two rising tones (low-rising CT23 vs. high-rising CT25), and one falling tone (low-falling CT21). The digits refer to tone transcriptions in Chao’s five-scale tone letters which are an analogy of a musical scale (Chao, 1930), with 5 being the highest and 1 being the lowest ‘relative’ pitch level of a speaker’s normalized pitch change. The three level tones in Cantonese (i.e., CT55–CT33–CT22) differ mainly in pitch height, while the other three contour tones (i.e., CT23–CT25–CT21) differ in both pitch height and direction. For Mandarin tone (MT), there are only four citation tones, each of which carries a distinct pitch contour (Wang, 1973; Yip, 2002): high-level MT55, high-rising MT35, high-falling MT51, and dipping or low-falling-rising MT214 (being realized as high-rising *MT35 at the non-final position when occurred before another dipping tone, and as low-falling *MT21 when the following tone is not a dipping tone). According to a corpus-based comparative study of Mandarin and Cantonese (Peng, 2006), the acoustic distribution of CT25 was very similar to that of MT35 according to tone chart, CT55 similar to MT55, and CT21 similar to *MT21, while there are no direct counterparts for two level CTs (CT33, CT22) and the low-rising CT (CT23) in Mandarin tonal system.

In the current cross-linguistic study, the productions when imitating CT in both Cantonese-speaking (native) and Mandarin-speaking (non-native) children with and without out ASD were analyzed and compared. The imitation task was adopted due to the following reasons. Firstly, imitation is a powerful form of learning commonly used by infants and children. The close relationship between imitation and language learning has been discussed in various studies (Lewis, 1957; Speidel & Nelson, 1989; Whitehurst & Vasta, 1975). During the process of speech acquisition, it is generally believed that imitation from adult models generates the most natural forms for its underlying mechanisms (Messum, 2008). Secondly, speech-language pathologists always make use of imitation in clinical training, and the establishment and reinforcement of imitative responses are regarded as a standard clinical practice (Rees, 1975). Thirdly, vocal imitation might be considered as a simple and primitive tool of communication, and mimicry is something that infants, children with cognitive and attentional deficits are able to do (Speidel & Nelson, 1989). Finally, it is often observed that echolalia, a form of verbal imitation, is one of the most common characteristics of communication in children with ASD (Prizant & Duchan, 1981). Some researchers proposed that autistic children could not accurately imitate the prosodic patterns of the adult stimuli, but rather tended to be monotonous or make an adjustment when they echoed speech (Paccia & Curcio, 1982; Pronovost et al., 1966). On the other hand, others claimed that autistic children were able to imitate the tone of voice and rhythm of other speakers as well as normal children (Frankel et al., 1987). Since the complex tonal system of Cantonese showed fine-grained pitch differences regarding both pitch height and direction, imitation of CT in both native and non-native autistic children offered a valuable chance to check the competence of imitating the subtle pitch changes in children with ASD, and to further illustrate how such imitative performance changes as a function of language experience.

As suggested by cross-linguistic processing models such as the Speech Learning Model (SLM; Flege, 1995, 2007), Perceptual Assimilation Model for suprasegmentals (PAM-S; So & Best, 2010, 2014), and the Similarity Differential Rate Hypothesis (SDRH; Major & Kim, 1996), there is an influence of the L1 phonological system on L2 speech processing, and the outcome of L2 processing can be related to the "cross-linguistic similarity" between an L2 item and its closest L1 counterpart. When imitating the Cantonese tonal models, Mandarin-speaking children would be likely to assimilate the acoustically similar L2 tones of CT25, CT55, and CT21 into the tones of MT35, MT55, and *MT21 respectively in their L1. In contrast, when there is no similarity between an L2 sound and its native L1 sound, the formation of mental representation for a novel and unfamiliar L2 category will occur (Flege, 2007). That's to say, for Mandarin-speaking child subjects, the three L2 lexical tones of CT55, CT25, and CT21 are familiar tonal stimuli, while the other three L2 tones of CT33, CT22, and CT23 were presumed to be unfamiliar for they have no direct acoustic counterparts in Mandarin (Peng, 2006; So & Best, 2010). Besides, the suprasegmental lexical tones are superimposed on the segmental components of each syllable, which might in turn exert an influence on the processing of L2 tones. For instance, for the healthy Mandarin-speaking adults, the native and familiar segments /fu/, /ji/ (existed in both Mandarin and

Cantonese) helped improve discrimination and/or identification accuracy of unfamiliar CT relative to the foreign and unfamiliar segments /si/ and /sɛ/ (only existed in Cantonese; note that Cantonese syllable /si/ is different from Mandarin /sɿ/) (Wang & Peng, 2014), indicating a top-down influence of phonological processing. In this study, the tonal familiarity and segmental familiarity of the speech models would be manipulated to investigate the influence of phonological knowledge on the performance of lexical tone imitation in children with ASD.

To this end, this cross-linguistic study aimed to evaluate the capacity of lexical tone and non-linguistic pitch imitation in Mandarin-speaking and Cantonese-speaking children with ASD. In the speech condition, CTs were adopted as the speech models due to a richer inventory of tonal types than MT (Jack Gandour, 1983; Peng, 2006). Both native and non-native children with and without ASD were recruited to test the influence of different language experience, and the familiarity of the speech materials was designed to differ at either the tonal or the segmental level (see *Materials* in Experiment 1). For the Cantonese-speaking children, they were familiar with all the suprasegmental and segmental components in speech models that were native. For the non-native Mandarin-speaking children, they were asked to imitate Cantonese syllables with familiar and unfamiliar tones, which were superimposed on familiar and unfamiliar segments. Besides, in this imitation study, nonspeech analogues were also generated sharing exactly the same pitch trajectories with the three level tones (CT55, CT33, CT22) and three contour tones (CT23, CT25, CT21) in Cantonese, in an effort to test whether the performance of atypical imitation in ASD was speech-specific or domain-general. The primary research questions addressed are the following:

(1) Would the syllable-level prosodic pitch produced by tone-language-speaking children with ASD show an increased pitch variation compared to TD children?

(2) When imitating the complex pitch contours of Cantonese tones, would native and non-native children with ASD be able to produce normal-like lexical tone productions that are acoustically comparable to those produced by TD peers?

(3) How did the top-down phonological knowledge (segmental familiarity: familiar vs. unfamiliar; tonal familiarity: familiar vs. unfamiliar) influence the lexical tone imitation accuracy in children with and without ASD?

4.2 Experiment 1: Acoustic Analyses of Lexical Tone and Non-Linguistic Pitch Imitation

4.2.1 Methods

4.2.1.1 Participants

In total, 104 child subjects participated in this study and completed all the tests (Table 4.1). Among which, there were 26 Cantonese-speaking children with ASD (CASD, two girls, M_{age} = 7.44 yr), 26 Cantonese-speaking TD children (CTD, one girl, M_{age} = 7.48 yr), 26 Mandarin-speaking children with ASD (MASD, one girl, M_{age} = 7.69 yr), and 26 Mandarin-speaking TD children (MTD, one girl, M_{age} = 7.65 yr). The Cantonese-speaking children with and without ASD were recruited from Hong Kong, and all spoke Cantonese as their first language at home and school with little exposure to Mandarin. The Mandarin-speaking participants were recruited from Shenzhen and used Mandarin as their first language with little exposure to Cantonese. All the child participants had no reported hearing impairment or no comorbidities such as developmental motor speech disorder. Permission to conduct this study was obtained from the Hong Kong Polytechnic University, ensuring appropriate adherence to informed consent procedures.

Table 4.1 *Descriptive characteristics of study samples in Experiment 1.*

(a)	CASD (n = 26)		CTD (n = 26)		t	p
	M (SD)	Range	M (SD)	Range		
CA in years	7.44 (1.28)	6;0–10;4	7.48 (1.22)	6;0–10;0	-0.14	.891
Language score	37.53 (10.52)	14–54	43.10 (7.75)	25–52	-2.17	.035*
Nonverbal IQ	104.50 (22.81)	70–143	107.96 (19.21)	80–140	-0.59	.557
(b)	MASD (n = 26)		MTD (n = 26)		t	p
	M (SD)	Range	M (SD)	Range		
CA in years	7.69 (1.45)	6;0–11;7	7.65 (1.08)	6;6–10;0	0.09	.928
Verbal IQ	97.95 (17.06)	71–125	106.23 (10.99)	82–130	-2.07	.045*
Nonverbal IQ	105.96 (14.60)	79–129	108.62 (12.06)	87–128	-0.72	.478

Note: CASD, Cantonese-speaking children with ASD; CTD, Cantonese-speaking TD children, MASD, Mandarin-speaking children with ASD; MTD, Mandarin-speaking TD children; CA, chronological age; *M*, mean; *SD*, standard deviation; *p < .05.

The clinical diagnosis of ASD was established according to the DSM-5 (American Psychiatric Association, 2013), and the ADOS-2 (Lord et al., 2012) by pediatricians and child psychiatrists with expertise in diagnosing ASD in local hospitals before enrollment. The autistic participants were high-functioning 6- to 11-year-old children without cognitive delays (nonverbal intelligence quotient (IQ) ⩾70; mean = 105.23) and without severe language delays (i.e., able to use full, complex sentences). The school-age children were chosen since neuro-typical preschoolers are still fine-tuning their control over coordinating pitch range, slope, and curvature in the production of native tonal categories (Peggy et al., 2019, 2020; Rattanasone et al., 2018; Wong, 2013). The obtained language score in Cantonese-speaking participants was averaged across with three subtests (Textual Comprehension Test, Expressive Nominal Vocabulary Test, and Test of Hong Kong Cantonese Grammar) in the *Hong Kong Cantonese Oral Language Assessment Scale* (T'sou et al., 2006); The nonverbal IQ in Cantonese-speaking participants was evaluated with *The Primary Test of Nonverbal Intelligence* (Ehrler & McGhee, 2008). The nonverbal and verbal IQ in Mandarin-speaking subjects were assessed with the *Wechsler Intelligence Scale for Children* (WISC-IV, Mandarin Chinese version) (Wechsler, 2003). As shown in Table 4.1, both Mandarin-speaking and Cantonese-speaking ASD group did not differ from corresponding TD group in terms of chronological age and nonverbal IQ, while slightly lagged behind TD children in the general language functioning ($p < .05$).

4.2.1.2 Materials

Table 4.2 *The speech models consisted of 24 Cantonese syllables.*

		Familiar Segment		Unfamiliar Segment	
		fu/fu/	ji/ji/	se/sɛ/	si/si/
Familiar Tone	**CT55**	夫	醫	些	詩
	CT25	苦	倚	寫	史
	CT21	扶	兒	蛇	時
Unfamiliar Tone	**CT33**	富	意	卸	嗜
	CT22	負	二	射	事
	CT23	婦	耳	社	市

Note: The segments in the second row were transcribed in Jyutping (Linguistic Society of Hong Kong [LSHK], 2002), with corresponding international phonetic symbols enclosed by backslashes. The bold strings in the second column indicated three level tones in Cantonese. The written forms of 24 Cantonese syllables were presented with traditional Chinese characters. Specifically, the familiar and unfamiliar distinctions in the table were specific to the non-native Mandarin-speaking participants.

The model stimuli contained 24 Cantonese syllables with three level tones (CT55, CT33, CT22) and three contour tones (CT23, CT25, CT21) superimposed on four Cantonese segments (fu, ji, se, si). Specifically, the tonal and segmental components of Cantonese syllables included both familiar and unfamiliar ones for the non-native Mandarin-speaking children (Table 4.2). These 24 Cantonese syllables were firstly recorded 10 times in a natural way from 10 Cantonese adult speakers (five females; five males) who were born and raised in Hong Kong. We picked out the speech samples spoken by two representative speakers (one female voice and one male voice) whose tonal productions were closer to the median of pitch height and slope among the 10 native speakers. Then, one speech sample (with the best voice quality) for each syllable was chosen from 10 repetitions by a phonetically trained native speaker based on the clarity and stability. Altogether, 48 speech stimuli (6 tones × 4 segments × 2 voice genders) recorded from one male voice and one female voice were selected as the speech models. As shown in Figure 4.1, the six tonal categories of speech models deviated from each other in the two dimensions of pitch height and pitch slope without overlapping in the tone chart. Moreover, the speech models were double checked by a Cantonese-speaking linguist to ensure they showed no perceived tone merge. The mean F0 (*SD*) for the speech models of female voice and male voice was 252 (52) Hz and 121 (23) Hz respectively. Finally, to generate the non-linguistic/nonspeech pitch models, F0 trajectories of the zero-onset syllable ji/ji/ (6 tones × 2 voice genders) were extracted to synthesize nonspeech models using equal-amplitude triangle waves, which have a different harmonic structure from that of speech sounds (Chen & Peng, 2016). The mean F0 (*SD*) for the nonspeech models of female voice and male voice was 250 (45) Hz and 119 (24) Hz respectively. The duration of both speech and nonspeech models was not normalized to make them sound more natural. Since the nonspeech stimuli sound lower perceptually when compared to the speech stimuli of the same intensity, for the purpose of matching the loudness level, the average intensity level of nonspeech models was set to 80 dB SPL, 15 dB higher than the speech models.

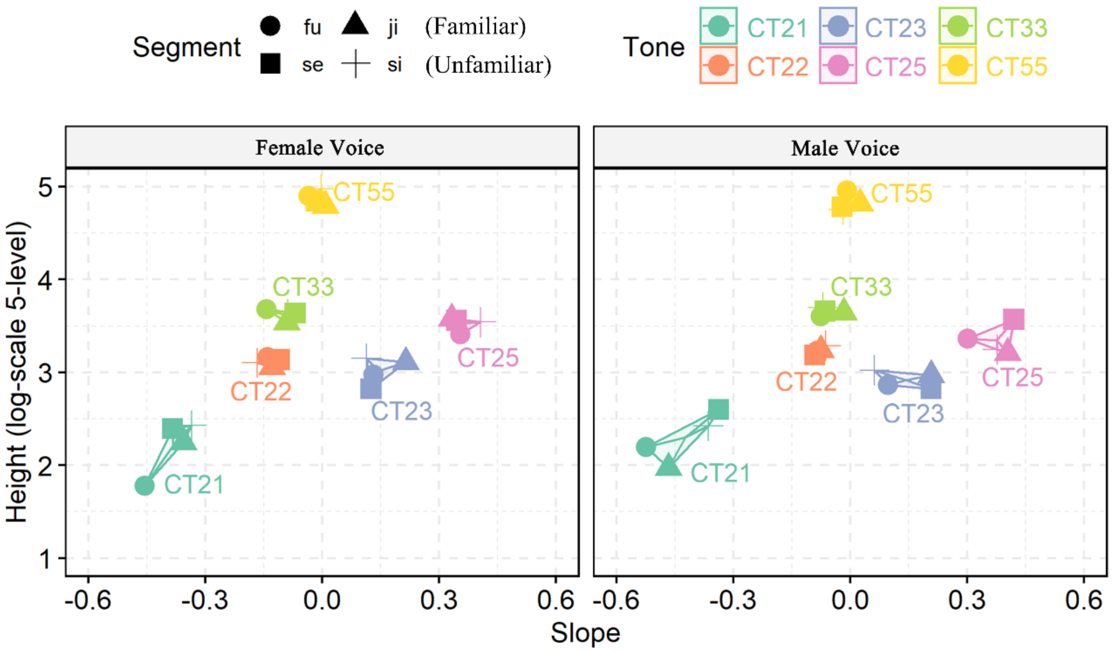


Figure 4.1 Two-dimensional tone charts for the speech models of female voice (left) and male voice (right). Height and slope represent the pitch height and pitch slope respectively. Six Cantonese tones are represented by various colors and four segments by different shapes, as shown in the legends.

4.2.1.3 Procedure

The experimenters were native speakers in each language background. The participants were firstly tested with their nonverbal IQ, verbal IQ/language ability (Table 4.1). During these cognitive and language tests, all the child participants showed no difficulties in understanding the verbal instructions, indicating that they were not deficient in perceiving the native speech sounds in connected and natural speech. For the imitation task, the stimulus presentation was implemented with E-Prime 2.0 program (Psychology Software Tools Inc., USA). The speech/nonspeech models were played in sound-attenuated rooms via bilateral loudspeakers (JBL CM220) located at 45 degrees to the left and right of the participant at a distance of approximately 1 meter. Before the formal experiment, a training session was provided by asking participants to imitate both the speech stimuli (Cantonese syllable ‘fan’ with six CT) and the nonspeech counterparts as accurately as possible, in order to familiarize children with the procedure and requirements. Training materials were produced by a new female talker, which were not involved in the formal test stimuli. The imitative productions in the training session were not recorded, while all the productions in the formal test were recorded using an external microphone (SHURE MV51) around 10 cm away from the mouth of the participant. The microphone was connected to a laptop computer through a USB audio interface with a sampling rate of 44,000 Hz. In the formal test, there were two testing blocks (speech block and nonspeech block) which were presented in a random order among participants. Within each testing block, the speech or nonspeech model stimuli were repeated three times and played randomly. The whole imitation task, including training and testing, lasted about 30 minutes for each participant.

4.2.1.4 Coding and Measurements

The recorded tokens of imitating speech/nonspeech models were coded offline in Praat (Boersma & Weenink, 2016). Acoustic measurement of F0 was derived through the automatic F0 tracking performed using ProsodyPro (Xu, 2013) on 20 equidistant points along the pitch contour. These F0 points were further checked and manually corrected for any “pitch halving” or “pitch doubling” errors. Then, 10% of both leftmost and rightmost F0 points were discarded, and only the middle 80% (i.e., 16 points) for each pitch trajectory were analyzed to decrease the tone-irrelevant variation (Peng, 2006). Besides, the intensity and duration values of each imitative token produced by child participants were also measured and acted as covariates in the statistical analyses of acoustic measures.

For the analyses of prosodic pitch pattern, three measures – pitch mean, pitch range, and pitch SD – were analyzed in the current study. The raw F0 values (in Hz) were transformed into semitones, with a reference frequency of 100 Hz (Rattanasone et al., 2018). In particular, the pitch range was calculated as the minimum F0 subtracted from maximum F0. These three pitch measures (in semitone) were calculated for each participant regardless of different tonal categories. The pitch mean was used to provide a general characterization of prosodic pitch (high vs. low); both pitch range and pitch SD were used to depict pitch variations. An expanded pitch range and/or larger pitch SD implied more pitch variations, and vice versa (Diehl et al., 2009).

For the analyses of lexical tone production, the pitch of each tonal category was transformed from Hz to log-scale 5-level values (Peng & Wang, 2005), consistent with the time-honored selection of number of levels for linguists to transcribe lexical tones (Chao, 1930). The log-scale 5-level value was adopted since it calculates each speaker's normalized linguistic pitch distribution and eliminates inter-speaker variations in absolute pitch differences. Furthermore, to better analyze the fine-grained and dynamic pitch changes of lexical tones which are nonlinear in nature, the second-order orthogonal polynomial models were adopted which is a multilevel regression technique designed for analyzing time course data (Mirman, 2014; Rattanasone et al., 2018; Tang et al., 2019). According to Mirman (2014), the polynomial function generates three 'time terms': the intercept term (i.e., pitch height), the first-order linear term (i.e., pitch slope), and the second-order quadratic term (i.e., pitch curvature). These three terms capture not only the height and slope of pitch contours, but also the steepness of the quadratic curvature. More specifically, the positive linear trend means a rising pitch contour, whereas negative means a falling contour; a larger absolute value of the linear trend represents a steeper slope. The positive quadratic trend indicates a concave F0 contour and negative indicates a convex contour, with a larger absolute value of quadratic trend indicating more curvy contours. In addition, for the acoustic comparison of two rising tonal pair in Cantonese (CT23 vs. CT25), they additionally showed a covert contrast at the temporal distinction of ‘inflection point’: The minimum F0 value appears slightly earlier in the high-rising CT25 compared to that in low-rising CT23 along the pitch contour (Mok et al., 2020). Positions of the inflection point were obtained by locating the lowest F0 point in the first two thirds of the rising pitch contour.

4.2.1.5 Statistical Analysis

First, the linear mixed-effect models (LMMs) in R (R Core Team, 2014) were used to analyzed three measures of prosodic pitch pattern. The package of lme4 (Bates et al., 2014) was used to create the LMMs. An advantage of LMM is that it is possible to fit models to large, unbalanced data, such as the production data by children with and without ASD. The visual inspection of Q-Q plots and plots of residuals revealed no obvious deviations from homoskedasticity after exclusion of extreme data by a model-based trimming. In each LMM of prosodic pitch analyses, the pitch mean/range/SD (in semitone) in speech/nonspeech condition was entered as the dependent variable, with *group* (ASD, TD), *voice gender* (female voice, male voice), *language* (Cantonese, Mandarin), and all possible interactions acting as fixed effects. When fitting the LMMs, the factors of *intensity* and *duration* were involved as controlled covariates, which were centered to reduce multicollinearity; *subject and item* were included as random effects.

Second, the growth curve analysis (Mirman, 2014) in R was adopted to analyze the lexical tone and non-linguistic pitch productions. The pitch contours (in log-scale 5-level value) measured over 16 normalized time points were modeled with a second-order orthogonal polynomial, with fixed effects of *group* (ASD, TD), *language* (Cantonese, Mandarin), and their interaction on all time terms. The model also included *subject* random effects on all time terms (intercept, linear, and quadratic terms). Besides, the centered *intensity* and *duration* were included as covariates. In speech condition, the second-order polynomial models were conducted for each tone (6 tones) and each type of segment (familiar and unfamiliar) separately (12 models in total: 6 for the familiar segment and 6 for the unfamiliar segment). In nonspeech condition, the second-order polynomial models were conducted for each tonal category (6 models in total).

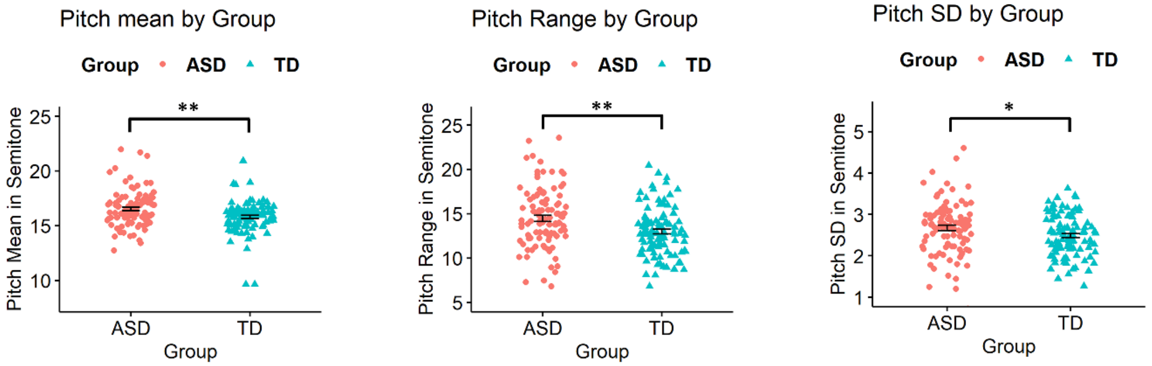
Third, for the analysis of inflection point, the generalized Poisson regression models (Consul & Famoye, 1992) were constructed in R, with *group* (ASD, TD), *language* (Cantonese, Mandarin), *tonal pair* (CT23, CT25), and all the possible interactions acting as fixed effects. The generalized Poisson regression model has been found useful in fitting the dependent variables of integer data. When fitting the regression models in speech/nonspeech condition, *subject and item* were included as random effects.

For all the generated LMMs, polynomial models, and Poisson regression models mentioned above, the random slopes and their intercepts for all the relevant fixed effects were included in the initial model to make it maximally generalizable across the data (Barr et al., 2013). The *p*-values of main effects and interaction effects were obtained using Satterthwaite's approximations in R package lmerTest (Kuznetsova et al., 2017). When a significant main effect of a multilevel factor or a significant interaction effect was detected, post-hoc pairwise comparisons were performed using the lsmeans package (Lenth, 2016) with Tukey adjustment.

4.2.2 Results

4.2.2.1 Prosodic Pitch Pattern

(a) Speech Condition



(b) Nonspeech Condition

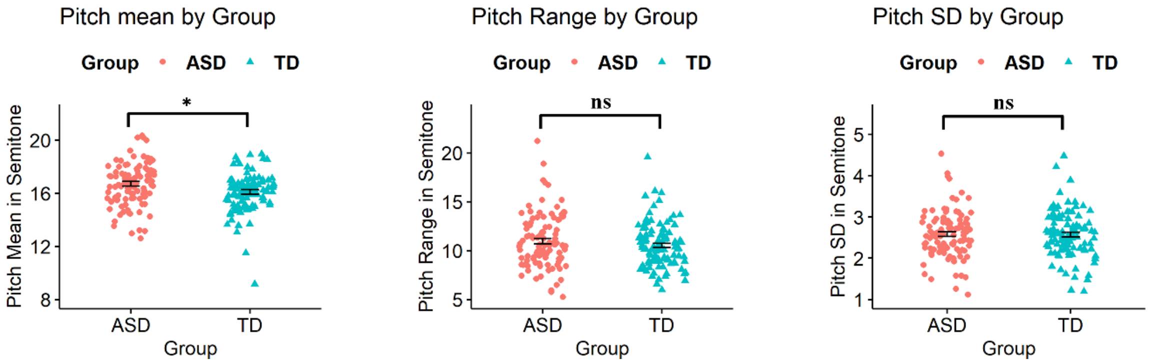


Figure 4.2 The pitch mean (left column), pitch range (middle column), and pitch SD (right column) produced by children with ASD and TD children when imitating models in (a) speech condition, and (b) nonspeech condition. The error bars were presented inside the jitters. $**p < .01$; $*p < .05$; ns: not significant.

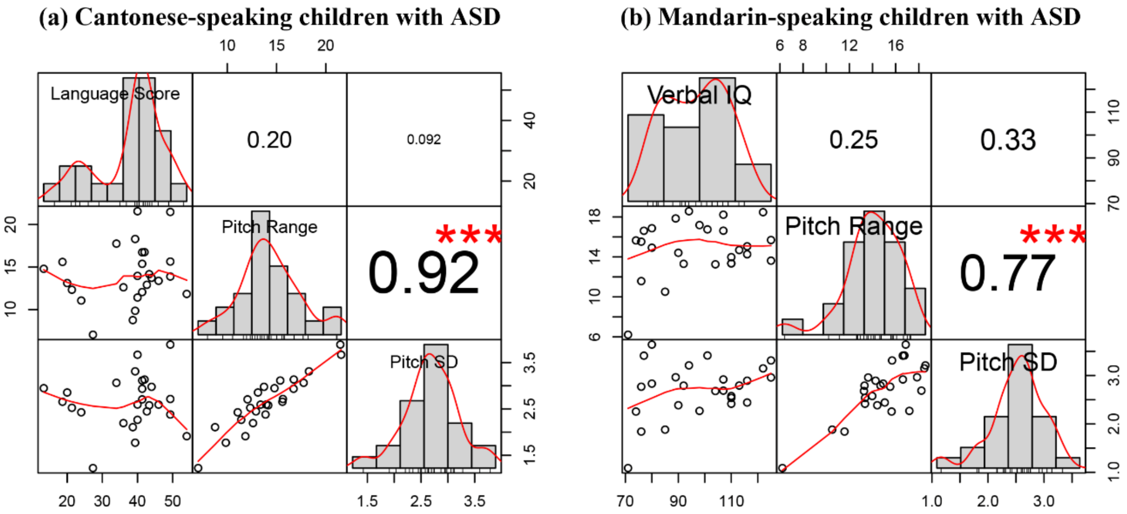
Pitch Mean. In speech condition, the LMM on the pitch mean (in semitone) showed significant main effects of *group* (χ^2 (1) = 9.77, $p < .01$), and *voice gender* (χ^2 (1) = 27.90, $p < .001$), while the main effect of *language* (χ^2 (1) = 0.00, p = .981) and all interaction effects did not reach significance (all *ps* > .05). As shown in the left column of Figure 4.2a, the ASD group regardless of language backgrounds generally demonstrated a higher average pitch (M = 16.5) when imitating the lexical tones compared to the TD children (*M* = 15.8). Moreover, as expected, both children with ASD and TD children enhanced their mean pitch when imitating the speech models of female voice (*M* = 16.5) relative to male voice (*M* = 15.9). Similarly, the LMM on the pitch mean (in semitone) in nonspeech condition also revealed significant main effects of *group* (χ^2 (1) = 4.87, $p < .05$), and *voice gender* (χ^2 (1) = 56.37, $p < .001$). That's to say, as presented in the left column of Figure 4.2b, the child participants with ASD (*M* = 16.7) also tended to exhibit a higher pitch mean at non-linguistic pitch imitation than their neuro-typical peers (*M* = 16.1). Moreover, child participants produced a relatively higher pitch mean when imitating the nonspeech stimuli containing the female pitch contours (M = 17.0) than the male pitch contours (M = 15.8).

Pitch Range. For the pitch range in speech condition, only a main effect of *group* (χ^2 (1) = 8.76, $p < .01$) was found. Neither the main effects of *voice gender* (χ^2 (1) = 0.22, $p = .637$), *language* (χ^2 (1) = 0.39, $p = .531$), nor any two-way and three-way interactions were significant (all *p*s > .05). The obtained pitch range (in semitone) when imitating the lexical tones in the ASD group was 14.5 and in the TD group was 13.0 (in the middle column of Figure 4.2a). The main effect of *group* suggested that both Mandarin-speaking and Cantonese-speaking participants with ASD generally produced a wider pitch range in the production of lexical tones, which might reveal an exaggerated pitch in ASD. For the pitch range in nonspeech condition, all the main effects and interaction effects fell short of significance (all *p*s > .05). As displayed in the middle column of Figure 4.2b, the autistic children (*M* = 11.0) showed comparable pitch range (in semitone) as the TD controls (*M* = 10.6) when imitating the non-linguistic pitch contours in nonspeech condition.

Pitch SD. LMM was performed on pitch SD in speech condition, and the statistical results showed significant main effects of *group* (χ^2 (1) = 4.42, $p < .05$), and *voice gender* (χ^2 (1) = 8.65, $p < .01$), while the main effect of *language* (χ^2 (1) = 0.01, $p = .908$) and all interaction effects did not reach statistical significance (all *p*s > .05). As illustrated in the right column of Figure 4.2a, there was a significant difference in pitch SD (in semitone) of imitating lexical tones between the two groups. Overall, children with ASD (*M* = 2.68) showed a significantly greater SD across F0 samples in speech condition than TD children (*M* = 2.49). Moreover, participants showed a larger pitch SD when imitating the CT models of female voice (*M* = 2.66) than male voice (*M* = 2.50). Next, in the nonspeech condition, only the main effect of *voice gender* was found to be marginally significant (χ^2 (1) = 2.97, *p* = .085). The imitations from the nonspeech models with female pitch contours (*M* = 2.62) showed a trend of receiving relatively higher pitch SD (in semitone) relative to the nonspeech models with male pitch contours (*M* = 2.53). The non-significant main effect of *group* (χ^2 (1) = 0.01, *p* = .940) suggested that, different from imitating the speech tones, autistic children (*M* = 2.58) generated a very similar pitch SD compared to the TD children (*M* = 2.57) when imitating the non-linguistic pitch contours (in the right column of Figure 4.2b).

Given that children with ASD showed greater pitch variations, as indicated by an expanded pitch range and a larger pitch SD, when imitating the lexical tones in speech condition (Figure 4.2a). However, in stark contrast, the imitative pitch range and pitch SD produced by autistic children in nonspeech condition were comparable to those by TD children (Figure 4.2b). In order to further examine whether the pitch variations of imitating speech tones were correlated with the language/verbal abilities in autistic children per se, we conducted Spearman's correlation in Cantonese- and Mandarin-speaking with ASD respectively. For CASD (Figure 4.3a), a very strong positive correlation was found between pitch range and pitch SD (*r* = .92, *p* < .001) as expected. However, the language score of CASD was not correlated with pitch range (*r* = .20, *p* = .330), or pitch SD (*r* = .09, *p* = .655). In a similar manner, there was a positive correlation between pitch range and pitch SD in MASD (*r* = .77, *p* < .001). However, neither the correlation between verbal

IQ and pitch range (*r* = .25, *p* = .223), nor between verbal IQ and pitch SD (*r* = .33, *p* = .104) reached significance in MASD when imitating the speech tones (Figure 4.3b).



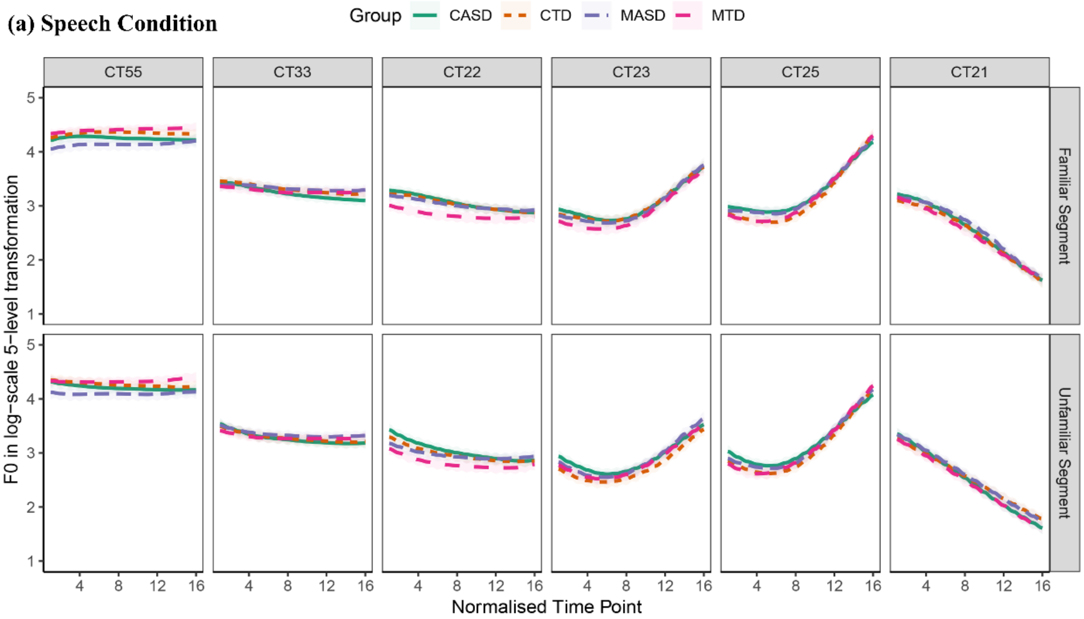
(a) Cantonese-speaking children with ASD

(b) Mandarin-speaking children with ASD

Figure 4.3 The correlations among language/verbal ability, pitch range, and pitch SD when imitating speech tones in (a) Cantonese-speaking children with ASD, and (b) Mandarin-speaking children with ASD. The correlation coefficient *r* was displayed by numbers in the squares, with larger font indicating a larger *r* value.

4.2.2.2 Linguistic Tone and Non-Linguistic Pitch Imitation

(a) Speech Condition



(b) Nonpeech Condition

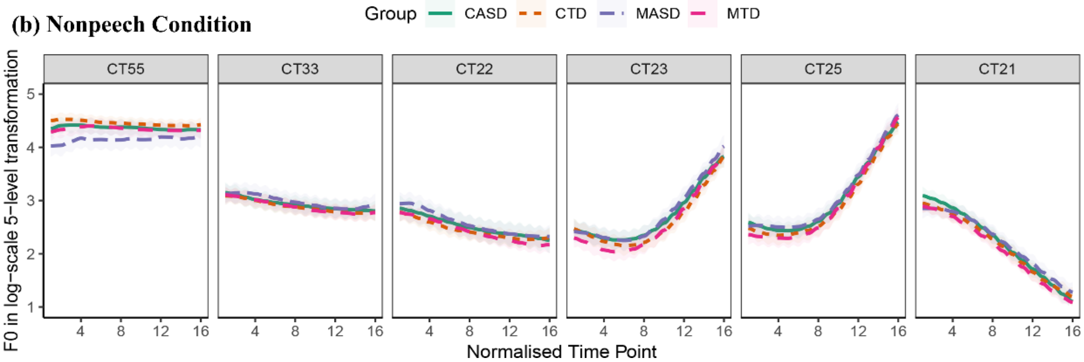


Figure 4.4 Pitch contours of (a) lexical tone and (b) non-linguistic pitch imitations produced by four subgroups (CASD, CTD, MASD, and MTD). The shades with light colors indicate standard error.

Figure 4.4 displayed the pitch contours along 16 time points produced by four subgroups of child participants (CASD, CTD, MASD, MTD) when imitating six CTs and the non-linguistic pitch models. The seemingly overlapping pitch contours across the four subgroups implied that all the child participants could generally produce the global tonal contours (Figure 4.4), consistent with high-level (CT55), mid-level (CT33), low-level (CT22), low-rising (CT23), high-rising (CT25), and low-falling (CT21) descriptions. But if we zoomed in on the fine-grained pitch differences, all the pitch trajectories showed dynamic changes in terms of pitch height, pitch slope, and pitch curvature. For instance, the three level tones (CT55, CT33, CT22) produced by Cantonese native speakers often have a slight falling contour in actual F0 realization (Mok et al., 2020). For the minimal pair of two rising tones (CT23, CT25), there is a dip in the F0 contour in the first half of the tone before the rising contour. In order to examine the fixed effects of *group* (ASD, TD), *language* (Cantonese, Mandarin), and their interaction on pitch height, slope, and curvature, the second-order orthogonal polynomial models were built for each tonal category. In the polynomial model, the intercept term, linear term (ot1), and quadratic term (ot2) capture the F0 contour's pitch height, pitch slope, and pitch curvature, respectively (Tables 4.3 & 4.4).

Level Tones. Table 4.3 shows the statistical results of fixed effects on the pitch height, slope, and curvature when imitating three level CTs. First, for the imitation of high-level CT55 (Table 4.3a), there was only a significant effect of *language* on the linear term (pitch slope) in both speech and nonspeech conditions. Compared to Mandarin-speaking children, the native Cantonese-speaking children tended to produce a relatively more falling F0 slope (Figure 4.4) when imitating the high-level CT55, in the familiar segment ($\beta = -0.13$, SE = 0.05, $t = -2.62$, $p = .01$), unfamiliar segment ($\beta = -0.21$, SE = 0.05, $t = -4.12$, $p < .001$), as well as in nonspeech condition ($\beta = -0.19$, SE = 0.09, $t = -2.14$, $p < .05$). Then, for the imitation of both mid-level CT33 and low-level CT22, the results merely revealed significant main effect of *language* on the linear term (pitch slope) in the speech condition. Specifically, when imitating the mid-level CT33 (Table 4.3b), Cantonese-speaking children showed more falling F0 slope, with significant negative estimates relative to Mandarin-speaking children in the familiar speech segment (β = -0.24, SE = 0.05, t = -4.56, p < .001), and unfamiliar segment (β = -0.20, SE = 0.05, t = -4.09, p < .001). Also, pitch trajectories of imitating low-level CT22 (Table 4.3c) tended to be more falling for Cantonese-speaking children in familiar speech segment (β = -0.21, SE = 0.06, t = -3.37, p = .001), and unfamiliar speech segment (β = -0.28, SE = 0.06, t = -5.04, p < .001). It should be noted that, when imitating the three level tones in both speech and nonspeech conditions, neither the main effect of *group* nor the interaction of *group* × *language* was found on the pitch height, slope, or curvature (Table 4.3).

Table 4.3 *The results of fixed effects on the intercept term, linear term, and quadratic term for each level tone (df=1).*

Level Tones	Time Terms	Fixed Effects	Speech Condition				Nonspeech Condition	
			Familiar Segment		Unfamiliar Segment			
			χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value
(a) CT55	Intercept Term	group	3.32	.068	1.82	.177	2.42	.120
		language	1.10	.294	1.84	.175	3.25	.072
		group:language	3.05	.081	2.28	.131	0.32	.570
	Linear Term	ot1:group	1.03	.310	0.41	.522	2.51	.113
		ot1:language	**6.66**	**.010****	**15.71**	**<.001*****	**4.50**	**.034***
		ot1:group:language	0.76	.385	0.12	.726	0.96	.327
	Quadratic Term	ot2:group	0.69	.408	0.00	.985	0.16	.692
		ot2:language	1.74	.188	0.00	.954	1.24	.266
		ot2:group:language	0.18	.675	0.31	.580	0.25	.616
(b) CT33	Intercept Term	group	0.16	.692	0.40	.525	1.11	.291
		language	0.17	.679	0.00	.998	0.27	.604
		group:language	0.14	.708	0.00	.967	0.09	.767
	Linear Term	ot1:group	1.87	.171	2.65	.103	0.21	.647
		ot1:language	**18.93**	**<.001*****	**15.55**	**<.001*****	0.04	.847
		ot1:group:language	0.77	.380	0.02	.894	0.04	.838
	Quadratic Term	ot2:group	0.43	.510	0.82	.366	0.00	.965
		ot2:language	0.21	.650	0.25	.617	0.00	.990
		ot2:group:language	0.99	.319	0.24	.624	0.00	.944
(c) CT22	Intercept Term	group	1.74	.188	2.33	.127	1.92	.166
		language	2.17	.141	2.77	.096	1.82	.177
		group:language	1.09	.297	0.08	.771	0.67	.413
	Linear Term	ot1:group	2.55	.110	0.55	.459	0.40	.526
		ot1:language	**10.79**	**.001****	**22.77**	**<.001*****	2.15	.142
		ot1:group:language	0.04	.841	3.26	.071	0.16	.692
	Quadratic Term	ot2:group	0.27	.603	0.23	.633	0.00	.973
		ot2:language	2.30	.129	0.05	.828	0.04	.837
		ot2:group:language	1.23	.267	0.53	.465	1.70	.192

***p < .001, **p < .01, *p < .05.

Table 4.4 *The results of fixed effects on the intercept term, linear term, and quadratic term for each contour tone (df=1).*

Contour Tones	Time Terms	Fixed Effects	Speech Condition				Nonspeech Condition	
			Familiar Segment		Unfamiliar Segment			
			χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value
(a) CT23	Intercept Term	group	1.56	.212	1.76	.185	2.74	.098
		language	2.40	.122	0.21	.648	0.05	.826
		group:language	0.01	.936	0.50	.481	0.01	.929
	Linear Term	ot1:group	0.89	.345	0.14	.713	0.07	.794
		ot1:language	2.38	.123	2.29	.130	3.43	.064
		ot1:group:language	0.06	.809	0.78	.376	0.26	.609
	Quadratic Term	ot2:group	1.17	.279	1.36	.244	1.69	.193
		ot2:language	0.03	.857	0.19	.662	0.02	.890
		ot2:group:language	0.20	.654	0.11	.746	0.24	.624
(b) CT25	Intercept Term	group	2.28	.131	2.16	.141	2.87	.090
		language	0.12	.732	0.66	.418	0.29	.588
		group:language	0.40	.526	0.23	.630	0.34	.558
	Linear Term	ot1:group	**5.68**	**.017***	**3.91**	**.048***	0.46	.497
		ot1:language	0.60	.440	2.73	.098	1.19	.275
		ot1:group:language	0.11	.744	0.03	.872	0.27	.604
	Quadratic Term	ot2:group	**4.00**	**.046***	1.00	.317	0.21	.645
		ot2:language	1.04	.308	0.12	.728	0.02	.902
		ot2:group:language	1.52	.218	1.73	.188	0.01	.911
(c) CT21	Intercept Term	group	1.84	.175	0.56	.453	2.66	.103
		language	0.17	.682	0.37	.541	2.44	.118
		group:language	0.30	.584	0.32	.571	0.00	.957
	Linear Term	ot1:group	0.84	.359	1.82	.177	0.01	.912
		ot1:language	0.41	.523	0.01	.919	1.34	.247
		ot1:group:language	0.23	.634	3.05	.081	2.62	.106
	Quadratic Term	ot2:group	2.23	.135	0.01	.905	1.00	.318
		ot2:language	0.12	.734	0.04	.843	0.31	.581
		ot2:group:language	1.50	.221	0.17	.679	0.29	.593

*$p < .05$.

Contour Tones. Table 4.4 shows the statistical results of fixed effects on the pitch height, slope, and curvature in the productions of three contour tones. When imitating the low-rising CT23 (Table 4.4a) or low-falling CT21 (Table 4.4c), none of the fixed effects on the time terms reached significance in both speech and nonspeech conditions. The results for high-rising CT25 showed a significant effect of *group* on the linear term (pitch slope) only in speech condition (Table 4.4b). That is, when imitating the high-rising CT25, both Mandarin- and Cantonese-speaking children with ASD produced rising contours with shallower slopes than age-matched

TD children in familiar segment (β = -0.29, SE = 0.12, t = -2.42, $p < .05$), and unfamiliar segment (β = -0.24, SE = 0.12, t = -2.01, $p < .05$). In addition, there was a significant negative effect of *group* on the quadratic term (pitch curvature) for children with ASD, suggesting that they produced a flatter F0 curve than TD children when imitating the high-rising CT25 in familiar segment (β = -0.13, SE = 0.07, t = -2.02, $p < .05$). All other fixed effects were not significant (see Table 4.4 for full results).

The Inflection Point of CT23 vs. CT25. Additionally, we compared the inflection points of the rising minimal pair (low-rising CT23 vs. high-rising CT25) using the generalized Poisson regression model, with *group*, *language*, *tonal pair* (CT23 vs. CT25), and all the possible interactions acting as fixed effects. In speech condition (Figure 4.5a), the regression model on inflection point showed a significant main effect of *tonal pair* (χ^2 (1) = 52.49, $p < .001$), while the other main effects and interaction effects did not reach significance (all *p*s > .05). All the child participants, regardless of language background and clinical condition, produced an earlier inflection position when imitating high-rising CT25 (M = 4.52) than the low-rising CT23 (M = 5.41) in speech condition. Similarly, in nonspeech condition (Figure 4.5b), only the significant main effect of *tonal pair* ($\chi2$ (1) = 12.41, $p < .001$) was found, with an earlier inflection position when imitating the nonspeech CT25 (M = 4.23) than CT23 (M = 5.10) for all the child participants. It should be noted that when imitating the two rising CTs in both speech and nonspeech conditions, neither the main effect of *group* nor its interaction effect on the inflection position was found to be significant (all *p*s > .05).

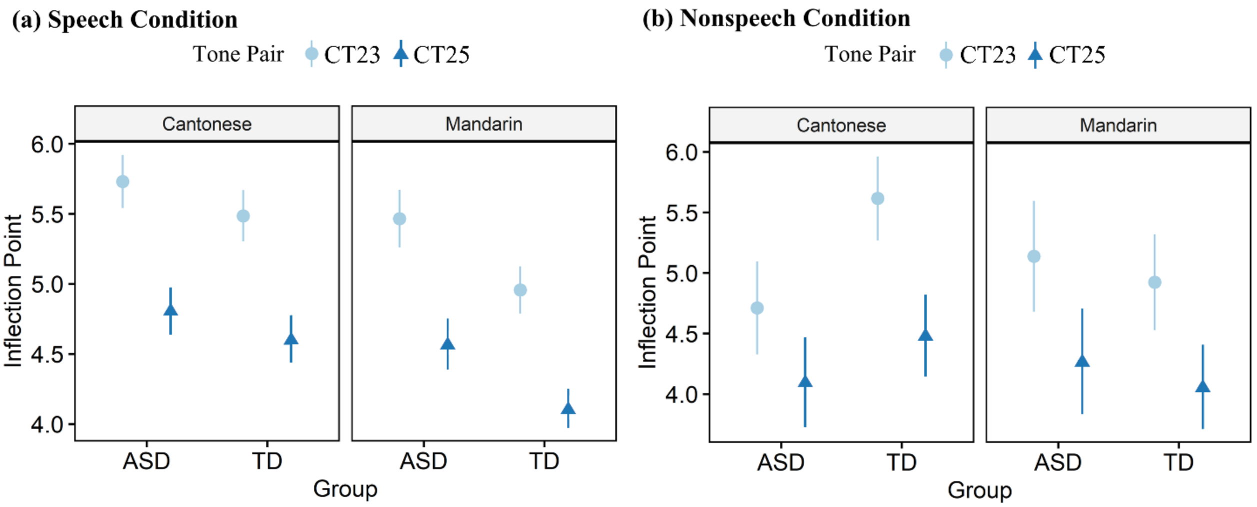


Figure 4.5 Inflection points of the imitative low-rising CT23 vs. high-rising CT25 in (a) speech condition, and (b) nonspeech condition.

4.3 Experiment 2: Identification of Low-Rising CT23 vs. High-Rising CT25

As shown in the acoustic of analyses of linguistic tone and non-linguistic pitch imitation in Experiment 1, the group difference (ASD vs. TD) was only detected during the imitation of high-rising CT25 in speech condition (Table 4.4b). Specifically, both Mandarin- and Cantonese-speaking children with ASD produced a shallower pitch slope and a flatter F0 curve in the imitation of high-rising CT25 relative to TD children. The minimal pair of two rising tones in Cantonese phonology mainly differ in terms of pitch slope, with high-rising CT25 showing a much steeper slope compared to the low-rising CT23 (Figure 4.1). However, given the categorical perception nature in perceiving native speech sounds (Liberman et al., 1957; Xu et al., 2006), the shallower pitch slope of the imitative high-rising CT25 produced by children with ASD does not necessarily entail identification difficulties for native speakers. To shed light on this issue, we further conducted an identification test (Experiment 2), by asking the native Cantonese-speaking adults to perceive and identify the minimal-pair tokens of two rising tones (CT23 vs. CT25), which were produced by both children with ASD and TD children from different language backgrounds. The perceptual analysis in Experiment 2 was performed to complement acoustic measurements in Experiment 1.

Based on the acoustic analyses of F0 contours, we predicted that native Cantonese-speaking adults might show a similar identification accuracy of the low-rising CT23 spoken by children with ASD and TD children, whereas a higher identification accuracy of the high-rising CT25 imitated by TD children relative to that imitated by autistic children. Alternatively, if the acoustic atypicality of a shallower pitch slope produced by children with ASD did not cause perceptual difficulties, the accuracy would be similar for the identification of CT25 produced by different groups (TD children and children with ASD).

4.3.1 Methods

4.3.1.1 Participants

In total, 16 neuro-typical undergraduate and graduate students in college (8 males; mean age = 24.6 years, SD = 2.9) whose first language is Cantonese participated in the identification test. They were not majoring in linguistics or psychology, and had no reported speech, language, or hearing disorders. None of the participants had received formal musical training over one year. All participants gave informed consent in compliance with the protocols approved by the Research Ethics Committee of Hong Kong Polytechnic University, and they were paid for their travel and time.

4.3.1.2 Stimuli and Procedure

Totally there were 1,664 syllables produced by all the child participants (CASD, CTD, MASD, and MTD) through imitating the speech models of CT25 and CT23, and these imitative syllables were included as the perceptual stimuli. The stimuli were not normalized in terms of intensity and duration, in an effort to keep these perceived sounds unmodified. Instead, the duration and intensity values were included as covariates in the statistical analysis to control for confounding factors. The stimuli were presented using E-prime 2.0, and were divided into four testing blocks based on four different carrying segments (fu, ji, se, si). The four testing blocks were counterbalanced across subjects. The perceptual stimuli were played in a random order within each testing block. Before the formal test, there was a practice block with the adult speech models of CT25 and CT23 included as the practice stimuli to familiarize subjects with the identification procedure. The subjects were asked to conduct a two-alternative forced choice (2AFC) identification task. After the presentation of each syllable, they would be asked to identify the target syllable as Cantonese character "婦" (CT23) or "苦" (CT25) in the block of "fu"; as "耳" (CT23) or "倚" (CT25) in the block of "ji"; as "社" (CT23) or "寫" (CT25) in the block of "se"; as "市" (CT23) or "史" (CT25) in the block of "si" by pressing corresponding keyboard buttons. The subjects were allowed to play the target syllable repeatedly till they were confident to make a judgement. The whole identification test, including the practice block, lasted approximately 1.5 hours for each subject.

4.3.1.3 Statistical Analysis

To analyze the identification accuracy, the generalized linear mixed-effects model (GLMM) was created in R using the lme4 package (Bates et al., 2014). For the construction of GLMM, the dichotomous response to each stimulus ("1" meaning correct response or "0" indicating incorrect response) was entered as the dependent measure, with *group* (ASD, TD), *language* (Cantonese, Mandarin), *segment* (familiar, unfamiliar), *tonal pair* (CT23, CT25), and all their possible interactions acting as fixed effects. Moreover, the centered duration and intensity values for each stimulus were included as the controlled covariates. When fitting GLMM, *subject* and *item* were included as random effects. A full structure of random effects was included in the initial model (Barr et al., 2013), which was compared with a simplified model that excluded a specific fixed factor using the ANOVA function in lmerTest package (Kuznetsova et al., 2017). When there was a significant interaction, post-hoc pairwise comparisons were performed using the lsmeans package (Lenth, 2016) with Tukey adjustment.

4.3.2 Results

Figure 4.6 showed box plots of the identification accuracy (%) across different conditions. The GLMM on identification accuracy revealed a significant three-way interaction of *group* × *language* × *tonal pair* (χ^2 (1) = 13.40, p < .001), as well as a significant three-way interaction of *group* × *segment* × *tonal pair* (χ^2 (1) = 8.32, p < .01). First, post-hoc pairwise comparisons for the interaction of *group* × *language* × *tonal pair* showed that native Cantonese speakers' identification accuracy was similar in the perception of high-rising CT25 stimuli produced by CASD and by CTD (β = -0.12, SE = 0.07, t = -1.65, p = .099), and those produced by MASD and by MTD (β = -0.01, SE = 0.07, t = -0.12, p = .904). Moreover, the identification accuracy was similar in the perception of low-rising CT23 stimuli produced by CASD and by CTD (β = -0.02, SE = 0.07, t = -0.25, p = .804), whereas the accuracy was much higher in the perception of CT23 stimuli produced by the non-native MTD than those produced by non-native MASD (β = -0.28, SE = 0.07, t = -3.89, p < .001). Then, post-hoc pairwise comparisons for the three-way interaction of *group* × *segment* × *tonal pair* showed that, native Cantonese speakers' identification accuracy in the perception of

CT23 stimuli produced by TD children was significantly higher compared to those produced by children with ASD when the carrying segment was familiar (β = -0.19, SE = 0.08, *t* = -2.38, *p* < .05), but similar when the carrying segment was unfamiliar (β = -0.10, SE = 0.08, *t* = -1.22, *p* = .224). These findings collectively indicated that the Cantonese-speaking adults showed a much higher identification accuracy in the perception of the low-rising CT23 stimuli with familiar segment which were imitated by MTD than those imitated by MASD. However, the accuracy did not differ towards the identification of high-rising CT25 stimuli when imitated by different groups (TD vs. ASD) from different language backgrounds (Figure 4.6).

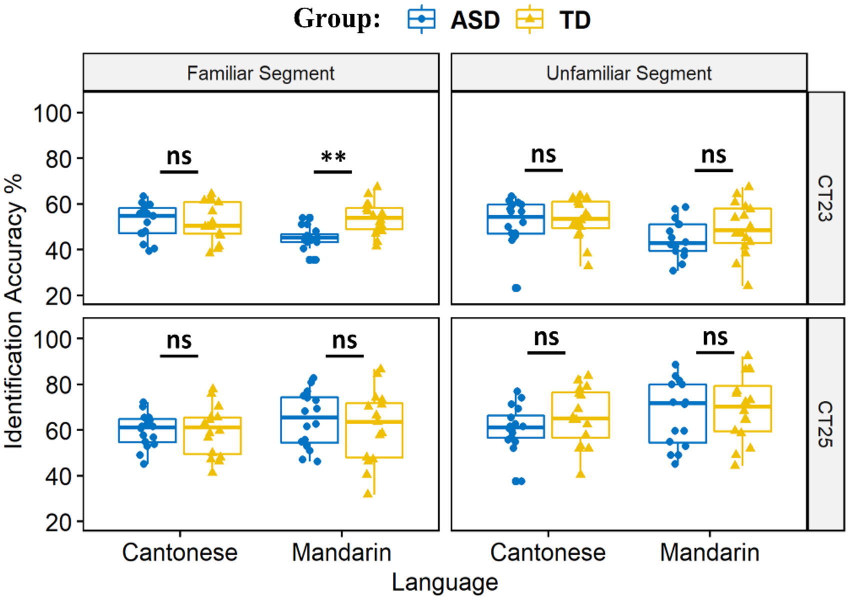


Figure 4.6 The identification accuracy (%) from Cantonese-speaking young adults in their perceptual judgements of the imitative CT23 vs. CT25 with the familiar and unfamiliar segments produced by CASD, CTD, MASD, and MTD. The bold line inside the boxes marks the median of identification accuracy, and the upper and lower boundaries of the box mark its upper and lower quartiles of accuracy. ***p* < .01; ns: not significant.

4.4 Discussion

The abnormalities of prosodic pitch productions have been noted since the earliest report of ASD (Kanner, 1943), but our full understandings of the language-specific features and the underlying mechanisms are currently inconclusive. Importantly, the changes in pitch also play a crucial role in distinguishing phonological contrasts and word meanings at the syllable level for tone language speakers. This study adopted an imitation task to investigate the prosodic pitch pattern and lexical tone production in tone-language-speaking children with ASD. Imitation from an adult model was widely used in clinical training, which was believed to be facilitative for the acquisition of speech and language (Ingersoll, 2008; Messum, 2008). The Cantonese lexical tones were chosen as speech models given their complex inventories with three level tones, and three contour tones. In the current cross-linguistic study, the pitch productions through imitating CT in both Cantonese-speaking (native) and Mandarin-speaking (non-native) children with ASD were analyzed. The age-matched neuro-typical children in each language background were also included as the reference group. The pitch mean, pitch range, and pitch SD (in semitone) were used to depict the variety of the prosodic F0 pattern at the syllable-level. Moreover, the relative F0 contour changes (in log-scale 5-level values) in terms of pitch height, slope, and curvature were used to analyze the fine-grained acoustic realizations of imitative lexical tones and non-linguistic pitch counterparts. Besides, tonal familiarity and segmental familiarity were major variables being manipulated to tap the phonological processing capacity in children with ASD. The major findings and relevant discussions were shown in the following parts.

4.4.1 Atypical Prosodic Pitch Pattern in Tone-Language-Speaking Children with ASD

The prosodic pitch pattern was investigated with pitch mean, pitch range, and pitch SD in the imitation of speech syllables and nonspeech sounds. When imitating speech models, both Cantonese- and Mandarin-speaking children with ASD showed a higher pitch mean, a larger pitch range, as well as a greater pitch SD than peers with TD (Figure 4.2a). When imitating nonspeech models, autistic children only produced a higher average pitch compared to the TD participants (Figure 4.2b). The group differences of the increased pitch of intonation in speakers with ASD have been found in some studies (Chan & To, 2016; Edelson et al., 2007; Sharda et al., 2010) but not others (Diehl et al., 2009; Nadig & Shaw, 2012). Most of the previous studies employing acoustic measurements focused on the pitch range and/or pitch SD of intonation at the sentence level. Contrary to the traditional stereotype of monotonic intonation in autism, the autistic children generally showed a significantly larger pitch range and/or pitch SD compared to TD children, indicating increased pitch variations of intonation in ASD group (Bonneh et al., 2011; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010). These previous studies only compared the F0 differences between ASD and TD groups, while neglected the influence from other prosodic features such as intensity and duration. Actually, the pitch-related parameters almost always involve concomitant variations in other prosodic features (Xu & Prom-on, 2019). After controlling for intensity and duration, the current findings corroborated the notion of increased pitch variations in ASD with the empirical evidence from a smaller prosodic unit at the syllable level. Especially, for individuals with autism who speak a tone language, the atypical prosodic feature of increased F0 variations emerged broadly, not only on the larger prosodic unit of intonation (Chan & To, 2016), but also on the smaller unit of lexical tone (this study).

The conclusion of increased F0 variation as a prominent feature of prosodic pitch pattern in ASD could be reached with high reliability and generalizability, since the same pattern was found consistently in various studies from low-functioning (Baltaxe, 1984; Fosnot & Jun, 1999) and high-functioning subgroups (this study; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Green & Tobin, 2009; Nadig & Shaw, 2012); from tonal (this study; Chan & To, 2016) and non-tonal (Bonneh et al., 2011; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010) language backgrounds; from a wide age ranges in children (this study; Bonneh et al., 2011; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Nadig & Shaw, 2012; Sharda et al., 2010), adolescents (Diehl et al., 2009), and even adults (Chan & To, 2016); from analyses in both spontaneous (Bonneh et al., 2011; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Green & Tobin, 2009; Nadig & Shaw, 2012; Sharda et al., 2010) and imitative speech samples (this study; Fosnot & Jun, 1999; Hubbard & Trauner, 2007). There was a concern about the influence of overall language/cognitive functioning on the prosodic abnormalities, since children with specific language impairment (Goffman, 1999; Marshall et al., 2009) or mental retardation (Shriberg & Widder, 1990) also revealed prosodic deficits. In this study, the ASD and TD groups were matched in terms of nonverbal IQ, although the general language functioning in ASD slightly lagged behind TD children. We have additionally performed correlation analysis between the general language functioning and pitch range/SD in CASD and MASD respectively (Figure 4.3), while no significant correlations were found from our study samples. Also, even in studies with matched comparison groups on variables of both IQ and general language functioning (Diehl et al., 2009; Nadig & Shaw, 2012), the group difference of increased pitch variation in speakers with ASD has been observed as well. Thus, the prosodic pitch differences produced by ASD and TD children tended to be specific to prosody, rather than an artifact of more general language and/or intellectual functioning. Such findings highlight the presence of prosodic pitch atypicalities even in very high-functioning and linguistically developed individuals with ASD, which could be a stigmatizing barrier to communication competence and social acceptance for speakers with ASD who evidence prosodic oddities.

What are the underlying mechanisms responsible for the atypical prosodic pitch pattern in individuals with ASD confirmed now across multiple studies? Some proposed that more variable prosody may be caused by a delayed developmental trajectory of speech in ASD (Sharda et al., 2010). The observations of increased pitch range/SD and pitch mean could also be discovered in speech directed to infants commonly known as ‘motherese’, which has distinct prosodic patterns characterized by heightened pitch and exaggerated pitch contours (Segal & Newman, 2015). Early intonation features of younger TD children under 2 years also mimic motherese-like features, but diminished gradually after 2–3 years of age (Eguchi, 1969). Thus, the increase in pitch variability in speakers with ASD might reflect prolonged mimicry of the prosodic pitch patterns of child-directed speech in this group, relative to TD children (Sharda et al., 2010). Some others labeled the atypical prosodic pitch pattern in ASD as aberrant rather than delayed development of speech prosody in ASD (Rapin & Dunn, 2003). This perspective could be supported by observing such an atypical pattern in adolescents and even in adults with ASD (Chan & To, 2016; Diehl et al., 2009). A possible explanation for the increased pitch variability in the ASD group is a disruption in basic speech production mechanisms that control pitch, which could stem from a deficit at the production level (Bonneh et al., 2011), or reflect speech compensation for auditory feedback perturbations to overcome a noisy channel supposed to transmit “efference copy” information

(Houde et al., 2007). More studies are needed to uncover the nature of abnormal supra-segmental aspects of speech production, or prosody in speakers with ASD.

4.4.2 Preserved Lexical Tone Imitation Skill in Cognitively Able Children with ASD

In this study, the complex CTs with changes in both pitch height and slope (Jack Gandour, 1981) were imitated by both Cantonese- and Mandarin-speaking children with and without ASD. As shown in Figure 4.4, all the native and non-native child participants at school age could accurately imitate the global tone contours for the high-level (CT55), mid-level (CT33), low-level (CT22), low-rising (CT23), high-rising (CT25), and low-falling (CT21) pitch trajectories, important for maintaining tone category distinctions. When we zoomed in on the fine-grained differences in pitch height, slope, and curvature between ASD and TD groups, the ASD group from both language backgrounds only produced a shallower slope and/or a flatter F0 curve in the production of high-rising CT25 relative to TD children. However, such fine-grained and within-category acoustic differences did not cause difficulties in native Cantonese speakers' perceptual judgement, as evidenced by a comparable accuracy of identifying the imitative CT25 stimuli produced by ASD and TD children (Figure 4.6). Moreover, even for the imitation of the covert contrast of the inflection points of CT23 vs. CT25, which was reliable acoustic difference not perceivable by naïve speakers (Edwards & Beckman, 2008), all the child subjects, including the non-native MASD, correctly produced an earlier inflection point for CT25 than CT23. Our findings offered strong supports to the notion that the echoed speech of autistic children could imitate complex tonal contours from the adult models accurately in a preserved manner (Frankel et al., 1987). Such imitation skill seemed to be unaffected by the linguistic status of children with ASD, as a lack of interaction effect of *group* × *language* in all the acoustic analyses. That is, even for the non-native Mandarin-speaking children with ASD, they could largely imitate pitch contours of unfamiliar/familiar tones superimposed on unfamiliar/familiar segments as well as native children. It appears to be the case that autistic children adopted a bottom-up mechanism when imitating the pitch contours at the syllable level, and they were largely intact at the processing of local pitch information, as suggested by the enhanced perceptual functioning in autism (K. Mottron & Burack, 2001).

However, we should be prudent in generalizing the current finding of preserved lexical tone imitation skill to each individual on the autistic spectrum especially those with intellectual disabilities or severe language delays. On the one hand, this study adopted a relatively simple task, which asked participants to imitate each lexical tone in isolation at the syllable level. Such a task may be obscuring group differences that would be present in more difficult tasks such as lexical tone imitation in connected speech. On the other hand, the autistic subjects in this study belonged to the cognitively able ones without cognitive impairment/severe language delay. Since there was a strong relationship between immediate imitation skill and language ability (Rogers et al., 2008; Toth et al., 2006), it was unclear whether low-functioning children with ASD with severe language delays would be able to imitate pitch contours of lexical tones that are acoustically comparable to those imitated by TD children.

4.4.3 Speech-Specific and Contour-Biased Lexical Tone Processing Atypicality in ASD

We have observed two speech-specific phenomena from the current imitation data. First, compared to TD peers, children with ASD showed increased pitch range/SD when imitating the speech tones, while exhibited similar pitch range/SD when imitating the nonspeech sounds. Second, autistic children showed some deviations from the TD children in coordinating pitch slope and curvature when imitating the high-rising CT25 superimposed on speech segments, but the two groups did not differ from each other when imitating the same pitch contour of CT25 embedded in nonspeech materials. The speech-specific imitation atypicality in ASD lends further support to the notion that children with ASD showed domain-specific pitch processing difficulties. Specifically, Mandarin-speaking children with ASD showed atypical or impaired processing of lexical tone (Wang et al., 2017; Yu et al., 2015) and intonation (statements vs. questions, Jiang et al., 2015), whereas they showed normal or even enhanced processing of the same pitch information in the domains of music and nonspeech. In line with extant findings in pitch processing, more and more research proposed a speech-specific viewpoint that speech and language learners with autism fail to engage or develop specialized networks for vocal processing and phonetic learning in speech sounds (Haesen et al., 2011; Kujala et al., 2013; Lindell & Hudry, 2013; O'connor, 2012).

In addition, the lexical tones can be divided into two types in general: contour tones and level tones. Contour tones change both pitch height and direction, whereas level tones remain at approximately a steady pitch (Yip, 2002). One recent study by Cheng et al. (2017) investigated the ability to discriminate 'level tones' embedded in real syllable, pseudo-syllable, and nonspeech in Cantonese-speaking individuals with ASD. However, no group differences were found across all three conditions. It is likely that the speech-specific lexical tone processing difficulties in ASD tended to be biased towards the processing of contour tones (high-rising tone vs. high-falling tone, Wang et al., 2017; Yu et al., 2015), while less severe in the processing of level tones (Cheng et al., 2017). However, the conclusion was far from clear in literature as none of the previous studies incorporated both level and contour lexical tones in one single study. In this study, children with and without ASD were asked to imitate both Cantonese level and contour tones. When imitating three level tones (CT55, CT33, CT22), both native and non-native children with ASD produced a comparable pitch height, slope, and curvature relative to TD children (Table 4.3). Only a main effect of *language* on the linear term (pitch slope) of level tones was found, with Cantonese-speaking children eliciting a more falling F0 slope compared to Mandarin-speaking ones. It was reported that Cantonese native speakers tended to produce a slightly falling contour in their actual realization of three level tones, especially for the low-level and mid-level tones (Mok et al., 2020; Zhang et al., 2018). Thus, the cross-linguistic differences in the acoustic realization of level tones could be attributed to the influence of long-term native language experience. Then, when imitating three contour tones (CT23, CT25, CT21), the two groups did not differ in the acoustic realizations of low-falling CT21, and low-rising CT23, whereas differed in terms of pitch slope and curvature of high-rising CT25. Although subsequent identification test proved that such fine-grained acoustic differences did not lead to perceptual ambiguity, the autistic children nevertheless exhibited some difficulties in coordinating exactly the same acoustic pitch trajectory of the more dynamic and fast-changing contour tones with a steeper slope (i.e., CT25 in this study). To conclude, the current findings pointed to a speech-specific and contour-biased lexical tone processing atypicality in people with ASD.

4.4.4 Top-Down Phonological Processing Deficits in Children with ASD

It was proposed that pitch processing capacity was intact or even superior at the bottom-up acoustic processing level but impaired due to a top-down phonological processing deficit in individuals with ASD (Jiang et al., 2015; Wang et al., 2017; Yu, 2018). The hypothesis was that in tonal models tapping the phonological processing abilities of the child participants, comparatively inferior imitation performance could arise from either the lack or the impairment of relevant phonological representations (Kuhl, 2011). As mentioned earlier, the non-native Mandarin-speaking children with ASD in our study could imitate the global tone contours of both familiar and unfamiliar tonal categories acoustically similar to TD children. It seemed that acoustic pitch realizations of lexical tone imitation persisted independently of speech familiarity, and were not influenced by the phonological status of the tonal categories and the linguistic status of the carrying syllables in children with ASD. However, as shown in the auditory perceptual judgment (Figure 4.6), native Cantonese adult speakers showed relatively higher accuracy in the identification of low-rising CT23 stimuli with familiar segment which were produced by MTD than those produced by MASD. Actually, MTD and MASD produced similar F0 realizations of CT23 with familiar segment, which means the different perceptual accuracy was affected by some other factors beyond F0 (the primary correlate of lexical tones). This is not surprising since several secondary cues, such as intensity profile, duration, and voice quality, also play a role in the perception of lexical tones (Zhang et al., 2012). In our statistical model, the duration and intensity have been entered as covariates. Thus, when imitating the non-native and unfamiliar CT23, MTD could utilize the phonological knowledge of native and familiar segment (/fu/ and /ji/) to produce a better voice quality of that syllable. However, in contrast, MASD failed to exploit such top-down phonological knowledge to compensate for the imitation of syllables with non-native tonal category. That is, MASD demonstrated compromised performance relative to MTD, when imitating unfamiliar Cantonese tonal stimuli that were superimposed on familiar segments (/fu/, /ji/), but comparatively normal performance when on unfamiliar segments (/si/, /sɛ/). These findings implied that the lexical tone processing difficulties in speech condition (Wang et al., 2017; Wu et al., 2020; Yu et al., 2015) reported in some children with ASD were caused by a phonological processing deficit rather than the acoustic pitch processing deficit.

4.4.5 Limitations and Future Research

There are several limitations in this study. First, the imitation task adopted in this study was simple which might be reliably performed in younger and low-functioning children with ASD. Future study could test low-functioning preschoolers with ASD who showed cognitive impairment/severe language delay. It would be meaningful to see how the prosodic pitch pattern and lexical tone imitation skill changed among different subgroups of the autistic spectrum and among different age groups in future studies. Second, in this cross-linguistic study, the nonverbal IQ and general language functioning among Cantonese- and Mandarin-speaking participants were evaluated with different testing materials. Our initial concern was to ensure the ASD group to be matched with TD group in terms of nonverbal IQ in children from each language background. But without unified measurement among all the child subjects, we could not yet include these factors as covariates in the statistical models. Unfortunately, there was no standard oral language assessment scale till now applicable to both Cantonese- and Mandarin-speaking children. Third, the evaluation could be strengthened by incorporating not only acoustic measures of imitation but also intelligibility assessment of the imitated sounds as rated by native speakers of Cantonese. Fourth, perceptual ability in child participants might be tested in future studies to investigate whether the perceptual and vocal imitation performance in ASD are closely related, or to some extent distinct since there might be distinct representations used to support speech imitation and perception tasks (Hutchins & Peretz, 2012). Fifth, the native Cantonese speakers' identification accuracy of CT23 vs. CT25 stimuli in Experiment 2 was surprisingly low. On the one hand, we did not control for the possibility that some native Cantonese speakers in Experiment 2 might merge the two rising tones in Cantonese (Fung & Lee, 2019; Mok et al., 2013). On the other hand, in the current identification study, we adopted a blocked-segment design that contained the imitative stimuli produced from different talkers in one single block. Subjects may be struggling to estimate the upper and lower F0 bounds of a particular voice within a block, thus unable to map each rising pitch stimulus to the corresponding tone category with reference to its relative position in that talker’s F0 range. Future identification study could, for example, present stimuli through a blocked-talker design and exclude the native adult participants who merged the two rising tones.

4.5 Conclusion

This study investigated the prosodic pitch pattern and lexical tone imitation skill in tone-language-speaking children with ASD in a cross-linguistic context. We found increased prosodic pitch variations in tone-language-speaking children with ASD, which was highly consistent with previous findings in autistic children from a non-tonal language background. The accumulating evidence of atypical prosodic pitch pattern contributes to the possibility of developing pitch-based measures as one of the biomarkers for early diagnosis of ASD (Bonneh et al., 2011). Moreover, the competence of imitating the complex F0 contours of Cantonese tones remained largely intact in both native and non-native children with ASD, which offers empirical evidence for the use of imitation in speech therapy for lexical tones. The atypical prosodic pitch pattern was detected when imitating speech tones, but not when imitating nonspeech counterparts. Similarly, children with ASD exhibited some difficulties in coordinating exactly the same pitch trajectory of high-rising CT25 superimposed on speech tones, but showed no difficulties on nonspeech sounds, indicating speech-specific pitch processing difficulties in ASD. Finally, our current observations lend further support to the notion that lexical tone processing difficulties in tone-language-speaking children with ASD occurred at the top-down phonological processing level.

Chapter 5: Adapted Melodic Intonation Therapy Facilitates Speech Learning in Tone-Language-Speaking Children with Autism: A Randomized Controlled Study

5.1 Introduction

The goal of this study is to integrate recent advances in speech therapy to facilitate speech sound acquisition in tone-language-speaking children with ASD. ASD is a neurodevelopmental disorder identified by a constellation of early-appearing social communication deficits and restricted, repetitive sensory-motor behaviours (American Psychiatric Association, 2013). A systematic review of epidemiological surveys commissioned by WHO estimated that the global prevalence of ASD was around 1% (Elsabbagh et al., 2012). Given the high incidence of ASD worldwide, it is necessary to improve our understanding of ASD from different cultures and language backgrounds, and to refine our treatment strategies.

It is well recognized that ASD is accompanied by some other difficulties—that are not part of the diagnostic criteria but can nevertheless exert a negative effect on communication and interaction with others (Lord et al., 2018). One such consideration is the developmental abnormalities in regards to delayed onset and development of the spoken language in the second and third year of life (R. Landa, 2007). There are large individual differences in terms of speech development in ASD. While some children with ASD showed a level of phonological development close to that of TD children (Bartak et al., 1975; Kjelgaard & Tager-Flusberg, 2001; Rapin & Dunn, 2003), others showed varying degrees of delayed speech sound development (Boucher, 1976; Rapin et al., 2009; Schoen et al., 2011; Shriberg et al., 2001; Wolk & Brennan, 2013) and exhibited atypical phonological processes (Cleland et al., 2010; Sheinkopf et al., 2000; Wolk et al., 2016; Wolk & Brennan, 2013; Wolk & Giesen, 2000). Specifically, there was a strong correlation between the severity of speech production difficulty and the degree of language deficit in children with ASD (Bartolucci et al., 1976; Schoen et al., 2011; Wolk & Brennan, 2013; Wolk & Giesen, 2000). Estimates vary, with 25%–46% of children with ASD reported as minimally verbal past the age of 5 years (Kasari et al., 2013; Norrelgen et al., 2015; Rose et al., 2016; Tager-Flusberg & Kasari, 2013), meaning that they have a limited repertoire of intelligible speech with a small vocabulary size. Besides, some low-functioning individuals with ASD even remain nonverbal with a complete absence of functional speech, and lack the ability to communicate with others using spoken language (Klinger et al., 2002; Koegel et al., 2009; Turner et al., 2006). While ASD is intrinsically a socially isolating disorder, the nonverbal and minimally verbal children with ASD are further exacerbated by their poor speech development, which adversely affects other aspects of language development and interpersonal communication abilities. Therefore, understanding speech production difficulties in children with ASD is crucial for devising and implementing early treatments, and the inclusion of the phonological component in treatment is recommended (Wolk et al., 2016; Wolk & Brennan, 2013).

Increased speech output is considered a positive prognostic indicator of outcomes for low-functioning children with ASD (Lord et al., 2006). However, the empirical reports on treatment strategies of enhancing speech output remained limited, most of which were byproducts of various didactic, naturalistic or developmental approaches targeted at improving functional communication in autism (see Paul, 2008 for a review). For those nonverbal or minimally verbal children with ASD, the augmentative and alternative communication (AAC) systems were often adopted by encompassing nonspeech means to make requests and interact with others, which included sign language, Voice Output Communication Aids, and Picture Exchange Communication System (Schlosser & Wendt, 2008; Tager-Flusberg & Kasari, 2013). It has been proved to be useful for helping preschool-aged children with limited spoken language skills become more verbal with the assistance of such AAC systems (Romski et al., 2010; Sulzer-Azaroff et al., 2009). Other methods try to recruit the ASD children's attention first and then introduce the verbal target, including the orienting cue (e.g., ‘‘high five’’ gesture, kisses, hugs) (Koegel et al., 2009), Rapid Motor Imitation Antecedent Training (Paul et al., 2013; Tsiouri & Greer, 2003), and PROMPT (Rogers et al., 2006). Recently, the growing availability of computer and the smartphone/tablet has generated a great deal of enthusiasm for their potential to help ameliorate speech deficits in children with ASD, such as the computer-assisted 3-D virtual pronunciation tutor (Chen et al., 2019), and the iPad with the Proloquo2Go application (King et al., 2014). Taken together, in consideration of autistic features especially among the low-functioning individual who begin treatment with limited or no spoken words, available behavioral interventions often tried to apply orienting cues or motor activities to attract attention, to add AAC modes of communication, and to present speech sounds with computerized “high-tech” solutions (Tager-Flusberg & Kasari, 2013).

Recent years have witnessed an increased demand for music therapy in special education settings, which is regarded as a promising intervention for individuals with ASD (James et al., 2015; Reschke-Hernández, 2011). Given the behavioral resemblance between singing and speaking (Schlaug et al., 2008), as well as neural overlap in responses to speech and musical stimuli (Peretz et al., 2015), researchers have begun to examine the therapeutic effects of singing/intonation, and how it can potentially ameliorate some of the speech deficits associated with neurological disorders such as ASD (Wan et al., 2010). Two earlier case studies have described the positive role of singing (intoned rather than spoken verbal stimulus) in facilitating speech development of autism (Hoelzley, 1993; S. B. Miller & Toca, 1979). Recently, the Melodic

Intonation Therapy (MIT; Albert et al., 1973; Sparks et al., 1974), initially designed for improving spoken language in left-hemisphere stroke patients with severe nonfluent aphasia, has been introduced to speech therapy in ASD. The music-based MIT approach involves the musical elements of both melody and rhythm through the use of pitched vocalization or singing in combination with left-hand rhythmic tapping to provide cueing for syllable production (Norton et al., 2009). An adapted version of MIT, called Auditory-Motor Mapping Training (AMMT), was initially proposed by Wan et al. (2011) to facilitate speech output for English-speaking nonverbal children with autism. AMMT combines intonation (singing) of bi-syllabic words or phrases and the use of a pair of tuned drums to activate bimanual motor activities. By using a single-subject multiple baseline design, following AMMT training of the first 15 sessions (each session lasted 45 minutes per day), all the six English-speaking nonverbal children with autism showed noticeable improvements in their ability to articulate several word approximations, most of which were maintained several weeks after the cessation of the treatment sessions (Wan et al., 2011). Furthermore, the efficacy of AMMT has been further corroborated in English-speaking minimally verbal children with autism (Chenausky et al., 2016) and one more-verbal child with autism (Chenausky et al., 2017) when compared with a control treatment.

A very recent review estimated that ASD prevalence in China was comparable to the Western world (about 108 per 10,000) using standardized case identification protocol (Sun et al., 2019). However, the situation of ASD in China lags considerably behind those in the West in terms of public awareness, education opportunities, and life outcomes of autistic people (Liu et al., 2016; Yu et al., 2020). It is compelling to come up with a language-specific training approach for children with ASD who live in a country hosting nearly 20% of the world’s population. Specifically, Mandarin Chinese is the official and widely spoken language in China and also used in some other countries/regions, which differs a lot from English phonology. For example, Mandarin is a syllable-timed tonal language that exploits variations in both pitch height and pitch direction at the syllable level to distinguish lexical meanings, while English is a stress-timed non-tonal language (Peggy Mok & Dellwo, 2008). There are four citation tones in Mandarin Chinese (Figure 5.1): high-level Tone 1 (T1, [55]), mid-rising Tone 2 (T2, [35]), low-falling-rising Tone 3 (T3, [214]), and high-falling Tone 4 (T4, [51]). These four lexical tones are essential elements of Mandarin speech sounds, and are used to differentiate lexical meanings. For instance, “i” spoken with the four distinct tones can respectively mean “doctor” (T1), “move” (T2), “rely on” (T3), and “easy” (T4). Thus, changing the lexical tone in a tonal language has same kind of effect as changing a vowel or a consonant. In contrast, variation in pitch contours in non-tonal languages mainly conveys different moods or intonations without changing the word content (Wang, 1973). Although tone-language-speaking children with ASD also showed superior nonspeech pitch processing skills the same as the English-speaking ones at the group level, they had difficulties in the perception of native lexical tones at both the neural and behavioral levels (Chen et al., 2016; Wang et al., 2017; Yu et al., 2015). In the closely related domain of speech production, one recent study investigated speech sound acquisition in Mandarin-speaking children with ASD using a picture naming task (Wu et al., 2020). When compared with age-matched TD children, the ASD group aged 3–6 years showed an apparent speech delay in the production of Mandarin initials/onsets (consonants), finals (mainly including monophthong, diphthong, triphthong, and nasal finals), as well as lexical tones. Thus, speech therapy in tone-language-speaking children with ASD should not only aim at enhancing consonant and vowel production (segmental elements of speech), but also target at improving lexical tone production additionally (supra-segmental element of speech).

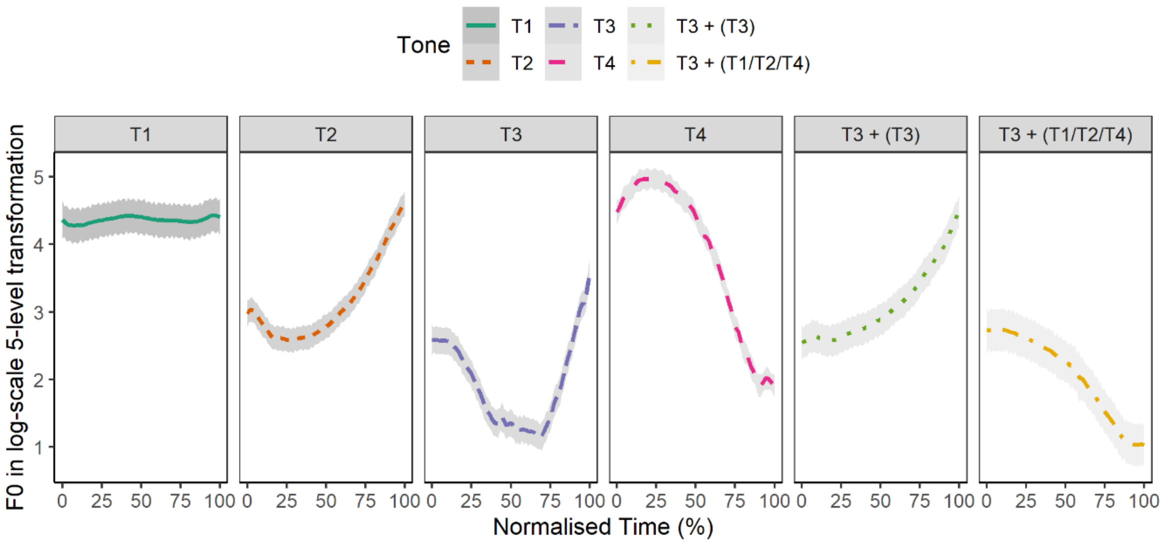


Figure 5.1 Pitch contours of four Mandarin citation tones (T1 [55], T2 [35], T3 [214], and T4 [51]) as well as two allophonic variants of T3 due to tone sandhi (full sandhi: *T3 [35] when occurred before another T3; half-T3: *T3 [21] when before T1/T2/T4). The five-level digits in square brackets are used to transcribe tones in Chao's tone letters (Chao, 1930). Grey shades indicate standard error.

Many children with ASD showed enhanced capacity of music processing (Applebaum et al., 1979; Heaton et al., 1998) and exhibited strong interests in learning and making music (Buday, 1995; Hairston, 1990). Since both music notes and lexical tones share the same psycho-acoustical attribute of pitch information, it would be reasonable to take advantage of the relative strength of musical skills to compensate for the relative weakness of speech sound especially lexical tone acquisition for tone language speaking children with ASD. As mentioned, the music-based therapy of MIT might offer an effective alternative to traditional speech therapy for children with speech impairment linked to autism (Wan et al., 2010). In addition to the singing/intonation of words/phrases, the other key component of MIT is the motor activities, by deploying tuned drums (or other tuned musical instruments) to facilitate auditory-motor mapping. In this study, we apply and assess the effect of the MIT-based spoken language treatment, with proven efficacy as an intervention for English-speaking children with ASD (Chenausky et al., 2016, 2017; Sandiford et al., 2013; Wan et al., 2011), on the Mandarin-speaking children with ASD. The application of MIT-based treatment for tone-language-speaking children with autism requires a few minor adjustments relative to that designed for non-tonal language speakers. For instance, in the AMMT system (Wan et al., 2011) for English-speaking children, the stressed English syllables were sung on the higher of the 2 fixed level pitches (at E^b or 311.127 Hz) while tapping one of the tuned drums, unaccented syllables on the lower level pitch (at C4 or 261.626 Hz) while tapping the other drum. However, given the dynamic changing feature of pitch contours among Mandarin citation tones and various forms of pitch variations due to tone sandhi of T3 (Figure 5.1), the tuned musical instrument needs to resemble the dynamic changes of pitch information in a training system designed for tone-language-speaking children with ASD. Thus, in the current study, we presented virtual piano, a common musical instrument, through a smartphone/iPad App. The therapist introduces each Mandarin disyllabic word by intoning the two syllables and simultaneously tapping the piano icons in the App tuned to the same two pitch contours of lexical tones for that particular word.

To this end, the present study evaluated the therapeutic potential of an adapted MIT in facilitating speech output for tone-language-speaking children with ASD. Moreover, as suggested by Wang (1978), children do not acquire speech sounds by learning individual phonemes one by one, but rather by learning speech categories through the learning of lexical items. Thus, in this study, our intervention based on a smartphone/iPad App, called Music-Mediated and Lexicon-

Integrated (MMLI) training, tries to combine phonology and word learning together. MMLI aimed to teach three components of Mandarin syllables – initials, finals, and lexical tones – through well-designed, high-frequency disyllabic lexical items (see Methods for more details), and by incorporating the key MIT elements through the association of motor activity (i.e., bimanual hand tapping) as well as intoned vocal output. The current training study used a randomized controlled trial to assess the efficacy of MMLI, an MIT-based treatment for facilitating spoken language in Mandarin-speaking nonverbal and low-verbal children with ASD, in comparison to a matched non-MIT-based control treatment, Speech Repetition Therapy (SRT). The control condition of SRT was also presented through a smartphone/iPad App, with a similar amount of speech sound input to children, and involves repeated, structured imitation of spoken stimuli, but does not involve hand tapping on the piano icons and the intoned verbal stimuli of piano-timbre nonspeech. That is to say, the SRT is designed in a similar manner to the conventional forms of speech therapy, while lacking the key elements of MIT (Chenausky et al., 2016, 2017). Specifically, the current training study compared the two training approaches with a randomized controlled design to address the following questions:

(1) Over intensive training sessions in nonverbal and low-verbal Mandarin-speaking children with ASD, would the MMLI lead to a greater improvement in lexical tone production, word acquisition, and target sounds in initial and final positions?

(2) Can the benefits of MMLI be retained after the cessation of daily training sessions and generalize to untrained items?

5.2 Methods

5.2.1 Participants

Table 5.1 *Characteristics of children with ASD and age-matched TD children.*

	ASD (n = 30)		TD (n = 30)		t	p
	M	*SD*	*M*	*SD*		
Age (in month)	67.80	15.06	67.57	16.12	0.56	.578
Language Ability	31.67	26.84	88.07	8.81	-11.81	<.001
Nonverbal IQ	59.53	12.32	105.83	17.34	-13.78	<.001
Working Memory	3.90	4.39	12.23	3.65	-12.00	<.001
% Words Correct	22.68	20.99	73.88	14.73	-12.30	<.001
% Initials Correct	18.69	16.71	69.74	13.81	-14.23	<.001
% Finals Correct	17.60	15.59	66.31	13.74	-14.05	<.001
% Tones Correct	17.59	15.71	68.38	14.74	-14.24	<.001

Thirty Mandarin-speaking nonverbal and low-verbal children with a diagnosis of ASD between the ages of 3 and 7 years were included and completed all the training sessions. They were recruited from Cangzhou Research Centre for Child Language Rehabilitation. Permission to conduct this study was obtained from the Hong Kong Polytechnic University, ensuring appropriate adherence to informed consent procedures. The clinical diagnosis of ASD was made based on the DSM-5 criteria for ASD (American Psychiatric Association, 2013), and confirmed with the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012), or Gilliam Autism Rating Scale–Second Edition (GARS-2; Gilliam, 2006) by pediatricians and child psychiatrists before enrollment in local hospitals. An additional 11 children with autism were found to be ineligible and were excluded from the current study because they either did not meet the inclusion criteria or could not regularly attend all the required training sessions. The nonverbal status was defined as having the complete absence of intelligible words before training, and low-verbal status as using expressive vocabulary of no more than 50 words, based on parent report as well as the pretest measures. All the 30 participants with ASD belong to low-functioning ones, and they indeed experienced significant language and cognitive delays compared to age-matched TD children (Table 5.1). Besides, other inclusion criteria were: 1) the ability to repeat at least two auditory sounds; 2) the ability to sit in a chair and take part in instructed activities for around 15 minutes at a time; 3) the ability to imitate gross motor activities such as clapping hands, and imitate oral motor movements; 4) without the following comorbidities including cerebral palsy or tuberous sclerosis, hearing/sight impairment, Down's syndrome, uncontrolled seizures, and organic impairment of oral or laryngeal structures.

A randomized control design was used in an effort to determine the effectiveness of the experimental treatment (MMLI) and control for various external factors. Prior to training, all the participants with ASD were assessed with their phonology and vocabulary size using a picture-naming test, the overall language ability (Ning, 2013), the nonverbal IQ using the *Primary Test of Nonverbal Intelligence* (PTONI, Ehrler & McGhee, 2008), and the working memory (Millman & Mattys, 2017). Upon entering the training sessions, the 30 children with ASD were randomly assigned to one of two treatment groups: the MIT-based group (MMLI group), and the traditional therapy group, Speech Repetition Therapy (SRT group). The MMLI group (*n* = 15, two girls) represented the experimental group, and the SRT group (*n* = 15, two girls) acted as the active control group. The two treatment groups did not differ from each other in terms of chronological age, language ability, nonverbal IQ, working memory, vocabulary size, and the speech production ability before training (all *p*s ⩾ .542, see Table 5.2 for a statistical description of the participants' characteristics in two treatment groups).

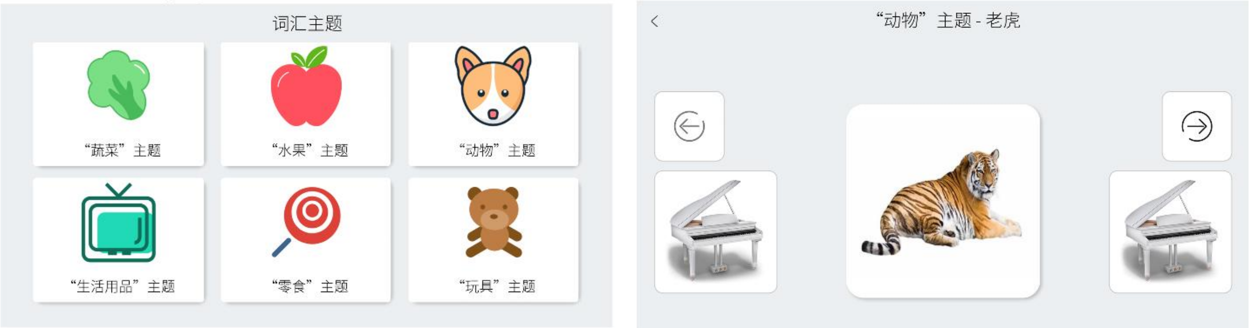
Table 5.2 *Characteristics of participants with ASD in two treatment groups.*

	MMLI (*n* = 15)		SRT (*n* = 15)		*t*	*p*
	M	*SD*	*M*	*SD*		
Age (in month)	66.13	15.32	69.47	15.14	-0.63	.542
Language Ability	33.07	25.35	30.27	29.08	0.47	.643
Nonverbal IQ	60.53	13.86	58.53	10.97	0.65	.528
Working Memory	3.93	4.38	3.87	4.55	0.05	.962
% Words Correct	22.44	23.48	22.92	19.02	-0.15	.887
% Initials Correct	18.49	17.81	18.90	16.17	-0.28	.785
% Finals Correct	17.88	17.05	17.32	14.60	0.22	.830
% Tones Correct	17.90	17.71	17.28	14.05	0.16	.872

5.2.2 Study Design

5.2.2.1 The Development of Smartphone/Ipad App

The App was developed based on the Ionic Framework (https://ionicframework.com/docs), which uses the UI toolkit for building mobile and desktop apps using web technologies. It is an open-source front-end framework for HTML5 hybrid mobile application, using AngularJS, Typescript, Html, Scss, Cordova, and related technologies such as Nodejs to help developers use the same source code to generate App files for both Android and iOS-based platforms. The primary function of the current App is to assist children's pronunciation training. The home page presented six themes, including vegetables, fruits, animals, daily necessities, snacks, and toys (Figure 5.2a). Under each theme, a total of 10 high-frequency lexical items were chosen, resulting in 60 trained lexical items in total. The 60 trained words were disyllabic nouns (except one onomatopoeia [uəŋ55 uəŋ55] imitating the sound of a bee) relevant to children's early-acquired vocabulary. These disyllabic items contained all the 21 initials, 39 finals, four citation tones as well as neutral tone, T3 tone sandhi in Mandarin phonology (see Appendix C for more details).



(a) The home page of six themes

(b) The interface of one lexical item: 老虎 (tiger)

Figure 5.2 The user interface of the App: (a) the home page of six themes, (b) the interface of one lexical item.

Then, for each lexical item, one corresponding picture was presented in the center of the interface as the visual cue, and two piano icons on the left and right sides by using the HTML img tag (Figure 5.2b). The natural speech sounds of 60 trained words (120 syllables) were recorded from one female announcer with standard Mandarin pronunciation. To construct the piano-timbre nonspeech sounds, first, a piano note (C4) with 261 Hz frequency was created, then the level pitch tier was replaced with the pitch contour extracted from each syllable (120 syllables in total) using the Pitch-Synchronous Overlap Add implanted in Praat (Boersma & Weenink, 2016). In this way, the piano-timbre nonspeech sounds share the same pitch contours as those in natural speech sounds (Figure 5.3). All the piano-timbre nonspeech sounds were normalized to be 500 ms, and equally for root-mean-square intensity level of 70 dB SPL. To confirm that piano-timbre nonspeech stimuli could no longer be perceived as speech, we recruited 8 native adult speakers naïve to the stimuli to rate all the samples using a 7-point Likert scale (7– definitely speech, 1– definitely not speech). The findings showed that natural speech stimuli received a mean score of 6.96 and piano-timbre nonspeech stimuli received 1.78. During MMLI training, when tapping the icons from left to right, the piano-timbre nonspeech sounds that match the pitch contours of natural lexical tones for the first and second syllables would be played, one tap per syllable. Audio playback was implemented using the ionic-native/native-audio library. As for the source code, please refer to the following address: https://github.com/introfei/VoiceTrain, and for compiling, packaging, and uploading issues, please refer to: https://github.com/introfei/Blog/issues/3 to see more details.

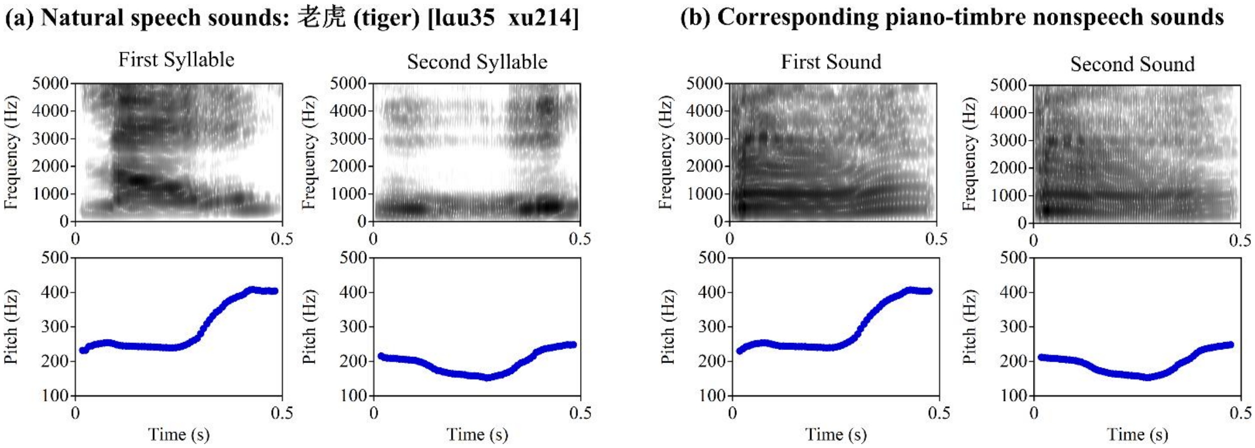


Figure 5.3 The spectrograms (the upper row) and pitch contours (the bottom row) of one lexical item "老虎" (tiger) [lɑu35 xu214] in (a) natural speech sounds, and (b) piano-timbre nonspeech sounds. The two types of sounds share exactly the same pitch contours with blue curves.

5.2.2.2 Treatment Protocol

Therapists were female undergraduate students from the Cangzhou Research Centre for Child Language Rehabilitation, and were trained specially to provide both MMLI training and SRT. In other words, if one therapist trained a subject from the MMLI group, she would also need to train another subject in the SRT group. For each training session, a warm-up activity around five minutes was firstly introduced before teaching lexical trials. The treatment protocol of MMLI and SRT sessions was shown in Table 5.3. They were conducted with intensive repetition in a highly structured environment. While SRT also presents verbal stimuli through the App interface and contains the same steps and speech outputs as MMLI, in SRT the verbal stimuli are spoken, not intoned; and there is no bimanual hand tapping on piano icons. To monitor the treatment fidelity, all treatment sessions were videotaped to evaluate therapists' adherence to the protocol. We reviewed five videotaped sessions selected at random for each child, and all reviewed sessions closely followed the MMLI or SRT treatment structure.

Table 5.3 *The warm-up stage at the beginning of each training session and the five-step structure of an MMLI trial vs. an SRT trial.*

	MMLI (Experimental Group)	**SRT (Control Group)**
Warming Up	Musical melodies without lyrics are played, and musical toys such as shaking maracas are introduced to facilitate their movements. Moreover, a rhythmic tapping of the foot is also used in time to music.	Playing checkboards without verbal or with minimally verbal instruction.
Steps	**MMLI Trial**	**SRT Trial**
1. Word Introduction	Therapist introduces the target word by showing a word picture (such as “tiger”) on the phone/iPad App and then **intoning (singing)** the word ‘‘[lɑu35 xu214]’’ by **tapping** the piano icons 1× per syllable.	Therapist introduces the target word by showing a word picture (such as “tiger”) on the phone/iPad screen and then **speaking** the word ‘‘[lɑu35 xu214]’’ without finger tapping.
2. Synchronous Production	Therapist produces target with the child. Therapist **intones and taps** ‘‘Let’s sing it together’’ and in synchrony with child ‘‘[lɑu35 xu214]’’.	Therapist produces target with the child. Therapist **speaks** “Let’s speak it together’’ and in synchrony with child ‘‘[lɑu35 xu214]’’.
3. Unison with Fading	Therapist and participant begin to **intone and tap** the target word together, but after the first syllable, the therapist stops while the child continues to intone and tap the next syllable. ‘‘[lɑu35] ____’’.	Therapist and participant begin to **speak** the target word together, but after the first syllable, the therapist stops while the child continues to produce the next syllable. ‘‘[lɑu35] ____’’.
4. Immediate Imitation	Therapist firstly **intones and taps** the target word alone. Afterwards, participant imitates the word, and therapist remains silent. ‘‘My turn first: [lɑu35 xu214]. Now your turn: _______’’.	Therapist firstly **speaks** the target word alone. Afterwards, participant imitates the word, and therapist remains silent. ‘‘My turn first: [lɑu35 xu214]. Now your turn: _______’’.
5. Independent Production	The child is further encouraged to independently **intone and tap** the target word once again. ‘‘__________’’	The child is further encouraged to independently **speak** the target word once again. ‘‘__________’’

5.2.2.3 Training Procedure

The therapy sessions were conducted in clinical treatment rooms at Cangzhou Research Centre for Child Language Rehabilitation. The child participants in both MMLI and SRT groups received short-term intensive training: 12 individual sessions 6 times per week (i.e., 6 days/sessions for each intervention phase), over a two-week period (i.e., two intervention phases). However, after the first-round training, 3 out of 15 nonverbal children with ASD from each treatment group demonstrated no progress at all, and still remained nonverbal. These six nonverbal participants were further trained for another 12 sessions in the second round, with 24 training sessions in total (Figure 5.4). Each training session began with a warm-up stage, and then followed by 10 lexical trials (each trial was repeated three times) of one specific theme. Each session lasted about 50–60 minutes, including the breaks, which occurred every ten to fifteen minutes, based on the child's stamina. The order of six training themes within one intervention phase was randomized using a Latin Square among participants. The training order in the second phase is a repetition of that in the first intervention phase for each child. While receiving MMLI or SRT, the child participants did not engage in any other speech therapy activities in regular school programs.

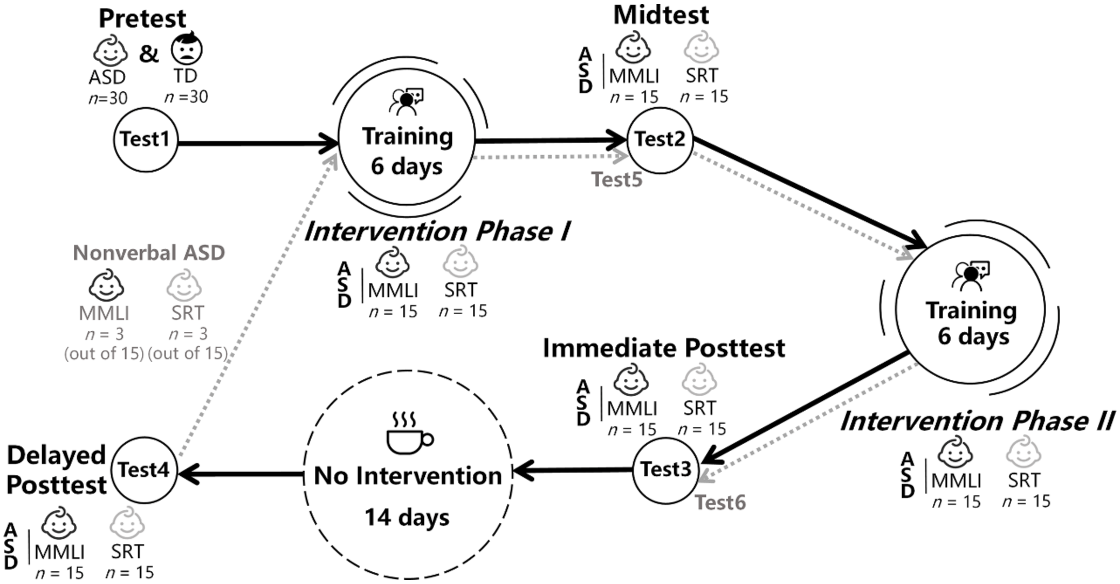


Figure 5.4 The probe assessments and training procedure for two treatment groups (MMLI and SRT).

The probe assessments were conducted in the Pretest (Test 1), Midtest (Test 2), Immediate Posttest (Test 3), and Delayed Posttest (Test 4). The Delayed Posttest was conducted at two weeks post-therapy to assess whether any changes observed during therapy persisted after treatment ended. Specifically, for those six nonverbal participants with ASD, the second-round Midtest (Test 5), and Immediate Posttest (Test 6) were performed additionally. The picture-naming method was used in the probe assessments, with trained (Set 1) and untrained (Set 2) picture stimuli intermixed and presented in random order. Set 1 consisted of 60 lexical items which were presented during probe assessments and also practiced during therapy sessions. Set 2 included 12 high-frequency items which were not practiced during treatment, but presented during the probes (see Appendix C). The untrained stimuli in Set 2 were used to assess the transfer of learning to novel stimuli. The child was encouraged to make more than one attempt if did not respond, or failed to pronounce the target word correctly. However, it should be noted that, during the probe assessments, neither feedback nor correct demonstration was provided. The spontaneous productions from each child were recorded in a quiet aural room for further analyses.

5.2.3 Outcome Measures

5.2.3.1 Word and Speech Production Measure

The outcome measures of interest were the child's word and speech productions when presented with the trained and untrained picture stimuli during the probe assessments. The recorded productions were transcribed offline by Mandarin-speaking coders who were totally blind to the current study design to minimize experimental bias. Before transcription, an expert majoring in linguistics picked out the best sample from each child's utterances if more than one attempt was produced for one specific target word. After selection, there were totally 3,634 words (7,268 syllables) produced by all the child participants from two training groups across all the probe assessments.

First, in terms of word production measure, the 3,634 produced words were randomized and presented to five native speakers of Mandarin (M_{age} = 22.45 year) with the E-Prime 2.0 program (Psychology Software Tools Inc., USA). They were asked to write down each word they heard with two Chinese characters (each Chinese character representing a morpheme in Mandarin word) one by one in the spreadsheet to evaluate the comprehensibility. Each coder needs around 15 hours to complete the transcription of words. The number of correctly coded characters was divided by the total number of characters to yield "Percent Words Correct".

Second, for speech production measure, the disyllabic word was split into two syllables, which were transcribed separately. The 7,268 produced syllables were randomized and further transcribed using International Phonetic Alphabet (IPA) by another five trained experts majoring in linguistics (M_{age} = 23.45 year). Especially for the tonal coding, the exact tonal categories were chosen from the following descriptions: high-level tone (T1), mid-rising tone (T2 or full sandhi: *T3 [35]), dipping tone (typical T3), high-falling tone (T4), low-falling tone (half-T3: *T3 [21]), and the neutral tone. Such stringent measures of phonetic transcription using IPA aimed to assess speech intelligibility (Munro & Derwing, 1995) in a more fine-grained and precise manner. Each coder needed around 40 hours in total to complete the transcription of all the initials, finals, and tones. Three outcome measures were included to evaluate the speech production capacity of three components of Mandarin syllables (initials, finals, and tones): Percent Initials Correct, Percent Finals Correct, and Percent Tones Correct which were calculated with the number of correctly transcribed initials, finals, and tones divided by the total number of syllables respectively. Kendall's Concordance Coefficient W was calculated for assessing agreement among raters (Legendre, 2005). The interrater reliability with Kendall's coefficients of 0.884 for word coding, 0.802 for initial transcription, 0.793 for final transcription, 0.828 for tone transcription was respectively reached, whose results exhibited relatively high inter-rater reliability.

5.2.3.2 User Experience

The user experience evaluation was executed after the completion of all the training sessions, which was rated by the therapists using a 5-point Likert scale based on the child's training performance (5– the highest degree, 1– the lowest degree). Such subjective observations evaluated the ways in which children with ASD approached different training methods. The user experience evaluation included five aspects: enjoyment, cooperation, consistency, interest, and motivation. The enjoyment refers to the degree of pleasure in the learning process; The cooperation means the degree of collaboration in learning a trial; The consistency indicates the continuity of the overall coordination throughout the learning process; The interest refers to the degree of interest in training materials; The motivation depicts the degree of the initiative before training sessions (whether the child wants to participate in training).

5.2.4 Statistical Analyses

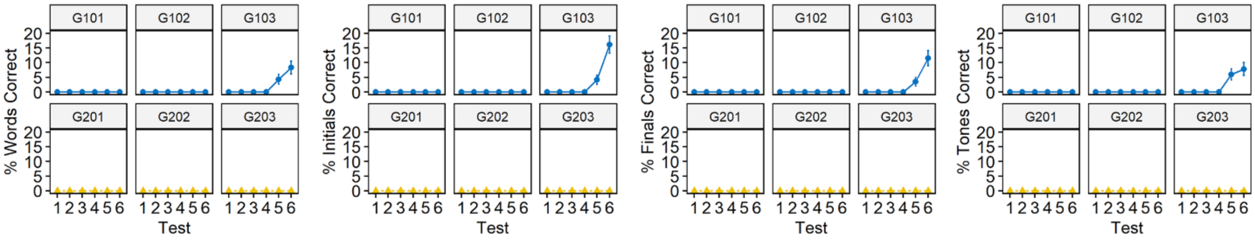
All the statistical analyses of outcome measures were performed in R (R Core Team, 2014). For the analyses of speech and word production accuracy, the generalized linear mixed-effects models (GLMMs) were created using the lme4 package (Bates et al., 2014). In each GLMM, the dichotomous response ("1" means correct or "0" means incorrect) was entered as the dependent measure, with *treatment group* (MMLI vs. SRT), *test*, and their two-way interaction acting as fixed effects. When fitting GLMMs, *subject* and *item* were included as random effects. A full structure of random effects was included in the initial model (Barr et al., 2013), which was compared with a simplified model that excluded a specific fixed factor using the ANOVA function in lmerTest package (Kuznetsova et al., 2017). Post-hoc pairwise comparisons were performed using the lsmeans package (Lenth, 2016) with Tukey adjustment. For the analyses of user experience, a generalized Poisson regression model (Consul & Famoye, 1992) was constructed in R, with *treatment group* (MMLI vs. SRT), *aspect* (motivation, consistency, interest, cooperation, and enjoyment), and their two-way interaction acting as fixed effects. The generalized Poisson regression model has been found useful in fitting over-dispersed as well as under-dispersed count data. Given that the nonverbal and low-verbal participants in this study received different amounts of training sessions, their results were reported separately.

5.3 Results

5.3.1 Outcomes in Nonverbal Participants with ASD

There were six nonverbal participants with ASD (*G101, G102, and G103* in MMLI group; *G201, G202, and G203* in SRT group), who have received 24 training sessions across four intervention phases in total (Figure 5.4). The probe assessment data were collected 6 times before, during, and after therapy (Figure 5.4): Test 1 (Pretest), Test 2 (First-round Midtest), Test 3 (First-round Immediate Posttest), Test 4 (First-round Delayed Posttest), Test 5 (Second-round Midtest), Test 6 (Second-round Posttest). As shown in Figure 5.5a, only one subject (G103) from the MMLI group began to acquire some words, initials, finals, and tones in the trained items during the second-round training (Test 5 and Test 6), while all the other 5 participants remained nonverbal even after 24 training sessions. None of the six nonverbal participants showed any improvement in the untrained novel items after training (Figure 5.5b). For the evaluation of user experience, the subject from the MMLI group (G103) who showed gains in speech and word acquisition also obtained relatively higher scores of user experience, especially in the aspect of enjoyment (Figure 5.6).

(a) Trained Items



(b) Untrained Items

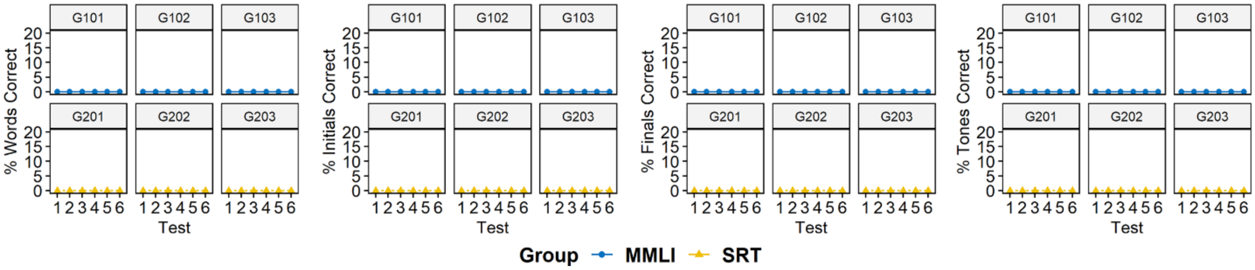


Figure 5.5 The obtained Percent Words Correct, Percent Initials Correct, Percent Finals Correct, and Percent Tones Correct for each nonverbal child by treatment group (MMLI and SRT) and

probe assessment (Tests 1-6) for the (a) Trained Items, and (b) Untrained Items. Error bars: +/- 1 Confidence Interval.

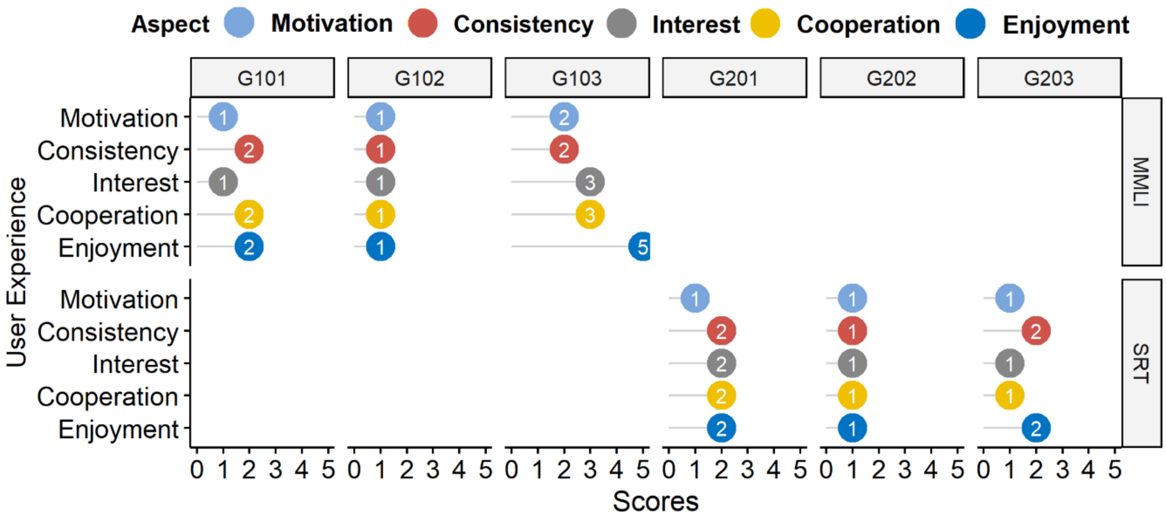


Figure 5.6 The scores of user experience for the six nonverbal participants in terms of motivation, consistency, interest, cooperation, and enjoyment. The obtained scores were shown by the numerical labels in the circles.

5.3.2 Outcomes in Low-Verbal Participants with ASD

There were totally 24 low-verbal participants with ASD (*n* = 12 in each treatment group) who have received 12 training sessions across two intervention phases (Figure 5.4). The probe assessment data were collected before, during, and after therapy: Test 1 (Pretest), Test 2 (Midtest), Test 3 (Immediate Posttest), and Test 4 (Delayed Posttest). Figure 5.7 shows the percentage of correct productions from two treatment groups in both trained and untrained items. The x-axis represents the probe assessment sessions (Tests 1-4) and the y-axis stands for the percentage of correct words, initials, finals, and tones respectively from left to right.

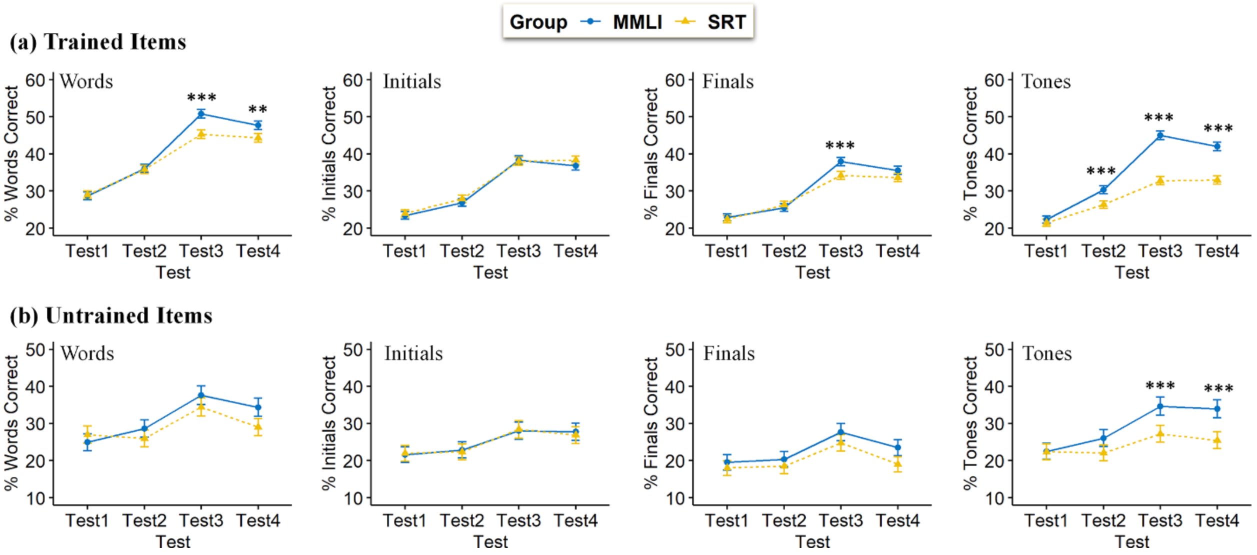


Figure 5.7 The production accuracy of words, initials, finals, and tones for the low-verbal participants by treatment group (MMLI and SRT) and probe assessment (Tests 1-4) in the (a) Trained Items, and (b) Untrained Items. ***$p < .001$; ** $p < .01$ after Tukey adjustment for the comparison of MMLI vs. SRT. Error bars: +/- 1 Confidence Interval.

5.3.2.1 Word Production Accuracy

First, for the trained items, the GLMM on word production accuracy showed a significant interaction of *treatment group* × *test* (χ^2 (3) = 34.58, $p < .001$), indicating that the two training groups showed different trajectories of word acquisition in the trained items. As shown in Figure 5.7a, post-hoc pairwise comparisons indicated that, compared to the baseline performance in Pretest (M_{MMLI} = 28.7%; M_{SRT} = 29.0%), both MMLI and SRT groups showed noticeable improvements in trained word production (all *p*s < .001) when tested at Midtest (M_{MMLI} = 36.1%; M_{SRT} = 35.8%), Immediate Posttest (M_{MMLI} = 50.8%; M_{SRT} = 45.3%), and follow-up assessment at 2 weeks later (M_{MMLI} = 47.7%; M_{SRT} = 44.3%). In terms of group difference at different timepoints, the two treatment groups performed similarly on Percent Words Correct in the Pretest (β = -0.015, SE = 0.038, t = -0.40, p = .688) and Midtest (β = 0.011, SE = 0.036, t = 0.29, p = .772), whereas the MMLI group produced a higher accuracy of the trained words than the control SRT group in the Immediate Posttest (β = 0.247, SE = 0.035, t = 7.03, $p < .001$) as well as Delayed Posttest (β = 0.103, SE = 0.035, t = 2.94, $p < .01$).

Second, for the untrained novel items, the GLMM on word production accuracy also revealed a significant interaction of *treatment group* × *test* (χ^2 (3) = 19.09, $p < .001$). Post-hoc analysis showed that compared to the performance in Pretest, the MMLI group produced more untrained words in Midtest (β = -0.31, SE = 0.11, t = -2.89, $p < .05$), Immediate Posttest (β = -1.02, SE = 0.11, t = -9.49, $p < .001$), and Delayed Posttest (β = -0.77, SE = 0.11, t = -7.18, $p < .001$). Over the same probe assessments, however, the SRT group only produced more untrained words in the Immediate Posttest (β = -0.54, SE = 0.10, t = -5.37, $p < .001$) right after 12 training sessions (Figure 5.7b), while there was no difference between the Pretest and Midtest (β = 0.08, SE = 0.10, t = 0.77, p = .868) or between the Pretest and Delayed Posttest (β = -0.15, SE = 0.10, t = -1.48, p = .450). For the between-group difference, among all the probe assessments (Tests 1-4), the MMLI group performed similarly as the control SRT group in terms of % Words Correct at the picture naming task in the novel words (all *p*s > .05).

5.3.2.2 Production Accuracy of Initials

The GLMM was performed on the production accuracy of initials in trained stimuli, and the statistical results showed a significant main effect of *test* (χ^2 (3) = 1143.70, $p < .001$). However, the GLMM did not reveal significant main effect of *treatment group* (χ^2 (1) = 1.33, p = .468) nor the interaction effect of *treatment group* * *test* (χ^2 (3) = 3.35, p = .341), indicating that the treatment methodology (MMLI vs. SRT) did not lead to outcome differences in the production accuracy of initials in the trained stimuli (Figure 5.7a). Further examination on the effect of *test* implied that in comparison to the production accuracy of initials in Pretest prior to therapy (M_{MMLI} = 23.4%; M_{SRT} = 24.0%), the low-verbal participants from both MMLI and SRT groups showed significant improvement at producing initials in the trained items (all *p*s < .001), at Midtest (M_{MMLI} = 26.9%; M_{SRT} = 27.9%), Immediate Posttest (M_{MMLI} = 38.4%; M_{SRT} = 38.0%), as well as Delayed Posttest (M_{MMLI} = 36.8%; M_{SRT} = 38.3%).

Then, the GLMM showed that neither the main effect of *treatment group* (χ^2 (1) = 0.05, *p* = .818) nor the interaction effect of *treatment group * test* (χ^2 (3) = 0.51, *p* = .918) reached significance on the accuracy of initials in the untrained items, while there was a main effect of *test* (χ^2 (3) = 56.01, *p* < .001). For both MMLI and SRT groups, as shown in Figure 5.7b, the number of correctly produced initials of untrained stimuli increased significantly after the whole 12 training sessions at Test 3 (β = -0.39, SE = 0.07, *t* = -6.03, *p* < .001) and at follow-up assessment at Test 4 (β = -0.34, SE = 0.07, *t* = -5.22, *p* < .001).

5.3.2.3 Production Accuracy of Finals

For the trained items in Set 1, the GLMM model on the accuracy of finals showed a significant two-way interaction of *treatment group* × *test* (χ^2 (3) = 17.61, *p* < .001). The two training groups differed after two intervention phases (Figure 5.7a), with MMLI group showing a higher production accuracy of trained finals than the matched control group at Test 3 (β = 0.18, SE = 0.04, *t* = 4.88, *p* < .001), while no group differences were found at Test 1, Test 2 and Test 4 (all *p*s > .05). Furthermore, compared to Pretest (M_{MMLI} = 22.9%; M_{SRT} = 22.4%), both treatment groups made significant progress in the trained finals over the course of treatment (all *p*s < .001), at Midtest (M_{MMLI} = 25.5%; M_{SRT} = 26.2%), Immediate Posttest (M_{MMLI} = 37.9%; M_{SRT} = 34.2%), and Delayed Posttest (M_{MMLI} = 35.6%; M_{SRT} = 33.7%).

For the untrained items in Set 2 (Figure 5.7b), the GLMM on production accuracy of finals did not reveal significant interaction of *treatment group* × *test* (χ^2 (3) = 3.33, p = .342), nor the main effect of *treatment group* (χ^2 (1) = 3.37, p = .066). The main effect of *test* was found to be significant (χ^2 (3) = 79.11, p < .001), with low-verbal participants from both treatment groups showing significant progress in number of correctly produced finals in untrained items only at the immediate posttest (β = -0.59, SE = 0.07, t = -7.97, p < .001).

5.3.2.4 Production Accuracy of Tones

The GLMM on the accuracy of trained tones showed a significant two-way interaction of *treatment group* × *test* (χ^2 (3) = 112.87, p < .001), indicating that the two treatment groups showed different trajectories of lexical tone acquisition in trained items (Figure 5.7a). In comparison to the performance in Pretest (M_{MMLI} = 22.4%; M_{SRT} = 21.5%), both MMLI and SRT groups showed significant improvement at tone production in the trained items (all ps < .001), at Midtest (M_{MMLI} = 30.3%; M_{SRT} = 26.3%), Immediate Posttest (M_{MMLI} = 45.0%; M_{SRT} = 32.8%), as well as Delayed Posttest (M_{MMLI} = 42.0%; M_{SRT} = 33.0%). Nevertheless, the growth rate was quite different, with participants who received MMLI training showing much higher production accuracies of trained tones compared with those receiving traditional SRT, at Midtest (β = 0.22, SE = 0.04, t = 5.62, p < .001), Immediate Posttest (β = 0.61, SE = 0.04, t = 16.26, p < .001), as well as Delayed Posttest (β = 0.45, SE = 0.04, t = 12.11, p < .001).

Then, GLMM was performed on the accuracy of tones in untrained items, and the statistical results showed a significant interaction of *treatment group* × *test* (χ^2 (3) = 14.27, p < .01). The low-verbal participants from MMLI group had significant progress in production of lexical tones in untrained items, at Immediate Posttest (β = -0.67, SE = 0.09, t = -7.57, p < .001), and Delayed Posttest (β = -0.64, SE = 0.09, t = -7.16, p < .001), while the matched participants from traditional

SRT group showed no progress over the course of treatment (all *p*s > .05). In terms of group difference (Figure 5.7b), the experimental group of MMLI obtained much higher accuracy of tones in untrained stimuli after 12 training sessions at Immediate Posttest ($\beta = 0.42$, SE = 0.12, $t = 3.41$, $p < .001$) and two weeks later at Delayed Posttest ($\beta = 0.49$, SE = 0.13, $t = 3.90$, $p < .001$).

5.3.2.5 User Experience

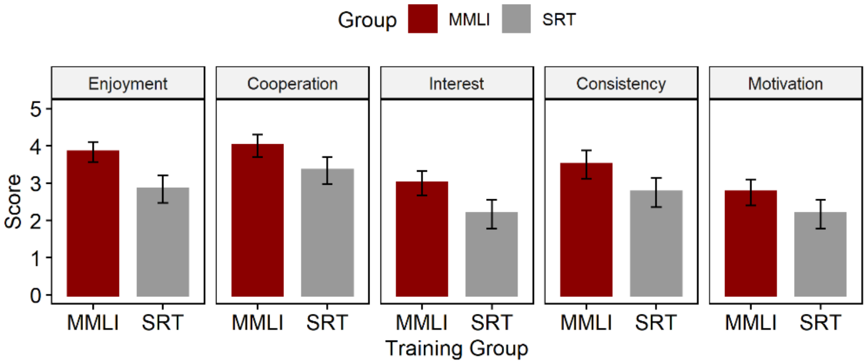


Figure 5.8 The average scores of user experience for the low-verbal participants in two treatment groups. Error bars: +/- 1 standard error.

The generalized Poisson regression model on scores of user experience for the low-verbal children with ASD ($n = 12$ in each treatment group) did not reveal significant interaction of *treatment group × aspect* (χ^2 (4) = 0.24, $p = .993$). There was a trend toward significance (χ^2 (4) = 8.22, $p = .084$) for the main effect of *aspect* ($M_{enjoyment} = 3.33$, $M_{cooperation} = 3.67$, $M_{interest} = 2.58$, $M_{consistency} = 3.12$, $M_{motivation} = 2.46$). Moreover, as shown in Figure 5.8, the main effect of *treatment group* ($M_{MMLI} = 3.42$, $M_{SRT} = 2.65$) was also found to be marginally significant (χ^2 (1) = 3.02, $p = .082$).

5.4 Discussion

Our current training App of MMLI aimed to integrate recent advances in speech therapy to facilitate speech sound acquisition in tone-language-speaking children with ASD. This randomized controlled study compared the efficacy of MIT-based therapy (MMLI) and non-MIT-based traditional therapy (SRT) for eliciting spoken language output in nonverbal and low-verbal children with ASD. The short-term intensive training of one-on-one, 50–60 min per day, 6 days per week setting was conducted for both treatment groups, which lasted two weeks (12 training sessions) and four weeks (24 training sessions) for the low-verbal and nonverbal children with ASD respectively. Efficacy was evaluated by user experience, as well as by % Words Correct, % Initials Correct, % Vowels Correct, and % Tones Correct, where the child's spontaneous production must be an exact match to the canonical form. For the low-verbal participants, while both MMLI and traditional speech therapy were found to be effective in enhancing the production skills, results suggested a faster rate of improvement in the production of words, vowels, and lexical tones in the trained items for the experimental group of MMLI. The higher production accuracy of words and lexical tones made from MMLI was maintained 2 weeks after the cessation of the treatment sessions. Moreover, the advantage of MMLI training transferred to the untrained novel items in terms of lexical tone production. For the six nonverbal participants, however, only one from the MMLI Group responded to treatment in the trained items, while others from both training groups showed no progress and remained nonverbal even after 24 training sessions. We will discuss these findings in the following parts.

5.4.1 Mechanisms Responsible for the Training Efficacy of MMLI

The MMLI resulted in greater improvements in most of the outcome measures than SRT for the Mandarin-speaking participants, which largely corroborated the efficacy of MIT-based training approach, with well-proven efficacy as a treatment for English-speaking children with autism

(Chenausky et al., 2016, 2017; Sandiford et al., 2013; Wan et al., 2011). The reported data here further proved that the two key elements of MIT – intonation and hand tapping – added greatly to MMLI's effectiveness in children from another language system. When tapping on the virtual piano presented through App, the nonspeech sounds with piano timbre would be generated, in an effort to mimic the music-making activities. Different from previous studies (Chenausky et al., 2016, 2017; Sandiford et al., 2013; Wan et al., 2011), our MMLI system did not utilize the real musical notes, but was modified to match various pitch variations of lexical tones in Mandarin phonology. Hoelzley (1993) proposes that the unique timbre of musical instrument may increase motivation and attention in autistic children. Besides, the structure of MMLI therapy requires participants to tap on one of two tuned piano icons, with the piano-timbre sounds played in sync with the moment when touching on the smartphone/iPad screen. This multimodal procedure fits well with one interesting theoretical speculation claiming that children with ASD tend to show an alternative learning path for language acquisition by orienting toward audiovisual synchrony (Jones et al., 2008; Klin et al., 2009). The increased motivation and attention towards MMLI training could be supported by relatively higher scores of user experience when compared to the traditional SRT approach (Figure 5.8). Furthermore, music making through bimanual tapping on the tuned piano icons is a multimodal activity that not only captures the autistic child's interest, but also primes and integrates the bilateral sensorimotor networks with shared motor, auditory and visual neural representations of the articulatory/hand movements (Bangert et al., 2006; Koelsch et al., 2002; Lahav et al., 2007; Meister et al., 2003).

5.4.2 Training Efficacy of Lexical Tones, Vowels, and Consonants for Low-Verbal Children with ASD

While MMLI holds promise for improving the spoken language in Mandarin-speaking children with ASD in general, the effectiveness of the MMLI was unbalanced among different components of syllables (i.e., initials, finals, and tones). A shown in Figure 5.7a, the low-verbal participants with ASD receiving MMLI began to show superiority over SRT in their ability to correctly articulate Mandarin lexical tones in trained items as early as Midtest after 6 training sessions, and such advantage further expanded after 12 training sessions at Immediate Posttest and was maintained at Delayed Posttest. In terms of the speech production of Mandarin finals which use vowel(s) as the whole final or as the nucleus, the low-verbal MMLI participants only experienced comparatively greater improvement than the matched SRT participants after 12 treatment sessions, and the treatment advantage for finals could not persist at the follow-up assessment. In terms of the speech production of Mandarin initials which were composed of consonants, the two treatment groups performed similarly in the trained items over all the probe assessments. Furthermore, for the untrained items (Figure 5.7b), MMLI produced significantly greater gains merely in the lexical tone acquisition in low-verbal children with ASD than a control therapy, SRT. Such generalization skills would be greatly beneficial to children with ASD, who show difficulty in transferring learned knowledge to a new context (Church et al., 2015; Happé & Frith, 2006). In a short conclusion, observed from the current data, the efficacy of MMLI was much higher in the training of lexical tones, followed by vowels, and then consonants.

The superior improvement on lexical tone acquisition should not be surprising given that relative to SRT, MMLI presented additional information of pitch contours embedded in piano-timbre nonspeech to participants. Music is one of the most meaningful and popular forms of nonspeech sound; like speech, it has developed to take advantage of the efficiencies of the human auditory system (Baldwin, 2012). In the research into auditory and speech processing, several studies demonstrated that children with ASD prefer and attend to musical or nonspeech stimuli over speech (Dawson et al., 1998; Kuhl et al., 2005). Moreover, a series of behavioral and neuroimaging studies have implied that, compared to TD controls, individuals with ASD have always demonstrated a pitch or melodic processing superiority in various musical and nonspeech stimuli (Bonnel et al., 2010; Ferri et al., 2003; Foxton et al., 2003; Gomot et al., 2002; Heaton, 2005; M. O'Riordan & Passetti, 2006), whereas conversely showed a speech-specific lexical tone processing difficulty in tone language speakers (Chen et al., 2016; Wang et al., 2017; Yu et al., 2015). Accumulating evidence pointed to a two-way transferability of pitch expertise across domains of music and speech (as lexical tones) in the neuro-typical children and adults (Bidelman et al., 2013; Chandrasekaran et al., 2009; Nan et al., 2018; Tang et al., 2016; Wong et al., 2007). By targeting the clinical population, the current data provided the first empirical evidence of using the relative strength of music, a ubiquitous nonspeech form, to compensate for the relative weakness of speech sounds especially lexical tone acquisition for tone language speaking children with ASD. One recent study (Nan et al., 2018) demonstrated that the six months of piano training not only enhanced the lexical tone discrimination, but also improved vowel and consonant discrimination in 4- to 5-year-old Mandarin-speaking TD children, suggesting strengthened common sound processing across domains underlying the benefits of musical training. In this training study, however, we failed to detect the benefits of music-based MMLI training on the acquisition of initials (consonants). On the one hand, the shorter, weaker, and more aperiodic consonants in speech sounds are likely to be impacted more in a co-occurring nonspeech background than the stable, periodic components of tones and vowels. On the other hand, since the syllable-initial consonants were acquired later than the vowels and tones in Mandarin-speaking

TD children (Hua & Dodd, 2000), the relatively short-term training in this study may be another potential factor leading to the failure of transfer effects on the late-acquired consonant production.

5.4.3 Training Efficacy for Nonverbal Children with ASD

The nonverbal participants in our study belong to the extremely low-functioning ones within the autism spectrum with severe language/cognitive impairment. They were completely nonverbal, despite having received extensive speech therapy (4-16 months) prior to recruitment. In the current study, except one nonverbal child with ASD responding to the MMLI treatment, the other five nonverbal participants could not correctly produce even one trained word after 24 training sessions (Figure 5.5), and meanwhile, they received pretty low scores on the user experience (Figure 5.6). It is unlikely due to the stringent measure of phonetic transcription, since these five nonverbal participants did not even spontaneously produce any verbal attempts during probe assessments. In contrast, as reported in Wan et al. (2011), the English-speaking nonverbal participants with autism who received similar MIT-based training, began to elicit some “word approximations” after 10-15 training sessions. It should be noted that the speech samples produced from the nonverbal participants in their study were imitations rather than spontaneous, should inform the degree and nature of progress (Wan et al., 2011). In our study, however, the spontaneous speech samples were collected from participants in a picture naming task without cueing or demonstration. Another possibility is that the nonverbal participants in our study were more severely impaired in terms of language and cognitive capacity compared to those in Wan et al. (2011). Given the extreme challenges these participants face, more patience and more training sessions should be delivered to the nonverbal children with ASD. Actually, both parents and therapists have observed an increase in the speechlike vocalizations during vocal play in daily life, which might be a precursor to speech development for these nonverbal participants.

5.4.4 Limitations

This study has several limitations. First, as with many other studies of autism, a limitation of the current training study is a small sample size, and replication in larger-scale randomized studies will be an important next step. Second, more training sessions should be implemented, to examine whether MMLI can reliably improve spoken language and articulation in the nonverbal children with ASD, and to check whether MMLI could also lead to a better performance in the acquisition of syllable-initial consonants relative to traditional control therapy. Third, in consideration of the huge heterogeneity of the autism spectrum, there are various types of speech disorder in ASD, such as developmental motor speech disorder (dysarthria or childhood apraxia of speech), consistent/inconsistent speech sound disorder, speech delay (alalia), or combinations of these (Chenausky et al., 2019). More in-depth investigation of speech therapy in different subtypes would help determine whether MMLI is effective for all the autistic children or only works well for certain subtypes. Understanding these mechanisms will help tailor the interventions, to select the most appropriate treatment that is personalized, and to make predictions about prognosis. Fourth, neuroimaging research will be necessary to examine the neural plasticity for autistic children which is induced with MMLI training, and to understand the neural processes underlying effective gains. Fifth, more future studies could help isolate the fundamental mechanisms (benefits attributable to the intonation and/or motor activities) leading to the effective gains from MMLI.

5.4.5 Clinical Implications

Taken together, the MIT-based training program of MMLI, notwithstanding its limitations, provided an effective training approach in accelerating the rate of word and speech sound acquisition, especially lexical tone acquisition for Mandarin-speaking children with ASD. The languages of the world exhibit great diversity. It has been suggest that around 60–70% of the world's languages are tonal (Yip, 2002), and more than 50% of the people use a tonal language (Fromkin, 1978). Thus, there is a high demand for MMLI, which could be modified and applied to help some other tone-language-speaking children with autism beyond Mandarin-speaking ones to better acquire the phonological category of lexical tones. With respect to the practical significance, the current MMLI approach is realized in the smartphone/iPad App, which is easily accessible and has the potential to be utilized remotely in the home environment as implemented by a parent or family member. This is important for speech therapy in autistic children from counties where the speech-language pathologists are in shortage. Finally, the success of MIT-based MMLI also lends support to the positive effects of music-based treatments in individuals with ASD (James et al., 2015; Reschke-Hernández, 2011; Salomon-Gimmon & Elefant, 2019; Sharda et al., 2018).

5.5 Conclusion

Using a randomized controlled design, this study compared the efficacy of MMLI, an MIT-based treatment, and traditional therapy in eliciting spoken language for tone-language-speaking children with autism. Relative to the control treatment, Mandarin-speaking with ASD showed higher improvement after receiving the MMLI training in terms of the lexical tone, final, and word acquisition in the trained items. Such enhanced training efficacy on lexical tone production remained at two weeks post-therapy, and even generalized to novel items that were not practiced. The results hold promise for the efficacy of MMLI to improve speech production in tone-language-speaking children with autism. Because the low-functioning children with autism had a very limited repertoire of speech sounds prior to treatment, the acquisition of speech sounds and words through MMLI is an important gain that provides a foundation for subsequent speech and language rehabilitation. More importantly, this study offers the first empirical evidence of utilizing the musical elements to facilitate lexical tone acquisition in the clinical population of ASD, which adds a new clinical perspective to our understanding of the close relationship between music and speech.

Chapter 6: General Discussion and Conclusions

Compared to non-tonal language speakers with ASD, tonal language users with ASD have been disproportionally under-explored in language and speech research. Motivated by the phonological roles of pitch in tone languages (such as Mandarin and Cantonese), recent studies with Chinese people with ASD have provided new information in this field. Following this line, this dissertation further investigated auditory and speech processing atypicalities and the treatment in tone-language-speaking individuals with ASD, with two behavioral studies (Chapters 3&4) and one treatment study (Chapter 5) included. As each language employs a unique set of phonological features, speech perception and production atypicalities in ASD may take different forms and mechanisms depending on the acoustic dimension (spectral or temporal), sound type (speech or nonspeech), and language background (tonal or non-tonal; tonal languages with different inventories). This dissertation directly compared the spectral vs. temporal processing, speech vs. nonspeech processing, as well as native vs. non-native lexical tone imitation among different studies. Furthermore, the training study made use of music, a ubiquitous nonspeech form, to help improve speech sound acquisition in tone-language-speaking children with autism, indicating a domain-transferred effect from nonspeech to speech domain. This chapter will provide a conclusion of the main findings as well as theoretical and practical implications. Limitations and future directions would also be discussed.

6.1 On the Nature of Speech Processing Difficulties in Tone Language Speakers with ASD

Previous studies have reported lexical tone developmental delays (Wu et al., 2020) and lexical tone perception difficulties (Chen et al., 2016; Wang et al., 2017; Yu et al., 2015) in tone language speakers with ASD. In line with Wu et al. (2020), the low-functioning Mandarin-speaking children with ASD who experienced severe language and cognitive delays indeed showed significantly lower production accuracy scores of consonants, vowels, and lexical tones relative to age-matched TD controls (Table 5.1). Given that the overall language ability and cognitive levels interfere with speech processing and acquisition (Bartolucci et al., 1976; Schoen et al., 2011; Wolk & Brennan, 2013; Wolk & Giesen, 2000), the first two behavioral studies in this dissertation were specially performed in high-functioning individuals with ASD without severe language/cognitive delays, and the relevant confounding factors were entered as covariates.

Different from previous CP findings (Chen et al., 2016; Wang et al., 2017), the high-functioning adolescents with ASD in Study 1 did perceive the lexical tones and VOT in a preserved CP pattern, as indicated by a much higher sensitivity to between-category pairs relative to within-category pairs in both types of continua (Figures 3.4 & 3.6). The preserved CP pattern in ASD was also found in the CP of vowels and consonants in both high-functioning children with autism and Asperger syndrome (You et al., 2017), and in high-functioning adults with ASD (Stewart et al., 2018). Further regression analyses showed that during the CP of lexical tones in ASD, the overall language ability was a significant predictor for the boundary width; the capacity of phonological working memory (digit span) was a contributing factor for the identification and discrimination of the lexical tones (Table 3.3). The close relationship between CP competence and language function was also observed in ASD with various cognitive abilities and different age ranges (Bishop et al., 2004; Chen et al., 2016; Constantino et al., 2007; Stewart et al., 2018). As shown by the above evidence, the impaired CP might not be observed in all the autistic individuals, but rather tend to be part of a shared vulnerability of language and/or cognitive impairment in a subgroup of low-functioning ASD. Similarly, some of the previous studies proposed that autistic children lack the flexible and complex expression of pitch variation (Bonneh et al., 2011; DePape et al., 2012; Green & Tobin, 2009), and they tend to make adjustment when they echoed speech (Paccia & Curcio, 1982; Pronovost et al., 1966). However, the current observation (Study 2) on lexical tone imitation showed that both Cantonese-speaking (native) and Mandarin-speaking (non-native) children with ASD could accurately imitate the global tone contours for the three Cantonese level tones (CT55, CT33, CT22) and three Cantonese contour tones (CT23, CT25, CT21), important for contrasting tonal categories (Figure 4.4). Even for the fine-grained 'growth curve analysis' (Mirman, 2014) in terms of pitch height, slope, and curvature, both native and non-native ASD groups only produced a shallower slope and/or a flatter F0 curve in the acoustic realizations of high-rising CT25 compared with TD groups. However, such allophonic differences did not lead to perceptual ambiguities, as evidenced by a comparable accuracy of identifying the imitative CT25 stimuli produced by ASD and TD children in native Cantonese adult perceivers (Figure 4.6). In a short conclusion, both CP of lexical tones and the complex tone imitation skills were largely preserved in high-functioning tone-language-speaking individuals with ASD.

However, the current observations of this dissertation also revealed some autism-associated speech processing difficulties, even in the subgroup of high-functioning ASD without severe intellectual disability/language disorder. First, although Mandarin-speaking adolescents with ASD showed a preserved CP of VOT with cross-boundary benefit, the degree of CP of VOT in ASD was greatly reduced relative to TD controls. This was reflected by a much wider boundary width (Figure 3.6c) and lower peakedness score (Figure 3.6e) in the ASD group during the perception of VOT continuum. In contrast, the degree CP of native lexical tones was typical-like in the Mandarin-speaking adolescents with ASD (Figures 3.3&3.5). Consequently, we speculated that the impaired auditory processing deficits of sound duration in autism are manifested profoundly and further persist into the higher-level phonological processing that involves the basic CP competence of VOT. Future training studies should aim specially at enhancing the perception and production of VOT among autistic individuals. Second, the ASD group showed a much higher boundary position (i.e., closer to the level tone) relative to TD controls in the identification of lexical tones in real word condition (Figure 3.2b), with a similar pattern called "psychophysical boundary" also observed in the non-tonal language speakers who had no tonal language experience (Wang, 1976). Such an atypical boundary shift in ASD might reflect a weaker influence from the higher-level semantic capture when performing the lexical tone identification task. Since the impaired word meaning processing has been found in individuals with ASD (Coffey-Corina et al., 2008; Henderson et al., 2011), it is hypothesized that lexical tone perception impairment in ASD might be related to impairment in lexical-semantic access, rather than associated with the basic acoustic pitch perception. Third, when imitating the non-native and unfamiliar tonal category of CT23, the Mandarin-speaking children with ASD failed to utilize the phonological knowledge of familiar segmental information (/fu/ and /ji/) to produce a better voice quality of that syllable, but TD children could. This point was reflected by relatively lower accuracy in the identification of CT23 stimuli with familiar segment, which were produced by MASD than those produced by MTD (Figure 4.6). Yet, MTD and MASD produced overlapping acoustic F0 realizations when imitating CT23 with familiar segment. Thus, these findings implied that lexical tone imitation difficulties were caused by deficient phonological processing of the carrying syllables rather than the low-level acoustic pitch imitation deficit.

The aforementioned autism-associated speech processing difficulties (reduced degree of CP of VOT; impaired semantic capture during lexical tone identification; deficient phonological processing of the carrying segments during lexical tone imitation) could be potentially explained by the atypical hemispheric lateralization in ASD for the processing of speech sounds and linguistic content. As mentioned, the accurate and complete perception of speech sounds in native speakers involves both the acoustic processing and phonological processing (Lancker, 1980). There has been accumulating evidence that although phonetic processing (including the lexical tone processing) engages both hemispheres in the neuro-typical brain, the phonological and semantic processing is left lateralized, and the pure acoustic processing tends to be processed more in the right hemisphere (Gandour et al., 2004; Hickok & Poeppel, 2007). Moreover, the temporal information (e.g., duration) was mainly processed in the left hemispheric, and spectral information (e.g., pitch, formant) processing was preferably right lateralized in human auditory cortices (Boemio et al., 2005; Schönwiesner et al., 2005; Zaehle et al., 2004). A series of auditory neuroimaging studies have found atypical patterns of either left hemisphere deficits and/or right hemisphere dominance in individuals with ASD (Bruneau et al., 1999; Gage et al., 2003; Haesen et al., 2011; Kasai et al., 2005; Mason et al., 2008; Roberts et al., 2008). The failure of left hemisphere activation or even rightward asymmetry might be responsible for a wide range of speech and language processing atypicalities in ASD (Eyler et al., 2012; Haesen et al., 2011; Lindell & Hudry, 2013). Thus, the neural specialization for auditory and speech processing in ASD is altered in a way that auditory spectral processing (right hemisphere function) is enhanced, but temporal processing (left hemisphere function) is impaired (Groen et al., 2009; Haesen et al., 2011; Huang et al., 2018); non-linguistic and acoustic pitch processing (right hemisphere function) is enhanced or preserved, whereas phonological and semantic processing (left hemisphere function) is impaired (Jiang et al., 2015; Wang et al., 2017; Yu et al., 2015).

6.2 On the Speech-Specific Pitch Processing Atypicalities in Tone Language Speakers with ASD

When the pitch information was superimposed on the spectrally and temporally complex speech stimuli, there is a trend of distinct patterns across different language speakers with ASD (i.e., tone language vs. non-tone language), pointing to a language-specific pitch processing pattern in ASD. For non-tonal language speakers, the speech pitch was perceived non-phonemically. Both behavioral and ERP studies have revealed a domain-general account of pitch processing superiority in non-tonal language speakers with ASD (Haesen et al., 2011; Pamela Heaton, Hudry, et al., 2008; Järvinen-Pasley, Pasley, et al., 2008; Järvinen-Pasley & Heaton, 2007; Kujala et al., 2010; Lepistö et al., 2005, 2006). For tone language speakers, the syllable-level pitch processing in speech (i.e., lexical tone processing) incorporates acoustic, phonetic, and phonological analyses, which was more demanding and complex. Two recently conducted studies (Wang et al., 2017; Yu et al., 2015) proposed a 'speech-specific' lexical tone perception difficulties in tone-language-speaking individuals with ASD. The related findings in this dissertation further lend support to this viewpoint, with extended evidence from high-functioning individuals with ASD, from two different tonal language backgrounds, and from both perception and production performance.

The related findings in this dissertation have observed three phenomena of domain-specificity in terms of pitch processing (Table 6.1). Firstly, although the boundary width (Figure 3.3) and peakedness score (Figure 3.5) did not differ between ASD and TD groups among speech and nonspeech pitch carriers, the boundary position (Figure 3.2) differed between two groups only in the speech condition of real word. Specifically, Mandarin-speaking adolescents with ASD showed a higher boundary position (i.e., closer to the level end) than TD controls in the real word condition, but similar boundary positions were shown in the nonspeech condition of IRN and pure tone. Secondly, both Mandarin-speaking and Cantonese-speaking children with ASD showed increased pitch range/SD than TD peers when imitating the speech tones, while exhibited comparable pitch range/SD when imitating the nonspeech sounds (Figure 4.2). That's to say, the atypical prosodic pitch pattern of increased variability was speech-specific. Thirdly, the fine-grained acoustic analyses showed that both native and non-native children with ASD showed some within-category deviations from the TD controls in coordinating pitch slope and curvature when imitating the high-rising pitch (CT25) embedded in speech. However, the two groups (ASD vs. TD) performed similarly in terms of pitch height/slope/curvature when imitating all the pitch contours embedded in nonspeech (Tables 4.3&4.4). Taken together, even high-functioning tone language speakers with ASD often show atypicalities in the lexical tone perception and imitation tasks, but perform as similar as TD controls when processing the nonspeech pitch counterparts. Based on relevant findings in pitch processing from different language backgrounds, more and more research proposed a speech-specific and language-specific viewpoint that tone language speakers with ASD fail to engage or develop specialized networks for lexical tone or intonation processing in speech context (this dissertation; Jiang et al., 2015; Wang et al., 2017; Yu, 2018; Yu et al., 2015)

Table 6.1 *Results summary of perception and imitation of pitch contours in the ASD group in comparison with the TD group in speech and nonspeech conditions.*

Condition	**Study 1**: CP of Lexical tones			**Study 2**: Lexical Tone & Non-linguistic Pitch Imitation					
	Pitch Perception of Real Word			Prosodic Pitch Pattern			Imitation of CT25		
	Position	Width	Peakedness	Mean	Range	SD	Height	Slope	Curvature
Speech	>	≈	≈	>	>	>	≈	<	<
Nonspeech	≈	≈	≈	>	≈	≈	≈	≈	≈

Note**:** >, *ASD > TD*; <, *ASD < TD*; ≈, *no difference.*

6.3 On the Theories Explaining Atypical Auditory and Speech Processing in ASD

The relationship between acoustic and phonological processing has been discussed for a long time. Some neuroimaging studies implied that they were represented differently in our human brain, with evidence from different processing areas along the auditory pathways (Okada et al., 2010; Wessinger et al., 2001; Zhang et al., 2011), and different patterns of brain lateralization (Gandour et al., 2004). However, more and more studies have pointed to a bidirectional interaction between acoustic analysis and phonological processing in speech perception (Hickok & Poeppel, 2007; Zhang et al., 2011). Especially, at the basic auditory processing level, autistic individuals showed unbalanced auditory processing capacity depending on the acoustic dimensions (spectral vs. temporal) (Alcántara et al., 2012; Groen et al., 2009; Yu, 2018), with enhanced acoustic processing of pitch (Bonnel et al., 2003; Foxton et al., 2003; Heaton et al., 2008; Heaton, 2003, 2005; Mottron et al., 2000; O'Riordan & Passetti, 2006), while reduced acoustic processing of sound duration (Brodeur et al., 2014; Falter et al., 2012; Maister & Plaisted-Grant, 2011; Martin et al., 2010; Szelag et al., 2004). By using the classic CP paradigm, this dissertation compared the phonological processing of linguistic pitch (lexical tone) and linguistic time (VOT) in native Mandarin speakers with ASD at the same time, which can help uncover whether and how lower-level acoustic processing could exert an influence on higher-level phonological processing. If the higher-level phonological processing—developed early in native speakers—is robust enough to resist the influence from acoustic sensitivity, it is possible that the degree of CP of lexical tone and VOT would be comparable in ASD. Alternatively, if the unbalanced acoustic processing capacity further extends to the phonological processing of native phonemes, it is likely that Mandarin-speaking adolescents with ASD would show different degrees of CP of lexical tone vs. VOT. The current findings fit with the latter speculation that the degree of CP of lexical tone in ASD was typical-like while the degree of CP of VOT was greatly reduced. The 'feed-forward mechanisms' (Binder, 2000; Scott & Wise, 2004) suggested that speech processing begins from the dorsal STG to downstream brain areas and then to more lateral and anterior regions, indicating that initial bottom-up acoustic processing might lay the foundation of phonological processing. Thus, the current findings provide behavioral evidence on the 'feed-forward mechanisms' from the clinical population of ASD.

A large body of research was dedicated to clarifying and refining theories of the communication deficits in ASD. As mentioned in the Introduction part, there were primarily three cognitive theories to explain the perceptual performance in ASD: a) ‘Weak Central Coherence’ (WCC) theory (Frith, 1989), b) ‘Social Theory’ (O’connor, 2012), and c) ‘Complexity Hypothesis’ (Bertone et al., 2005). According to these theories, the difficulties individuals with ASD experience may not be caused solely by the social relevance of the stimuli, the local/global level of processing, or the complexity of the presented information. A full understanding of the cognitive explanation of information processing in ASD is still needed with more research from different modalities, and in a wider range of participants with ASD from different language backgrounds. Related findings in this dissertation regarding lexical tone processing in ASD may shed light on this issue to some extent, by manipulating the pitch carriers with varying levels of spectro-temporal complexity or phonemic/semantic relevance (Table 3.2). In terms of pitch perception in Study 1, Mandarin-speaking participants with ASD showed a higher boundary position (i.e., closer to the level end) relative to TD controls only in the ‘real word’ condition, indicating a compressed perceptual space for the high-level tone in ASD group. This might be attributed to the weaker influence from semantic activation of the real word “eight” with high-level tone in ASD. In terms of pitch imitation in Study 2, the non-native Mandarin-speaking children with ASD failed to exploit the top-down phonological knowledge to compensate for the imitation of syllables with non-native tonal category. Taken together, these findings from both lexical tone perception and imitation revealed the global and top-down processing deficiency in ASD. Such findings might be related to the ‘Weak Central Coherence’ theory (Frith, 1989; Happé & Frith, 2006) from auditory modality. Furthermore, the ‘Social Theory’ (O’connor, 2012) could be used to explain the speech-specific pitch processing atypicalities in ASD to some extent, since the signals of spoken language contain relatively more social information relative to the nonspeech materials. But future studies should focus on the processing of more socially relevant speech types such as IDS to examine the potential link between social competence and speech processing among autistic individuals. Finally, as shown in Study 1, the autistic participants exhibited better within-category pitch discrimination performance in the pure tone condition, compared with the other three types of pitch carriers (real word, nonword, and IRN), which were more complex in terms of spectro-temporal complexity. Such findings might shed light on the 'Neural Complexity Hypothesis' (Samson et al., 2006), which proposes that individuals with ASD tend to perform better with spectro-temporally simple sounds, but have difficulties in dealing with spectro-temporally complex auditory information.

6.4 On the Effectiveness of Music-Assisted Speech Therapy in Tone Language Speakers with ASD

Music therapy has been regarded as a promising intervention strategy for individuals with ASD to promote communication, social-emotional, perceptuo-motor, and behavioral skills (Gold et al., 2006; James et al., 2015; LaGasse, 2017; Reschke-Hernández, 2011; Srinivasan & Bhat, 2013). Although music and speech belong to different domains with different neural representations, an increasing number of neuroimaging studies have detected a large neural overlap in responses to speech and musical stimuli (see Peretz et al. 2015 for a review), which implies a close relationship between music and speech processing. Indeed, several studies in literature have proved the efficacy of using music-assisted speech therapy to facilitate speech sound acquisition for non-tonal language speakers with ASD (Chenausky et al., 2016, 2017; Hoelzley, 1993; Miller & Toca, 1979; Sandiford et al., 2013; Wan et al., 2010). For tone-language-speaking individuals with ASD, most of them showed difficulties in lexical tone processing and acquisition beyond vowel and consonant

(Chen et al., 2016; Wang et al., 2017; Wu et al., 2020; Yu et al., 2015). Even for the high-functioning children and adolescents with ASD, the first two studies in this dissertation found speech-specific lexical tone processing atypicalities in tone language speakers with ASD. However, both the pitch perception and imitation skills were typical-like when the pitch contours were embedded in various nonspeech materials. Moreover, several behavioral and neuroimaging studies have shown that individuals with ASD have always demonstrated a pitch perception superiority in various musical and nonspeech stimuli compared to TD controls (Bonnel et al., 2010; Ferri et al., 2003; Foxton et al., 2003; Gomot et al., 2002; Heaton, 2005; O'Riordan & Passetti, 2006). Since both music (a ubiquitous nonspeech form) and lexical tone (speech form) share the same psycho-acoustical attribute of pitch, it would be reasonable take advantage of the relative strength of musical skills to compensate for the relative weakness of speech sound especially lexical tone acquisition for tone language speaking children with ASD.

To fill this research gap, Study 3 in this dissertation used a randomized controlled trial to evaluate the efficacy of MMLI, an MIT-based treatment for facilitating spoken language in Mandarin-speaking nonverbal and low-verbal children with ASD, in comparison to a matched non-MIT-based control treatment. Compared with control treatment, MMLI additionally involved musical making activities by hand tapping on the piano icons and the intoned verbal stimuli of piano-timbre nonspeech. Specifically, the piano-timbre nonspeech contained the same pitch contours of natural lexical tones for the first and second syllables of each word. Training results showed that the music-assisted training approach of MMLI not only increased the learning motivation and attention in ASD (Figures 5.6&5.8), but also leaded to a faster rate of improvement in the acquisition of words, vowels, and especially lexical tones in the trained items (Figures 5.5&5.7). Besides, the efficacy of MMLI was much higher in the training of lexical tones relative to vowels and consonants, since the benefits of MMLI on lexical tone training could generalize to the untrained novel items. The current data provided the first empirical evidence of using a music-assisted training approach to successfully accelerate the lexical tone and other speech sound acquisition in tone language speakers with ASD. Further, the domain-transferred effect from nonspeech pitch to speech tone acquisition in ASD adds new evidence to the transferability of pitch processing across the domains of music and lexical tones. Since more than half of the world's population speak a tonal language (Fromkin, 1978), there is a high demand for the music-assisted MMLI, which could be modified and applied to aid some other tone-language-speaking children with ASD to better acquire the lexical tones.

6.5 Limitations and Future Directions

The limitations for the three studies of this dissertation have been discussed in each chapter, respectively. Besides, this dissertation has some other outstanding limitations, and there are still many unsolved research gaps that merit further investigations in terms of auditory and speech processing mechanisms and treatment especially in tone language speakers with ASD.

First, it is important to note that not all the current findings in this dissertation may apply to each individual on the autistic spectrum with severe language, cognitive, or adaptive deficits, and with different age groups. One major weakness of the three studies in this dissertation is the heterogeneity of the subject demographic profiles with relatively narrow and non-overlapping age groups and cognitive/linguistic abilities. Recently, there has been a concerted effort to parse the huge heterogeneity of the autistic spectrum into meaningful subgroups (Happé & Frith, 2020; Zheng et al., 2019), and the same approach should be introduced in the research field of auditory and speech processing in ASD. In this regard, it would be necessary and meaningful to explore how the speech processing capacity (such as the basic CP of speech sounds) was varied among different subgroups of ASD and different age groups in future work.

Second, all the three experiments in this dissertation are behavioral studies. Thus, some of the cited references regarding neurophysiological data might not be proper to support the current behavioral findings. For instance, speech training studies have shown that the mismatch response is an index of neural sensitivity to sound contrasts without requiring focused attention or judgment, which may not necessarily be consistent with behavioral responses that depend on attentional processing for proper judgment. In particular, with the help of neuroimaging techniques, future research is needed to uncover the neurophysiological mechanisms underlying speech and lexical tone processing in tone language speakers with ASD, as well as its potential relationship with behavioral correlates.

Third, this dissertation only focused on the syllabic-level speech processing of lexical tones and VOT in ASD and free from noises and social functions. On the one hand, future work should extend the syllabic speech processing to the sentence level, such as investigating the role of tone contour in sentence perception for tone language speakers with ASD. In many cases, the optimal and quiet listening condition during auditory and speech processing is not always guaranteed for individuals with ASD. The capacity to extract acoustic or phonetic information from a target speaker among a background of competing speakers or environmental noise is modulated by sensory-cognitive interaction such as attention, working memory, and language (Anderson & Kraus, 2010). Such capacity has been found to be struggling for ASD reported in very limited studies (Alcántara et al., 2004; Mamashli et al., 2017; Russo et al., 2009). Future studies need to compare the speech and nonspeech (such as lexical tone and nonspeech counterparts) perception in ASD with background speech noise containing spectral and/or temporal dips. On the other hand, the socially-relevant speech perception, such as the processing of IDS, might act as one of the potential early markers of risk for ASD in the auditory domain (Filipe et al., 2018). Although in literature, there was some behavioral evidence for reduced attention to IDS in children with ASD (e.g., Filipe et al., 2018; Klin, 1991; Kuhl et al., 2005; Paul et al., 2007), the underlying neurophysiological bases for the between-group differences in processing different aspects of IDS-specific acoustic and phonetic features still remain unclear. Future studies in the field of auditory and speech processing need to pay more attention to the neural representation of vowel formant exaggeration in IDS, as well as the prosodic change in IDS and its interaction with lexical tone variation for tone language speakers with ASD.

Fourth, isolated lexical tone processing has been well studied (this dissertation; Chen et al., 2016; Wang et al., 2017; Yu et al., 2015). However, in real life, we seldom perceive lexical tones in isolation at the syllable level. To the best of our knowledge, no study has ever focused on the effects of contextual information on target tone processing in tone language speakers with ASD. In the actual speech communication, the exact F0 values of lexical tones are highly variable across utterances and talkers. The term “tone normalization” has been used to describe the processes by which listeners recognize the same tone produced by different talkers or the same talker in different conditions (Francis et al., 2006).). Previous studies have consistently demonstrated a ‘contrastive context effect’ in TD individuals on the processing of both Cantonese level tones (Francis et al., 2006; Wong & Diehl, 2003; Zhang et al., 2013), as well as Mandarin contour tones (Chen & Peng, 2016; Huang & Holt, 2009). Relevant for this aspect is the WCC theory, which postulates that people with ASD might be impaired in extracting and combining contextual information (Foxton et al., 2003; Frith, 1989; Happé & Frith, 2006). It is interesting to further investigate whether tone language speakers with ASD would show deficits in extracting the contextual speech or nonspeech pitch information to facilitate tone normalization.

Fifth, in the current training study (Study 3), although the music-assisted MMLI showed a superior training efficacy compared to the traditional method, some individuals with ASD especially the nonverbal ones showed no improvement after receiving MMLI. More in-depth investigation of speech therapy using MMLI in different subtypes would help determine whether MMLI is effective for all the autistic children or only works well for certain subtypes. For these small subgroups of children with ASD, consequently, other styles of training approach should be explored and supported to make the training user-adaptive and skill-adaptive, which would deepen our understanding of the learning mechanisms in ASD. Also, the training efficacy of MMLI should be replicated in larger-scale randomized studies and should be applied to some other tone language speakers with ASD beyond Mandarin-speaking children with ASD.

6.6 Conclusions

This dissertation directly compared the spectral vs. temporal processing, speech vs. nonspeech processing, as well as native vs. non-native lexical tone imitation in tone language speakers with ASD. The degree of CP was much higher for the perception of lexical tones than the perception of VOT in ASD, reflecting the influence from lower-level acoustic processing of pitch and time. Moreover, the relevant findings in this dissertation further lend support to the notion of speech-specific lexical tone processing difficulties in ASD, with extended evidence from two different tonal language backgrounds (Mandarin and Cantonese), and from both perception and imitation performance. The bottom-up acoustic pitch processing was largely intact in ASD during the perception as well as imitation of pitch contours, while even the high-functioning individuals with

ASD without severe language/cognitive delay showed deficits at the top-down phonological processing. Taken together, speech processing atypicality in ASD can be not only cue-specific and domain-specific but also language-specific. Finally, the treatment study offers the first empirical evidence of utilizing the musical elements to facilitate lexical tone acquisition in the clinical population of ASD, which deepened our understanding of the domain-transferred effect between music and speech.

Appendices

Appendix A. The Items Used in the Task of Forward Digit Span

Item	Trial	Test	Response	Score
1	1 2	4-7 3-0		
2	1 2	5-8-4 7-6-8		
3	1 2	3-1-6-0 4-8-9-7		
4	1 2	2-0-5-8-6 2-3-7-9-4		
5	1 2	5-1-2-0-6-7 9-3-6-7-2-8		
6	1 2	2-9-0-3-4-7-6 8-6-2-4-7-0-3		
7	1 2	2-4-1-8-6-7-9-0 1-4-6-2-7-9-3-8		
8	1 2	5-6-2-4-7-9-0-3-1 4-3-6-5-9-8-5-6-0		

Appendix B. The Items and Their Corresponding *Pinyin* Transcriptions in the Nonword Repetition Task

One Syllable		Two Syllables		Three Syllables	
Practice	*miang4*	Practice	*zuai3miang4*	Practice	*rei3fao2shong 1*
	tia4		*cuang4ria3*		*tei1zuai3miang4*
	ruang2		*fao2shong1*		*suai4chang4ruai3*
Trial 1	*rai4*	Trial 1	*bong1nua2*	Trial 1	*sua3piong4buai1*
Trial 2	*fao3*	Trial 2	*shong4fao3*	Trial 2	*piang3fian4suai3*
Trial 3	*fiang1*	Trial 3	*suang2fiang1*	Trial 3	*sei3zuai3miang4*
Trial 4	*fong4*	Trial 4	*tiang3fong4*	Trial 4	*nua3cuang4ria3*
Trial 5	*niong2*	Trial 5	*bia4mong3*	Trial 5	*tiang3fao2shong1*
Trial 6	*pia4*	Trial 6	*liong3tiu4*	Trial 6	*cua3mia2zua3*
Trial 7	*fiong3*	Trial 7	*tia4fiong3*	Trial 7	*nia1diong1ruai4*
Trial 8	*zuang4*	Trial 8	*fong1tiang2*	Trial 8	*liong4diang3rua4*
Trial 9	*diang4*	Trial 9	*rua1diang4*	Trial 9	*tiong2tua1lua2*
Trial 10	*biong2*	Trial 10	*nia1biong2*	Trial 10	*tiu1ria2cuang3*
Trial 11	*fie2*	Trial 11	*tei1fie2*	Trial 11	*suai4nua3bong4*
Trial 12	*ria2*	Trial 12	*miang1zuai2*	Trial 12	*rei3rei3cuai4*
Trial 13	*fua3*	Trial 13	*tiong2fua3*	Trial 13	*fian1chei3pua2*
Trial 14	*mong2*	Trial 14	*buai4piong3*	Trial 14	*bia3piu2cei1*
Trial 15	*mua1*	Trial 15	*miong2zuang1*	Trial 15	*zua2biu2niong3*
Trial 16	*ruai1*	Trial 16	*fiang4suang1*	Trial 16	*biang4pua1chei2*
Trial 17	*mia1*	Trial 17	*zua2mia1*	Trial 17	*diong2niong2biu1*
Trial 18	*cuang3*	Trial 18	*piang3bua2*	Trial 18	*rua1biong1nia4*
Trial 19	*buai1*	Trial 19	*piong4buai1*	Trial 19	*fia2biang3mua4*
Trial 20	*zuai3*	Trial 20	*fian4suai3*	Trial 20	*miong3ruang2dua1*

Note: Numbers show an example of a minimal quartet (Tones 1-4), using a common transcription convention of suffixing the tonal number to the *Pinyin* transcription of each syllable.

Appendix C. Word Lists of Six Themes for Picture-Naming Test

(a) Theme I: Vegetables

Set 1: Trained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
土豆 (potato)	*t* [t^{h}]	*u* [u]	**T3 [21]**	*d* [t]	*ou* [ou]	T4 [51]
茄子 (eggplant)	*q* [tɕ^{h}]	*ie* [iɛ]	T2 [35]	*z* [ts]	-*i* [ɿ]	[neutral tone]
白菜 (cabbage)	*b* [p]	*ai* [ai]	T2 [35]	*c* [ts^{h}]	*ai* [ai]	T4 [51]
萝卜 (radish)	*l* [l]	*uo* [uo]	T2 [35]	*b* [p]	*o* [o]	[neutral tone]
辣椒 (chilli)	*l* [l]	*a* [ᴀ]	T4 [51]	*j* [tɕ]	*iao* [iɑu]	T1 [55]
玉米 (corn)	Ø	*ü* [y]	T4 [51]	*m* [m]	*i* [i]	T3 [214]
大蒜 (garlic)	*d* [t]	*a* [ᴀ]	T4 [51]	*s* [s]	*uan* [uan]	T4 [51]
黄瓜 (cucumber)	*h* [x]	*uang* [uaŋ]	T2 [35]	*g* [k]	*ua* [uᴀ]	T1 [55]
木耳 (agaric)	*m* [m]	*u* [u]	T4 [51]	Ø	*er* [ɚ]	T3 [214]
芹菜 (celery)	*q* [tɕ^{h}]	*in* [in]	T2 [35]	*c* [ts^{h}]	*ai* [ai]	T4 [51]
Set 2: Untrained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
南瓜 (pumpkin)	*n* [n]	*an* [an]	T2 [35]	*g* [k]	*ua* [uᴀ]	T1 [55]
洋葱 (onion)	Ø	*iang* [iaŋ]	T2 [35]	*c* [ts^{h}]	*ong* [ʊŋ]	T1 [55]

(b) Theme II: Fruits

Set 1: Trained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
苹果 (apple)	*p* [p^{h}]	*ing* [iŋ]	T2 [35]	*g* [k]	*uo* [uo]	T3 [214]
香蕉 (banana)	*x* [ɕ]	*iang* [iaŋ]	T1 [55]	*j* [tɕ]	*iao* [iɑu]	T1 [55]
葡萄 (grape)	*p* [p^{h}]	*u* [u]	T2 [35]	*t* [t^{h}]	*ao* [ɑu]	[neutral tone]
橙子 (orange)	*ch* [tʂ^{h}]	*eng* [əŋ]	T2 [35]	*z* [ts]	-*i* [ɿ]	[neutral tone]
西瓜(watermelon)	*x* [ɕ]	*i* [i]	T1 [55]	*g* [k]	*ua* [uᴀ]	T1 [55]
草莓 (strawberry)	*c* [ts^{h}]	*ao* [ɑu]	**T3 [21]**	*m* [m]	*ei* [ei]	T2 [35]
芒果 (mango)	*m* [m]	*ang* [aŋ]	T2 [35]	*g* [k]	*uo* [uo]	T3 [214]
菠萝 (pineapple)	*b* [p]	*o* [o]	T1 [55]	*l* [l]	*uo* [uo]	T2 [35]
石榴(pomegranate)	*sh* [ʂ]	-*i* [ʅ]	T2 [35]	*l* [l]	*iu* [iou]	[neutral tone]
桂圆 (longan)	*g* [k]	*ui* [uei]	T4 [51]	Ø	*üan* [yæn]	T2 [35]
Set 2: Untrained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone

柠檬 (lemon)	*n* [n]	*ing* [iŋ]	T2 [35]	*m* [m]	*eng* [əŋ]	T2 [35]
樱桃 (cherry)	Ø	*ing* [iŋ]	T1 [55]	*t* [t^h]	*ao* [ɑu]	T2 [35]

(c) Theme III: Animals

Set 1: Trained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
老虎 (tiger)	*l* [l]	*ao* [ɑu]	**T3 [35]**	*h* [x]	*u* [u]	T3 [214]
大象 (elephant)	*d* [t]	*a* [ᴀ]	T4 [51]	*x* [ɕ]	*iang* [iaŋ]	T4 [51]
熊猫 (panda)	*x* [ɕ]	*iong* [jʊŋ]	T2 [35]	*m* [m]	*ao* [ɑu]	T1 [55]
斑马 (zebra)	*b* [p]	*an* [an]	T1 [55]	*m* [m]	*a* [ᴀ]	T3 [214]
蝴蝶 (butterfly)	*h* [x]	*u* [u]	T2 [35]	*d* [t]	*ie* [iɛ]	T2 [35]
乌龟 (tortoise)	Ø	*u* [u]	T1 [55]	*g* [k]	*ui* [uei]	T1 [55]
嗡嗡 (buzz)	Ø	*ueng* [uəŋ]	T1 [55]	Ø	*ueng* [uəŋ]	T1 [55]
孔雀 (peacock)	*k* [k^h]	*ong* [ʊŋ]	**T3 [21]**	*q* [tɕ^h]	*üe* [yɛ]	T4 [51]
蚊子 (mosquito)	Ø	*uen* [uən]	T2 [35]	*z* [ts]	*-i* [ɿ]	[neutral tone]
企鹅 (penguin)	*q* [tɕ^h]	*i* [i]	**T3 [21]**	Ø	*e* [ɤ]	T2 [35]
Set 2: Untrained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
兔子 (rabbit)	*t* [t^h]	*u* [u]	T4 [51]	*z* [ts]	*-i* [ɿ]	[neutral tone]
小狗 (puppy)	*x* [ɕ]	*iao* [iɑu]	**T3 [35]**	*g* [k]	*ou* [ou]	T3 [214]

(d) Theme IV: Daily Necessities

Set 1: Trained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
牙刷 (toothbrush)	Ø	*ia* [iᴀ]	T2 [35]	*sh* [ʂ]	*ua* [uᴀ]	T1 [55]
毛巾 (towel)	*m* [m]	*ao* [ɑu]	T2 [35]	*j* [tɕ]	*in* [in]	T1 [55]
筷子 (chopsticks)	*k* [k^h]	*uai* [uai]	T4 [51]	*z* [ts]	*-i* [ɿ]	[neutral tone]
水杯 (glass)	*sh* [ʂ]	*ui* [uei]	**T3 [21]**	*b* [p]	*ei* [ei]	T1 [55]
拖把 (mop)	*t* [t^h]	*uo* [uo]	T1 [55]	*b* [p]	*a* [ᴀ]	T3 [214]
马桶 (toilet)	*m* [m]	*a* [ᴀ]	**T3 [35]**	*t* [t^h]	*ong* [ʊŋ]	T3 [214]
枕头 (pillow)	*zh* [tʂ]	*en* [ən]	**T3 [21]**	*t* [t^h]	*ou* [ou]	[neutral tone]
手套 (gloves)	*sh* [ʂ]	*ou* [ou]	**T3 [21]**	*t* [t^h]	*ao* [ɑu]	T4 [51]
衣架 (hanger)	Ø	*i* [i]	T1 [55]	*j* [tɕ]	*ia* [iᴀ]	T4 [51]
台灯 (lamp)	*t* [t^h]	*ai* [ai]	T2 [35]	*d* [t]	*eng* [əŋ]	T1 [55]
Set 2: Untrained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
雨伞 (umbrella)	Ø	*ü* [y]	**T3 [35]**	*s* [s]	*an* [an]	T3 [214]

闹钟(alarm clock)	*n* [n]	*ao* [ɑu]	T4 [51]	*zh* [tʂ]	*ong* [ʊŋ]	T1 [55]

(e) Theme V: Snacks

Set 1: Trained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
饼干 (cookies)	*b* [p]	*ing* [iŋ]	**T3 [21]**	*g* [k]	*an* [an]	T1 [55]
软糖 (fudge)	*r* [ʐ]	*uan* [uan]	**T3 [21]**	*t* [t^{h}]	*ang* [aŋ]	T2 [35]
面包 (bread)	*m* [m]	*ian* [ian]	T4 [51]	*b* [p]	*ao* [ɑu]	T1 [55]
牛奶 (milk)	*n* [n]	*iu* [iou]	T2 [35]	*n* [n]	*ai* [ai]	T3 [214]
海苔 (seaweed)	*h* [x]	*ai* [ai]	**T3 [21]**	*t* [t^{h}]	*ai* [ai]	T2 [35]
橙汁 (orange juice)	*ch* [tʂ^{h}]	*eng* [əŋ]	T2 [35]	*zh* [tʂ]	*-i* [ʅ]	T1 [55]
瓜子(melon seeds)	*g* [k]	*ua* [uᴀ]	T1 [55]	*z* [ts]	*-i* [ɿ]	T3 [214]
薯片 (potato chips)	*sh* [ʂ]	*u* [u]	**T3 [21]**	*p* [p^{h}]	*ian* [ian]	T4 [51]
麻花 (bread twist)	*m* [m]	*a* [ᴀ]	T2 [35]	*h* [x]	*ua* [uᴀ]	T1 [55]
核桃 (walnut)	*h* [x]	*e* [ɤ]	T2 [35]	*t* [t^{h}]	*ao* [ɑu]	[neutral tone]
Set 2: Untrained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
蛋糕 (cake)	*d* [t]	*an* [an]	T4 [51]	*g* [k]	*ao* [ɑu]	T1 [55]
香肠 (sausage)	*x* [ɕ]	*iang* [iaŋ]	T1 [55]	*ch* [tʂ^{h}]	*ang* [aŋ]	T2 [35]

(f) Theme VI: Toys

Set 1: Trained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
积木(toy blocks)	*j* [tɕ]	*i* [i]	T1 [55]	*m* [m]	*u* [u]	T4 [51]
气球 (balloon)	*q* [tɕ^{h}]	*i* [i]	T4 [51]	*q* [tɕ^{h}]	*iu* [iou]	T2 [35]
拼图 (puzzle)	*p* [p^{h}]	*in* [in]	T1 [55]	*t* [t^{h}]	*u* [u]	T2 [35]
彩泥 (color-mud)	*c* [ts^{h}]	*ai* [ai]	**T3 [21]**	*n* [n]	*i* [i]	T2 [35]
滑梯 (slide)	*h* [x]	*ua* [uᴀ]	T2 [35]	*t* [t^{h}]	*i* [i]	T1 [55]
秋千 (swing)	*q* [tɕ^{h}]	*iu* [iou]	T1 [55]	*q* [tɕ^{h}]	*ian* [ian]	T1 [55]
魔方(magic cube)	*m* [m]	*o* [o]	T2 [35]	*f* [f]	*ang* [aŋ]	T1 [55]
风车 (windmill)	*f* [f]	*eng* [əŋ]	T1 [55]	*ch* [tʂ^{h}]	*e* [ɤ]	T1 [55]
铃铛 (bell)	*l* [l]	*ing* [iŋ]	T2 [35]	*d* [t]	*ang* [aŋ]	[neutral tone]
水枪 (water gun)	*sh* [ʂ]	*ui* [uei]	**T3 [21]**	*q* [tɕ^{h}]	*iang* [iaŋ]	T1 [55]
Set 2: Untrained Words	First Syllable			Second Syllable		
	Initial	Final	Tone	Initial	Final	Tone
足球 (football)	*z* [ts]	*u* [u]	T2 [35]	*q* [tɕ^{h}]	*iu* [iou]	T2 [35]
口哨 (whistle)	*k* [k^{h}]	*ou* [ou]	**T3 [21]**	*sh* [ʂ]	*ao* [ɑu]	T4 [51]

Note. *Italics* represent *Pinyin*, an alphabetic phonological coding system widely used in Mainland China, and the corresponding transcriptions of International Phonetic Alphabet (IPA) in square brackets. The symbol of *‘Ø’* means zero onset in Mandarin initials. There are four Mandarin citation tones, traditionally characterized as Tone 1 (T1), Tone 2 (T2), Tone 3 (T3), and Tone 4 (T4). T1 [55] is a high-level tone, T2 [35] is a mid-rising tone, T3 [214] is a dipping (i.e., falling-rising) tone, and T4 [51] is a high-falling tone, with digits in square brackets referring to tone transcriptions in Chao’s tone letters (Chao, 1930). The bold tone numbers indicate the surface tone realizations of T3 after the process of tone sandhi.

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic Intonation Therapy for Aphasia. *Archives of Neurology*, *29*(2), 130–131. https://doi.org/10.1001/archneur.1973.00490260074018

Alcántara, José I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

Alcántara, José Ignacio, Cope, T. E., Cope, W., & Weisblatt, E. J. (2012). Auditory temporal-envelope processing in high-functioning children with Autism Spectrum Disorder. *Neuropsychologia*, *50*(7), 1235–1251. https://doi.org/10.1016/j.neuropsychologia.2012.01.034

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington, VA : American Psychiatric Publishing.

Anderson, S., & Kraus, N. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *Journal of the American Academy of Audiology*, *21*(9), 575–585. https://doi.org/10.3766/jaaa.21.9.3

Applebaum, E., Egel, A. L., Koegel, R. L., & Imhoff, B. (1979). Measuring musical abilities of autistic children. *Journal of Autism and Developmental Disorders*, *9*(3), 279–285. https://doi.org/10.1007/BF01531742

Baldassi, S., Pei, F., Megna, N., Recupero, G., Viespoli, M., Igliozzi, R., Tancredi, R., Muratori, F., & Cioni, G. (2009). Search superiority in autism within, but not outside the crowding regime. *Vision Research*, *49*(16), 2151–2156. https://doi.org/10.1016/j.visres.2009.06.007

Baldwin, C. L. (2012). *Auditory Cognition and Human Performance: Research and Applications*. CRC Press.

Baltaxe, C. A. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research*, *27*(1), 97–105. https://doi.org/10.1044/jshr.2701.97

Baltaxe, C. A. M., & D'Angiola, N. (1992). Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders*, *22*(1), 1–21. https://doi.org/10.1007/BF01046399

Baltaxe, C. A. M., & Guthrie, D. (1987). The use of primary sentence stress by normal, aphasic, and autistic children. *Journal of Autism and Developmental Disorders*, *17*(2), 255–271. https://doi.org/10.1007/BF01495060

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H.-J., & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*(3), 917–926. https://doi.org/10.1016/j.neuroimage.2005.10.044

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, *30*(2), 285–297. https://doi.org/10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, Mass. : MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

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

Bartak, L., Rutter, M., & Cox, A. (1975). A comparative study of infantile autism and specific developmental receptive language disorder. *The British Journal of Psychiatry*, *126*(2), 127 – 145. https://doi.org/10.1192/bjp.126.2.127

Bartolucci, G., Pierce, S., Streiner, D., & Eppel, P. (1976). Phonological investigation of verbal autistic and mentally retarded subjects. *Journal of Autism and Childhood Schizophrenia*, *6*(4), 303–316. https://doi.org/10.1007/BF01537908

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain*, *128*(10), 2430–2441. https://doi.org/10.1093/brain/awh561

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(2–3), 183–216. https://doi.org/10.1177/00238309030460020701

Beukelman, D. R., & Mirenda, P. (1998). *Augmentative and alternative communication*. Baltimore: Paul H. Brookes.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0060676

Binder, J. (2000). The new neuroanatomy of speech perception. *Brain*, *123*(12), 2371–2372. https://doi.org/10.1093/brain/123.12.2371

Bishop, D. V. M., Maybery, M., Wong, D., Maley, A., Hill, W., & Hallmayer, J. (2004). Are phonological processing deficits part of the broad autism phenotype? *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *128B*(1), 54–60. https://doi.org/10.1002/ajmg.b.30039

Blakemore, S.-J., Tavassoli, T., Calò, S., Thomas, R. M., Catmur, C., Frith, U., & Haggard, P. (2006). Tactile sensitivity in Asperger syndrome. *Brain and Cognition*, *61*(1), 5–13. https://doi.org/10.1016/j.bandc.2005.12.013

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. https://doi.org/10.1038/nn1409

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (Version 6.0. 14) [Computer program]*. http://www.praat.org/

Boets, B., Verhoeven, J., Wouters, J., & Steyaert, J. (2015). Fragile spectral and temporal auditory processing in adolescents with autism spectrum disorder and early language delay. *Journal of Autism and Developmental Disorders*, *45*(6), 1845–1857. https://doi.org/10.1007/s10803-014-2341-1

Bolinger, D. L. (1972). *Intonation*. Harmondsworth: Penguin.

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, *4*, 237. https://doi.org/10.3389/fnhum.2010.00237

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., Burack, J. A., & Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, *48*(9), 2465–2475. https://doi.org/10.1016/j.neuropsychologia.2010.04.020

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Bosseler, A., & Massaro, D. W. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of Autism and Developmental Disorders*, *33*(6), 653–672. https://doi.org/10.1023/B:JADD.0000006002.82367.4f

Boucher, J. (1976). Articulation in early childhood autism. *Journal of Autism and Childhood Schizophrenia*, *6*(4), 297–302. https://doi.org/10.1007/BF01537907

Brodeur, D. A., Green, C. G., Flores, H., & Burack, J. A. (2014). Time estimation among low-functioning individuals with autism spectrum disorders: Evidence of poor sensitivity to variability of short durations. *Autism Research*, *7*(2), 237–244. https://doi.org/10.1002/aur.1364

Bruneau, N., Roux, S., Adrien, J. L., & Barthélémy, C. (1999). Auditory associative cortex dysfunction in children with autism: Evidence from late auditory evoked potentials (N1 wave–T complex). *Clinical Neurophysiology*, *110*(11), 1927–1934. https://doi.org/10.1016/S1388-2457(99)00149-2

Buday, E. M. (1995). The effects of signed and spoken words taught with music on sign and speech imitation by children with autism. *Journal of Music Therapy*, *32*(3), 189–202. https://doi.org/10.1093/jmt/32.3.189

Callahan, K., Shukla-Mehta, S., Magee, S., & Wie, M. (2010). ABA versus TEACCH: The case for defining and validating comprehensive treatment models in autism. *Journal of Autism and Developmental Disorders*, *40*(1), 74–88. https://doi.org/10.1007/s10803-009-0834-0

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: From the subcortical roots to cortical deviance detection. *Trends in Hearing*, *22*, 1–33. https://doi.org/10.1177/2331216518784822

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8–16. https://doi.org/10.1002/ana.20319

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, *29*(1), 57–66. https://doi.org/10.1023/A:1025970600181

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech–sound-selective auditory impairment in children with autism: They can perceive but do not attend. *Proceedings of the National Academy of Sciences*, *100*(9), 5567–5572. https://doi.org/10.1073/pnas.0835631100

Chan, K. K. L., & To, C. K. S. (2016). Do individuals with high-functioning autism who speak a tone language show intonation deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792. https://doi.org/10.1007/s10803-016-2709-5

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*(1), 1–9. https://doi.org/10.1016/j.bandl.2008.02.001

Chang, Y.-C., & Locke, J. (2016). A systematic review of peer-mediated interventions for children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *27*, 1–10. https://doi.org/10.1016/j.rasd.2016.03.010

Chao, Y. R. (1930). A system of tone letters. *Le Maiîre Phonétique*, *45*, 24–27.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, F., Zhang, H., Wang, S. Y., & Peng, G. (2019). Intrinsic cues and vowel categorical perception. *Linguistic Sciences*, *18*(4), 410−425. https://doi.org/10.7509/j.linsci.201808.032304

Chen, Fei, & Peng, G. (2016). Context effect in the categorical perception of Mandarin tones. *Journal of Signal Processing Systems*, *82*(2), 253–261. https://doi.org/10.1007/s11265-015-1008-2

Chen, Fei, Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, *44*(6), 1413–1434. https://doi.org/10.1017/S0305000916000581

Chen, Fei, Wang, L., Peng, G., Yan, N., & Pan, X. (2019). Development and evaluation of a 3-D virtual pronunciation tutor for children with autism spectrum disorders. *PLOS ONE*, *14*(1), e0210858. https://doi.org/10.1371/journal.pone.0210858

Chen, Fei, Yan, N., Pan, X., Yang, F., Ji, Z., Wang, L., & Peng, G. (2016). *Impaired categorical perception of Mandarin tones and its relationship to language ability in autism spectrum disorders*. 233–237. https://doi.org/10.21437/Interspeech.2016-1133

Chenausky, K., Brignell, A., Morgan, A., & Tager-Flusberg, H. (2019). Motor speech impairment predicts expressive language in minimally verbal, but not low verbal, individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, *4*, 2396941519856333. https://doi.org/10.1177/2396941519856333

Chenausky, K., Norton, A., Tager-Flusberg, H., & Schlaug, G. (2016). Auditory-motor mapping training: Comparing the effects of a novel speech treatment to a control treatment for minimally verbal children with autism. *PLOS ONE*, *11*(11), e0164930. https://doi.org/10.1371/journal.pone.0164930

Chenausky, K. V., Norton, A. C., & Schlaug, G. (2017). Auditory-motor mapping training in a more verbal child with autism. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00426

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch perception in tone language-speaking adults with and without autism spectrum disorders. *I-Perception*, *8*(3), 2041669517711200. https://doi.org/10.1177/2041669517711200

Cheung, H., Chung, K. K. H., Wong, S. W. L., McBride-Chang, C., Penney, T. B., & Ho, C. S. H. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry*, *50*(6), 726–733. https://doi.org/10.1111/j.1469-7610.2008.02001.x

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*, *49*(3), 507–517. https://doi.org/10.1016/j.neuropsychologia.2010.11.042

Church, B. A., Rice, C. L., Dovgopoly, A., Lopata, C. J., Thomeer, M. L., Nelson, A., & Mercado, E. (2015). Learning, plasticity, and atypical generalization in children with autism. *Psychonomic Bulletin & Review*, *22*(5), 1342–1348. https://doi.org/10.3758/s13423-014-0797-9

Cleland, J., Gibbon, F. E., Peppé, S., O'Hare, A., & Rutherford, M. (2010). Phonetic and phonological errors in children with high functioning autism and Asperger syndrome. *International Journal of Speech-Language Pathology*, *12*(1), 69–76. https://doi.org/10.3109/17549500903469980

Coffey-Corina, S., Padden, D., & Kuhl, P. K. (2008). ERPs to words correlate with behavioral measures in children with Autism Spectrum Disorder. *The Journal of the Acoustical Society of America*, *123*(5), 3742–3742. https://doi.org/10.1121/1.2935280

Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*, *45*(4), 719–726. https://doi.org/10.1111/j.1469-7610.2004.00266.x

Constantino, J. N., Yang, D., Gray, T. L., Gross, M. M., Abbacchi, A. M., Smith, S. C., Kohn, C. E., & Kuhl, P. K. (2007). Clarifying the associations between language and social development in autism: A study of non-native phoneme recognition. *Journal of Autism and Developmental Disorders*, *37*(7), 1256–1263. https://doi.org/10.1007/s10803-006-0269-9

Consul, P. C., & Famoye, F. (1992). Generalized poisson regression model. *Communications in Statistics - Theory and Methods*, *21*(1), 89–109. https://doi.org/10.1080/03610929208830766

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. JSTOR. https://doi.org/10.2307/1130766

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Curtin, S., & Vouloumanos, A. (2013). Speech preference is associated with autistic-like behavior in 18-months-olds at risk for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*(9), 2114–2120. https://doi.org/10.1007/s10803-013-1759-1

Cutler, A., & Isard, S. D. (1980). The production of prosody. In *Language production* (Vol. 1, pp. 245–269). London: Academic Press.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, *28*(6), 479–485. https://doi.org/10.1023/A:1026043926488

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–424. https://doi.org/10.1207/s15326942dn2703_6

DePape, A.-M. R., Hall, G. B. C., Tillmann, B., & Trainor, L. J. (2012). Auditory processing in high-functioning adolescents with autism spectrum disorder. *PLoS ONE*, *7*(9), e44084. https://doi.org/10.1371/journal.pone.0044084

Diehl, Joshua J., Watson, D., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(3), 385–404. https://doi.org/10.1017/S0142716409090201

Diehl, Joshua John, & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics*, *34*(1), 135–161. https://doi.org/10.1017/S0142716411000646

Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R., & Behrmann, M. (2012). Unreliable evoked responses in autism. *Neuron*, *75*(6), 981–991. https://doi.org/10.1016/j.neuron.2012.07.026

Droucker Danielle, Curtin Suzanne, & Vouloumanos Athena. (2013). Linking infant-directed speech and face preferences to language outcomes in infants at risk for autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *56*(2), 567–576. https://doi.org/10.1044/1092-4388(2012/11-0266)

Edelson, L., Grossman, R., & Tager-Flusberg, H. (2007). Emotional prosody in children and adolescents with autism. *Poster Session Presented at the Annual International Meeting for Autism Research*. Poster session presented at the annual international meeting for Autism Research, Seattle, WA.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics*, *22*(12), 937–956. https://doi.org/10.1080/02699200802330223

Eguchi, S. (1969). Development of speech sounds in children. *Acta Otolaryngol*, *257*, 1–51.

Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary test of nonverbal intelligence*. Austin, TX: Pro-Ed.

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691. https://doi.org/10.1016/j.rasd.2010.09.001

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M. T., & Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179. https://doi.org/10.1002/aur.239

Eyler, L. T., Pierce, K., & Courchesne, E. (2012). A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, *135*(3), 949–960. https://doi.org/10.1093/brain/awr364

Falter, C. M., Noreika, V., Wearden, J. H., & Bailey, A. J. (2012). More consistent, yet less sensitive: Interval timing in autism spectrum disorders. *Quarterly Journal of Experimental Psychology*, *65*(11), 2093–2107. https://doi.org/10.1080/17470218.2012.690770

Fay, W., & Schuler, A. L. (1980). *Emerging language in autistic children*. Baltimore, MD: University Park Press.

Feng, Y. (2018). *Development of categorical perception in Mandarin-speaking children* [Master Dissertation]. Central China Normal University.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209–221.

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680. https://doi.org/10.1016/S1388-2457(03)00153-6

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger Syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders*, *44*(8), 1972–1981. https://doi.org/10.1007/s10803-014-2073-2

Filipe, M. G., Watson, L., Vicente, S. G., & Frota, S. (2018). Atypical preference for infant-directed speech as an early marker of autism spectrum disorders? A literature review and directions for further research. *Clinical Linguistics and Phonetics*, *32*(3), 213–231. https://doi.org/10.1080/02699206.2017.1342694

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge: Cambridge University Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353–382). Berlin: Walter de Gruyter.

Flippin, M., Reszka, S., & Watson, L. R. (2010). Effectiveness of the Picture Exchange Communication System (PECS) on communication and speech for children with autism spectrum disorders: A meta-analysis. *American Journal of Speech-Language Pathology*, *19*(2), 178–195. https://doi.org/10.1044/1058-0360(2010/09-0022)

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Fosnot, S. M., & Jun, S. (1999). *Prosodic characteristics in children with stuttering or autism during reading and imitation*. Paper presented at the 14th international congress of phonetic sciences.

Foxton, J. M., Stewart, M. E., Barnard, L., Rodgers, J., Young, A. H., O'Brien, G., & Griffiths, T. D. (2003). Absence of auditory 'global interference' in autism. *Brain*, *126*(12), 2703–2709. https://doi.org/10.1093/brain/awg274

Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America*, *119*(3), 1712–1726. https://doi.org/10.1121/1.2149768

Frankel, F., Simmons, J. Q., & Richey, V. E. (1987). Reward value of prosodic features of language for autistic, mentally retarded, and normal children. *Journal of Autism and Developmental Disorders*, *17*(1), 103–113. https://doi.org/10.1007/BF01487263

Frith, U. (1989). *Autism: Explaining the Enigma*. Blackwell, Oxford.

Frith, U. (2003). *Autism: Explaining the Enigma* (2nd ed.). Oxford: Blackwell.

Fromkin, V. (1978). *Tone: A linguistic survey*. New York: Academic Press.

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal of the Acoustical Society of America*, *146*(5), EL424–EL430. https://doi.org/10.1121/1.5133661

Gage, N. M., Siegel, B., Callen, M., & Roberts, T. P. L. (2003). Cortical sound processing in children with autism disorder: An MEG investigation. *NeuroReport*, *14*(16), 2047–2051.

Gandour, Jack. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*(1), 20–36. JSTOR.

Gandour, Jack. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175. https://doi.org/10.1016/S0095-4470(19)30813-7

Gandour, Jackson, Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage*, *23*(1), 344–357. https://doi.org/10.1016/j.neuroimage.2004.06.004

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children’s repetition of nonwords. *Applied Psycholinguistics*, *12*(3), 349–367. https://doi.org/10.1017/S0142716400009267

Gebauer, L., Skewes, J., Hørlyck, L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical, 6*, 370–378. https://doi.org/10.1016/j.nicl.2014.08.025

Gillberg, C. (1991). Outcome in autism and autistic-like conditions. *Journal of the American Academy of Child & Adolescent Psychiatry*, *30*(3), 375–382. https://doi.org/10.1097/00004583-199105000-00004

Gillespie-Lynch, K., Sepeta, L., Wang, Y., Marshall, S., Gomez, L., Sigman, M., & Hutman, T. (2012). Early childhood predictors of the social competence of adults with autism. *Journal of Autism and Developmental Disorders*, *42*(2), 161–174. https://doi.org/10.1007/s10803-011-1222-0

Gilliam, J. E. (2006). *Gilliam Autism Rating Scale: GARS 2*. Austin, TX: PRO-ED.

Goffman, L. (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits: Kinematic, acoustic, and transcription evidence. *Journal of Speech, Language, and Hearing Research: JSLHR*, *42*(6), 1499–1517. https://doi.org/10.1044/jslhr.4206.1499

Gogolla, N., LeBlanc, J. J., Quast, K. B., Südhof, T. C., Fagiolini, M., & Hensch, T. K. (2009). Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *Journal of Neurodevelopmental Disorders*, *1*(2), 172–181. https://doi.org/10.1007/s11689-009-9023-x

Gogolla, N., Takesian, A. E., Feng, G., Fagiolini, M., & Hensch, T. K. (2014). Sensory integration in mouse insular cortex reflects GABA circuit maturation. *Neuron*, *83*(4), 894–905. https://doi.org/10.1016/j.neuron.2014.06.033

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*(6), 1096–1106. https://doi.org/10.1007/s10803-006-0252-5

Gold, C., Wigram, T., & Elefant, C. (2006). Music therapy for autistic spectrum disorder. *Cochrane Database of Systematic Reviews*, *2*. https://doi.org/10.1002/14651858.CD004381.pub2

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gomot, M., Blanc, R., Clery, H., Roux, S., Barthelemy, C., & Bruneau, N. (2011). Candidate electrophysiological endophenotypes of hyper-reactivity to change in autism. *Journal of Autism and Developmental Disorders*, *41*(6), 705–714. https://doi.org/10.1007/s10803-010-1091-y

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: Electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(5), 577–584. https://doi.org/10.1017/S0048577202394058

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult … but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315. https://doi.org/10.1080/17549500903003060

Groen, W. B., van Orsouw, L., Huurne, N. ter, Swinkels, S., van der Gaag, R.-J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750. https://doi.org/10.1007/s10803-008-0682-3

Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *NeuroImage*, *83*, 637–645. https://doi.org/10.1016/j.neuroimage.2013.02.080

Güçlü, B., Tanidir, C., Mukaddes, N. M., & Unal, F. (2007). Tactile sensitivity of normal and autistic children. *Somatosensory & Motor Research*, *24*(1–2), 21–33. https://doi.org/10.1080/08990220601179418

Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(2), 701–714. https://doi.org/10.1016/j.rasd.2010.11.006

Hairston, M. (1990). Analyses of responses of mentally retarded autistic and mentally retarded nonautistic children to art therapy and music therapy. *Journal of Music Therapy*, *27*(3), 137–150. https://doi.org/10.1093/jmt/27.3.137

Hamann, C. (2017). Phonological working memory and language development: What are the measures and what do they measure? *Applied Psycholinguistics*, *38*(6), 1313–1321. https://doi.org/10.1017/S0142716417000261

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Happé, F., & Frith, U. (2020). Annual Research Review: Looking back to look forward – changes in the concept of autism and implications for future research. *Journal of Child Psychology and Psychiatry*, *n/a*(n/a). https://doi.org/10.1111/jcpp.13176

Heaton, P., Pring, L., & Hermelin, B. (2001). Musical processing in high functioning children with autism. *Annals of the New York Academy of Sciences*, *930*(1), 443–444. https://doi.org/10.1111/j.1749-6632.2001.tb05765.x

Heaton, Pamela. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 543–551. https://doi.org/10.1111/1469-7610.00143

Heaton, Pamela. (2005). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 787–793. https://doi.org/10.1007/s10803-005-0024-7

Heaton, Pamela, Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception: An Interdisciplinary Journal*, *15*(3), 291–305. https://doi.org/10.2307/40285769

Heaton, Pamela, Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology*, *25*(6), 771–782. https://doi.org/10.1080/02643290802336277

Heaton, Pamela, Williams, K., Cummins, O., & Happé, F. (2008). Autism and pitch processing splinter skills: A group and subgroup analysis. *Autism*, *12*(2), 203–219. https://doi.org/10.1177/1362361307085270

Heikkinen, J., Jansson-Verkasalo, E., Toivanen, J., Suominen, K., Väyrynen, E., Moilanen, I., & Seppänen, T. (2010). Perception of basic emotions from speech prosody in adolescents with Asperger’s syndrome. *Logopedics Phoniatrics Vocology*, *35*(3), 113–120. https://doi.org/10.3109/14015430903311184

Henderson, L. M., Clarke, P. J., & Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *52*(9), 964–973. https://doi.org/10.1111/j.1469-7610.2011.02393.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

Hesling, I., Dilharreguy, B., Peppé, S., Amirault, M., Bouvard, M., & Allard, M. (2010). The integration of prosodic speech in high functioning autism: A preliminary fMRI study. *PLOS ONE*, *5*(7), e11571. https://doi.org/10.1371/journal.pone.0011571

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hitoglou, M., Ververi, A., Antoniadis, A., & Zafeiriou, D. I. (2010). Childhood autism and auditory system abnormalities. *Pediatric Neurology*, *42*(5), 309–314. https://doi.org/10.1016/j.pediatrneurol.2009.10.009

Hoelzley, P. D. (1993). Communication potentiating sounds: Developing channels of communication with autistic children through psychobiological responses to novel sound stimuli. *Canadian Journal of Music Therapy*, *1*, 54–76.

Hoonhorst, I., Medina, V., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2011). Categorical perception of voicing, colors and facial expressions: A developmental study. *Speech Communication*, *53*(3), 417–430. https://doi.org/10.1016/j.specom.2010.11.005

Houde, J., Nagarajan, S., & Heinks-Maldonado, T. (2007). Dynamic cortical imaging of speech compensation for auditory feedback perturbations. *The Journal of the Acoustical Society of America*, *121*(5), 3045–3045. https://doi.org/10.1121/1.4781744

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

Huang, D., Yu, L., Wang, X., Fan, Y., Wang, S., & Zhang, Y. (2018). Distinct patterns of discrimination and orienting for temporal processing of speech and nonspeech in Chinese children with autism: An event-related potential study. *European Journal of Neuroscience, 47*(6), 662–668. https://doi.org/10.1111/ejn.13657

Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *The Journal of the Acoustical Society of America*, *125*(6), 3983–3994. https://doi.org/10.1121/1.3125342

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research*, *36*(2), 159–173. https://doi.org/10.1007/s10936-006-9037-4

Hutchins, S., & Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain and Language*, *123*(3), 234–239. https://doi.org/10.1016/j.bandl.2012.09.011

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–951.

Ingersoll, B. (2008). The social role of imitation in autism: Implications for the treatment of imitation deficits. *Infants & Young Children*, *21*(2), 107–119. https://doi.org/10.1097/01.IYC.0000314482.24087.14

James, R., Sigafoos, J., Green, V. A., Lancioni, G. E., O'Reilly, M. F., Lang, R., Davis, T., Carnett, A., Achmadi, D., Gevarter, C., & Marschik, P. B. (2015). Music therapy for individuals with autism spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders*, *2*(1), 39–54. https://doi.org/10.1007/s40489-014-0035-4

Jansson-Verkasalo, E., Ceponiene, R., Kielinen, M., Suominen, K., Jäntti, V., Linna, S.-L., Moilanen, I., & Näätänen, R. (2003). Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neuroscience Letters*, *338*(3), 197–200. https://doi.org/10.1016/S0304-3940(02)01405-2

Järvinen-Pasley, A., & Heaton, P. (2007). Evidence for reduced domain-specificity in auditory processing in autism. *Developmental Science*, *10*(6), 786–793. https://doi.org/10.1111/j.1467-7687.2007.00637.x

Järvinen-Pasley, A., Pasley, J., & Heaton, P. (2008). Is the linguistic content of speech less salient than its perceptual features in autism? *Journal of Autism and Developmental Disorders*, *38*(2), 239–248. https://doi.org/10.1007/s10803-007-0386-0

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The relationship between form and function level receptive prosodic abilities in autism. *Journal of Autism and Developmental Disorders*, *38*(7), 1328–1340. https://doi.org/10.1007/s10803-007-0520-z

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of melodic contour and intonation in autism spectrum disorder: Evidence from Mandarin speakers. *Journal of Autism and Developmental Disorders*, *45*(7), 2067–2075. https://doi.org/10.1007/s10803-015-2370-4

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry*, *65*(8), 946–954. https://doi.org/10.1001/archpsyc.65.8.946

Joseph, R. M., Keehn, B., Connolly, C., Wolfe, J. M., & Horowitz, T. S. (2009). Why is visual search superior in autism spectrum disorder? *Developmental Science*, *12*(6), 1083–1096. https://doi.org/10.1111/j.1467-7687.2009.00855.x

Junge, C., & Cutler, A. (2014). Early word recognition and later language skills. *Brain Sciences*, *4*(4), 532–559. https://doi.org/10.3390/brainsci4040532

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kanner, Leo. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia*, *1*(2), 119–145. https://doi.org/10.1007/BF01537953

Kasai, K., Hashimoto, O., Kawakubo, Y., Yumoto, M., Kamio, S., Itoh, K., Koshida, I., Iwanami, A., Nakagome, K., Fukuda, M., Yamasue, H., Yamada, H., Abe, O., Aoki, S., & Kato, N. (2005). Delayed automatic detection of change in speech sounds in adults with autism: A magnetoencephalographic study. *Clinical Neurophysiology*, *116*(7), 1655–1664. https://doi.org/10.1016/j.clinph.2005.03.007

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. *Proceedings of International Conference on Asia-Pacific Signal and Information Processing Association*, 111–121.

Keehn, B., Brenner, L., Palmer, E., Lincoln, A. J., & Müller, R.-A. (2008). Functional brain organization for visual search in ASD. *Journal of the International Neuropsychological Society*, *14*(6), 990–1003. https://doi.org/10.1017/S1355617708081356

Kéïta, L., Mottron, L., & Bertone, A. (2010). Far visual acuity is unremarkable in autism: Do we need to focus on crowding? *Autism Research*, *3*(6), 333–341. https://doi.org/10.1002/aur.164

King, M. L., Takeguchi, K., Barry, S. E., Rehfeldt, R. A., Boyer, V. E., & Mathews, T. L. (2014). Evaluation of the iPad in the acquisition of requesting skills for children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *8*(9), 1107–1120. https://doi.org/10.1016/j.rasd.2014.05.011

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*(2–3), 287–308. https://doi.org/10.1080/01690960042000058

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience & Biobehavioral Reviews*, *31*(7), 1003–1016. https://doi.org/10.1016/j.neubiorev.2007.03.005

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42. https://doi.org/10.1007/BF02206995

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, *459*, 257. https://doi.org/10.1038/nature07868

Klinger, L., Dawson, G., & Renner, P. (2002). Autistic disorder. In *Child Psychopathology* (2nd ed., pp. 409–454). New York: Guilford Press.

Koegel, R. L., Shirotova, L., & Koegel, L. K. (2009). Brief report: Using individualized orienting cues to facilitate first-word acquisition in non-responders with autism. *Journal of Autism and Developmental Disorders*, *39*(11), 1587–1592. https://doi.org/10.1007/s10803-009-0765-9

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, *17*(2), 956–966. https://doi.org/10.1006/nimg.2002.1154

Koerner, T. K., & Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sciences*, *7*(3), 26. https://doi.org/10.3390/brainsci7030026

Kubicek, C., Gervain, J., Hillairet de Boisferon, A., Pascalis, O., Lœvenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior and Development*, *37*(4), 644–651. https://doi.org/10.1016/j.infbeh.2014.08.010

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K. (2011). Who’s talking? *Science*, *333*(6042), 529–530. https://doi.org/10.1126/science.1210277

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science*, *8*(1), F1–F12. https://doi.org/10.1111/j.1467-7687.2004.00384.x

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

Kujala, T., Kuuluvainen, S., Saalasti, S., Jansson-Verkasalo, E., Wendt, L. von, & Lepistö, T. (2010). Speech-feature discrimination in children with Asperger syndrome as determined with the multi-feature mismatch negativity paradigm. *Clinical Neurophysiology*, *121*(9), 1410–1419. https://doi.org/10.1016/j.clinph.2010.03.017

Kujala, T., Lepistö, T., & Näätänen, R. (2013). The neural basis of aberrant speech and audition in autism spectrum disorders. *Neuroscience & Biobehavioral Reviews*, *37*(4), 697–704. https://doi.org/10.1016/j.neubiorev.2013.01.006

Kujala, Teija. (2007). The role of early auditory discrimination deficits in language disorders. *Journal of Psychophysiology*, *21*(3–4), 239–250. https://doi.org/10.1027/0269-8803.21.34.239

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

LaGasse, A. B. (2017). Social outcomes in children with autism spectrum disorder: A review of music therapy outcomes. *Patient Related Outcome Measures*, *8*, 23–32. https://doi.org/10.2147/PROM.S106267

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*(2), 308–314. https://doi.org/10.1523/JNEUROSCI.4822-06.2007

Lai, M.-C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *The Lancet*, *383*(9920), 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1

Lancker, D. van. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Paper in Linguistics*, *13*(2), 201–277. https://doi.org/10.1080/08351818009370498

Landa, R. (2007). Early communication development and intervention for children with autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*(1), 16–25. https://doi.org/10.1002/mrdd.20134

Landa, R. J. (2008). Diagnosis of autism spectrum disorders in the first 3 years of life. *Nature Clinical Practice Neurology*, *4*(3), 138–147. https://doi.org/10.1038/ncpneuro0731

Leclercq, D., Duffau, H., Delmaire, C., Capelle, L., Gatignol, P., Ducros, M., Chiras, J., & Lehéricy, S. (2010). Comparison of diffusion tensor imaging tractography of language tracts and intraoperative subcortical stimulations: Clinical article. *Journal of Neurosurgery*, *112*(3), 503–511. https://doi.org/10.3171/2009.8.JNS09558

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235–3248. https://doi.org/10.1121/1.2990713

Legendre, P. (2005). Species associations: The Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, *10*(2), 226. https://doi.org/10.1198/108571105X46642

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

Lepistö, T., Kajander, M., Vanhala, R., Alku, P., Huotilainen, M., Näätänen, R., & Kujala, T. (2008). The perception of invariant speech features in children with autism. *Biological Psychology*, *77*(1), 25–31. https://doi.org/10.1016/j.biopsycho.2007.08.010

Lepistö, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Näätänen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neurophysiology*, *117*(10), 2161–2171. https://doi.org/10.1016/j.clinph.2006.06.709

Lepistö, Tuulia, Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052

Lewis, M. M. (1957). *How Children Learn to Speak*. London: Harrap.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lindell, A. K., & Hudry, K. (2013). Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. *Neuropsychology Review*, *23*(3), 257–270. https://doi.org/10.1007/s11065-013-9234-5

Linguistic Society of Hong Kong [LSHK]. (2002). *Hong Kong JyutPing Character Table* (2nd ed.). Hong Kong: Linguistic Society of Hong Kong.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, *62*(2–4), 70–87. https://doi.org/10.1159/000090090

Liu, Y., Li, J., Zheng, Q., Zaroff, C. M., Hall, B. J., Li, X., & Hao, Y. (2016). Knowledge, attitudes, and perceptions of autism spectrum disorder in a stratified sampling of preschool teachers in China. *BMC Psychiatry*, *16*(1), 142. https://doi.org/10.1186/s12888-016-0845-2

López-Barroso, D., Catani, M., Ripollés, P., Dell’Acqua, F., Rodríguez-Fornells, A., & Diego-Balaguer, R. de. (2013). Word learning is mediated by the left arcuate fasciculus.

Proceedings of the National Academy of Sciences, *110*(32), 13168–13173. https://doi.org/10.1073/pnas.1301696110

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1), 1–23. https://doi.org/10.1038/s41572-019-0138-4

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *The Lancet*, *392*(10146), 508–520. https://doi.org/10.1016/S0140-6736(18)31129-2

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry*, *63*(6), 694–701. https://doi.org/10.1001/archpsyc.63.6.694

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule: ADOS–2*. Los Angeles, CA: Western Psychological Services.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *120*(6), 3978–3987. https://doi.org/10.1121/1.2363927

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User’s Guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Maister, L., & Plaisted-Grant, K. C. (2011). Time perception and its relationship to memory in Autism Spectrum Conditions. *Developmental Science*, *14*(6), 1311–1322. https://doi.org/10.1111/j.1467-7687.2011.01077.x

Major, R. C., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning*, *46*(3), 465–496. https://doi.org/10.1111/j.1467-1770.1996.tb01244.x

Mamashli, F., Khan, S., Bharadwaj, H., Michmizos, K., Ganesan, S., Garel, K.-L. A., Hashmi, J. A., Herbert, M. R., Hämäläinen, M., & Kenet, T. (2017). Auditory processing in noise is associated with complex patterns of disrupted functional connectivity in autism spectrum disorder. *Autism Research*, *10*(4), 631–647. https://doi.org/10.1002/aur.1714

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language & Communication Disorders*, *44*(4), 466–488. https://doi.org/10.1080/13682820802591643

Martin, J. S., Poirier, M., & Bowler, D. M. (2010). Brief report: Impaired temporal reproduction performance in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*(5), 640–646. https://doi.org/10.1007/s10803-009-0904-3

Martzoukou, M., Papadopoulou, D., & Kosmidis, M.-H. (2017). The comprehension of syntactic and affective prosody by adults with autism spectrum disorder without accompanying cognitive deficits. *Journal of Psycholinguistic Research*, *46*(6), 1573–1595. https://doi.org/10.1007/s10936-017-9500-4

Mason, R. A., Williams, D. L., Kana, R. K., Minshew, N., & Just, M. A. (2008). Theory of Mind disruption and recruitment of the right hemisphere during narrative comprehension in autism. *Neuropsychologia*, *46*(1), 269–280. https://doi.org/10.1016/j.neuropsychologia.2007.07.018

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Mazefsky, C. A., & Oswald, D. P. (2007). Emotion perception in Asperger's Syndrome and high-functioning autism: The importance of diagnostic criteria and cue intensity. *Journal of Autism and Developmental Disorders*, *37*(6), 1086–1095. https://doi.org/10.1007/s10803-006-0251-6

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

McCleery, J. P., Tully, L., Slevc, L. R., & Schreibman, L. (2006). Consonant production patterns of young severely language-delayed children with autism. *Journal of Communication Disorders*, *39*(3), 217–231. https://doi.org/10.1016/j.jcomdis.2005.12.002

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McKavanagh, R., Buckley, E., & Chance, S. A. (2015). Wider minicolumns in autism: A neural basis for altered processing? *Brain*, *138*(7), 2034–2045. https://doi.org/10.1093/brain/awv110

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics*, *38*(4), 493–503. https://doi.org/10.1016/j.wocn.2010.06.002

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., & Töpper, R. (2003). Motor cortex hand area and speech: Implications for the development of language. *Neuropsychologia*, *41*(4), 401–406. https://doi.org/10.1016/S0028-3932(02)00179-3

Mesa-Gresa, P., Gil-Gómez, H., Lozano-Quilis, J.-A., & Gil-Gómez, J.-A. (2018). Effectiveness of virtual reality for children and adolescents with autism spectrum disorder: An evidence-based systematic review. *Sensors*, *18*(8), 2486. https://doi.org/10.3390/s18082486

Messum, P. R. (2008). *The role of imitation in learning to pronounce* [Doctoral thesis]. University of London.

Miller, L. (1989). *Musical savants: Exceptional skill in the mentally retarded*. Hillsdale, NJ: Erlbaum.

Miller, S. B., & Toca, J. M. (1979). Adapted melodic intonation therapy: A case study of an experimental language program for an autistic child. *The Journal of Clinical Psychiatry*, *40*(4), 201–203.

Millman, R. E., & Mattys, S. L. (2017). Auditory verbal working memory as a predictor of speech perception in modulated maskers in listeners with normal hearing. *Journal of Speech, Language, and Hearing Research*, *60*(5), 1236–1245. https://doi.org/10.1044/2017_JSLHR-S-16-0105

Minshew, N. J., & Goldstein, G. (1998). Autism as a disorder of complex information processing. *Mental Retardation and Developmental Disabilities Research Reviews*, *4*(2), 129–136. https://doi.org/10.1002/(SICI)1098-2779(1998)4:2<129::AID-MRDD10>3.0.CO;2-X

Mirman, D. (2014). *Growth curve analysis and visualization using R*. FL: Chapman & Hall/CRC.

Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*, *79*(1), 344–351. https://doi.org/10.3758/s13414-016-1195-3

Mok, P. P. K., Fung, H. S. H., & Li, V. G. (2019). Assessing the link between perception and production in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1243–1257. https://doi.org/10.1044/2018_JSLHR-S-17-0430

Mok, P. P. K., Li, V. G., & Fung, H. S. H. (2020). Development of phonetic contrasts in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, *63*(1), 95–108. https://doi.org/10.1044/2019_JSLHR-19-00152

Mok, P., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370. https://doi.org/10.1017/S0954394513000161

Mok, Peggy, & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Speech Prosody 2008*, 423–426.

Mottron, K., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The Development of Autism* (pp. 131–148). Mahwah, NJ: Lawrence Erlbaum Associates.

Mottron, L., Peretz, I., & Ménard, E. (2000). Local and global processing of music in high-functioning persons with autism: Beyond central coherence? *Journal of Child Psychology and Psychiatry*, *41*(8), 1057–1065. https://doi.org/10.1111/1469-7610.00693

Mottron, Laurent, Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

Nakai, Y., Takashima, R., Takiguchi, T., & Takada, S. (2014). Speech intonation in children with autism spectrum disorder. *Brain and Development*, *36*(6), 516–522. https://doi.org/10.1016/j.braindev.2013.07.006

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D. E., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*(28), E6630–E6639. https://doi.org/10.1073/pnas.1808412115

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*(4), 643–655. https://doi.org/10.1037/0012-1649.42.4.643

Ning, C. Y. (2013). *Test of Language Ability in Mandarin-speaking Preschoolers*. Institute of Linguistics, Tianjin Normal University: Tianjin University Press.

Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., Gillberg, C., & Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, *19*(8), 934–943. https://doi.org/10.1177/1362361314556782

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic Intonation Therapy. *Annals of the New York Academy of Sciences*, *1169*(1), 431–436. https://doi.org/10.1111/j.1749-6632.2009.04859.x

O'connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

O'Haire, M. E. (2013). Animal-assisted intervention for autism spectrum disorder: A systematic literature review. *Journal of Autism and Developmental Disorders*, *43*(7), 1606–1622. https://doi.org/10.1007/s10803-012-1707-5

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., & Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486–2495. https://doi.org/10.1093/cercor/bhp318

Orefice, L. L. (2019). Outside-in: Rethinking the etiology of autism spectrum disorders. *Science*, *366*(6461), 45–46. https://doi.org/10.1126/science.aaz3880

Orefice, L. L., Zimmerman, A. L., Chirila, A. M., Sleboda, S. J., Head, J. P., & Ginty, D. D. (2016). Peripheral mechanosensory neuron dysfunction underlies tactile and behavioral deficits in mouse models of ASDs. *Cell*, *166*(2), 299–313. https://doi.org/10.1016/j.cell.2016.05.033

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 719–730. https://doi.org/10.1037/0096-1523.27.3.719

O'Riordan, M., & Passetti, F. (2006). Discrimination in autism within different sensory modalities. *Journal of Autism and Developmental Disorders*, *36*(5), 665–675. https://doi.org/10.1007/s10803-006-0106-1

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, *14*(2), 239–251. https://doi.org/10.1017/s0954579402002031

Ouimet, T., Foster, N. E. V., Tryfon, A., & Hyde, K. L. (2012). Auditory-musical processing in autism spectrum disorders: A review of behavioral and brain imaging studies. *Annals of the New York Academy of Sciences*, *1252*(1), 325–331. https://doi.org/10.1111/j.1749-6632.2012.06453.x

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*(5), 489–504. https://doi.org/10.1525/mp.2009.26.5.489

Paccia, J. M., & Curcio, F. (1982). Language processing and forms of immediate echolalia in autistic children. *Journal of Speech, Language, and Hearing Research*, *25*(1), 42–47. https://doi.org/10.1044/jshr.2501.42

Panagos, J. M., & Prelock, P. (1997). Prosodic analysis of child speech. *Topics in Language Disorders*, *17*(4), 1–10.

Paul, R. (2008). Interventions to improve communication in autism. *Child and Adolescent Psychiatric Clinics of North America*, *17*(4), 835–856. https://doi.org/10.1016/j.chc.2008.06.011

Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(2), 418–431. https://doi.org/10.1007/s10803-012-1583-z

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(6), 861–869. https://doi.org/10.1007/s10803-005-0031-8

Paul Rhea, Chawarska Katarzyna, Fowler Carol, Cicchetti Domenic, & Volkmar Fred. (2007). "Listen my children and you shall hear": Auditory preferences in toddlers with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1350–1364. https://doi.org/10.1044/1092-4388(2007/094)

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*(1), 134–154. JSTOR.

Peng, G., Deutsch, D., Henthorn, T., Su, D., & Wang, W. S.-Y. (2013). Language experience influences non-linguistic pitch perception. *Journal of Chinese Linguistics*, *41*(2), 447–467.

Peng, G., & Wang, W. S. Y. (2005). Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication*, *45*(1), 49–62. https://doi.org/10.1016/j.specom.2004.09.004

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140090. https://doi.org/10.1098/rstb.2014.0090

Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *5*(1), 60–69. https://doi.org/10.1016/j.rasd.2010.03.011

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*(6), 1385–1398. https://doi.org/10.3758/APP.71.6.1385

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced Visual Search for a Conjunctive Target in Autism: A Research Note. *Journal of Child Psychology and Psychiatry*, *39*(5), 777–783. https://doi.org/10.1111/1469-7610.00376

Posar, A., & Visconti, P. (2018). Sensory abnormalities in children with autism spectrum disorder. *Jornal de Pediatria*, *94*(4), 342–350. https://doi.org/10.1016/j.jped.2017.08.008

Prizant, B. M., & Duchan, J. F. (1981). The functions of immediate echolalia in autistic children. *Journal of Speech and Hearing Disorders*, *46*(3), 241–249. https://doi.org/10.1044/jshd.4603.241

Pronovost, W., Wakstein, M. P., & Wakstein, D. J. (1966). A longitudinal study of the speech behavior and language comprehension of fourteen children diagnosed atypical or autistic. *Exceptional Children*, *33*(1), 19–26. https://doi.org/10.1177/001440296603300104

Puts, N. A. J., Wodka, E. L., Harris, A. D., Crocetti, D., Tommerdahl, M., Mostofsky, S. H., & Edden, R. A. E. (2017). Reduced GABA and altered somatosensory function in children with autism spectrum disorder. *Autism Research*, *10*(4), 608–619. https://doi.org/10.1002/aur.1691

Puts, N. A. J., Wodka, E. L., Tommerdahl, M., Mostofsky, S. H., & Edden, R. A. E. (2014). Impaired tactile processing in children with autism spectrum disorder. *Journal of Neurophysiology*, *111*(9), 1803–1811. https://doi.org/10.1152/jn.00890.2013

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain and Development*, *25*(3), 166–172. https://doi.org/10.1016/S0387-7604(02)00191-2

Rapin, I., Dunn, M. A., Allen, D. A., Stevens, M. C., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology*, *34*(1), 66–84. https://doi.org/10.1080/87565640802564648

Rattanasone, N. X., 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*, 1–10.

Raven, J., & Court, J. (1998). *Raven manual: Section 3. Standard progressive matrices*. Oxford: Oxford Psychologists Press.

Rees, N. S. (1975). Imitation and language development: Issues and clinical implications. *Journal of Speech and Hearing Disorders*, *40*(3), 339–350. https://doi.org/10.1044/jshd.4003.339

Reschke-Hernández, A. E. (2011). History of music therapy treatment interventions for children with autism. *Journal of Music Therapy*, *48*(2), 169–207. https://doi.org/10.1093/jmt/48.2.169

Rimland, B., & Fein, D. (1988). Special talents of autistic savants. In L. Obler & D. Fein (Eds.), *The Exceptional Brain*. Guilford Press, New York.

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *55*(3), 683–694. https://doi.org/10.1044/1092-4388(2011/10-0263)

Roberts, T. P. L., Schmidt, G. L., Egeth, M., Blaskey, L., Rey, M. M., Edgar, J. C., & Levy, S. E. (2008). Electrophysiological signatures: Magnetoencephalographic studies of the neural correlates of language impairment in autism spectrum disorders. *International Journal of Psychophysiology*, *68*(2), 149–160. https://doi.org/10.1016/j.ijpsycho.2008.01.012

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Robertson, C. E., Martin, A., Baker, C. I., & Baron-Cohen, S. (2012). Atypical integration of motion signals in autism spectrum conditions. *PLOS ONE*, *7*(11), e48173. https://doi.org/10.1371/journal.pone.0048173

Robertson, C. E., Ratai, E.-M., & Kanwisher, N. (2016). Reduced GABAergic action in the autistic brain. *Current Biology*, *26*(1), 80–85. https://doi.org/10.1016/j.cub.2015.11.019

Robertson, C. E., Thomas, C., Kravitz, D. J., Wallace, G. L., Baron-Cohen, S., Martin, A., & Baker, C. I. (2014). Global motion perception deficits in autism are reflected as early as primary visual cortex. *Brain*, *137*(9), 2588–2599. https://doi.org/10.1093/brain/awu189

Rogers, S. J., Hayden, D., Hepburn, S., Charlifue-Smith, R., Hall, T., & Hayes, A. (2006). Teaching young nonverbal children with autism useful speech: A pilot study of the Denver Model and PROMPT interventions. *Journal of Autism and Developmental Disorders*, *36*(8), 1007–1024. https://doi.org/10.1007/s10803-006-0142-x

Rogers, S. J., Young, G. S., Cook, I., Giolzetti, A., & Ozonoff, S. (2008). Deferred and immediate imitation in regressive and early onset autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(4), 449–457. https://doi.org/10.1111/j.1469-7610.2007.01866.x

Romski, M., Sevcik, R. A., Adamson, L. B., Cheslock, M., Smith, A., Barker, R. M., & Bakeman, R. (2010). Randomized comparison of augmented and nonaugmented language interventions for toddlers with developmental delays and their parents. *Journal of Speech, Language, and Hearing Research*, *53*(2), 350–364. https://doi.org/10.1044/1092-4388(2009/08-0156)

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, *60*(5), 464–477. https://doi.org/10.1111/jir.12284

Russo, N., Zecker, S., Trommer, B., Chen, J., & Kraus, N. (2009). Effects of background noise on cortical encoding of speech in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(8), 1185–1196. https://doi.org/10.1007/s10803-009-0737-0

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194. https://doi.org/10.1023/A:1015497629971

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview—Revised*. Los Angeles: Western Psychological Services.

Rutter, Michael, Greenfeld, D., & Lockyer, L. (1967). A five to fifteen year follow-up study of infantile psychosis: Ii. Social and behavioural outcome. *The British Journal of Psychiatry*, *113*(504), 1183–1199. https://doi.org/10.1192/bjp.113.504.1183

Salomon-Gimmon, M., & Elefant, C. (2019). Development of vocal communication in children with autism spectrum disorder during improvisational music therapy. *Nordic Journal of Music Therapy*, *28*(3), 174–192. https://doi.org/10.1080/08098131.2018.1529698

Samson, F., Mottron, L., Jemel, B., Belin, P., & Ciocca, V. (2006). Can spectro-temporal complexity explain the autistic pattern of performance on auditory tasks? *Journal of Autism and Developmental Disorders*, *36*(1), 65–76. https://doi.org/10.1007/s10803-005-0043-4

Sandiford, G. A., Mainess, K. J., & Daher, N. S. (2013). A pilot study on the efficacy of melodic based communication therapy for eliciting speech in nonverbal children with autism. *Journal of Autism and Developmental Disorders*, *43*(6), 1298–1307. https://doi.org/10.1007/s10803-012-1672-z

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19–25. https://doi.org/10.1037/a0020740

Scheerer, N. E., Shafai, F., Stevenson, R. A., & Iarocci, G. (2020). Affective prosody perception and the relation to social competence in autistic and typically developing children. *Journal of Abnormal Child Psychology*, *48*(7), 965–975. https://doi.org/10.1007/s10802-020-00644-5

Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 315–323. https://doi.org/10.1525/mp.2008.25.4.315

Schlosser, R. W., & Wendt, O. (2008). Effects of augmentative and alternative communication intervention on speech production in children with autism: A systematic review. *American Journal of Speech-Language Pathology*, *17*(3), 212–230. https://doi.org/10.1044/1058-0360(2008/021)

Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research*, *4*(3), 177–188. https://doi.org/10.1002/aur.183

Schönwiesner, M., Rübsamen, R., & Cramon, D. Y. V. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*(6), 1521–1528. https://doi.org/10.1111/j.1460-9568.2005.04315.x

Schwartz, S., Shinn-Cunningham, B., & Tager-Flusberg, H. (2018). Meta-analysis and systematic review of the literature characterizing auditory mismatch negativity in individuals with autism. *Neuroscience & Biobehavioral Reviews*, *87*, 106–117. https://doi.org/10.1016/j.neubiorev.2018.01.008

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, *92*(1), 13–45. https://doi.org/10.1016/j.cognition.2002.12.002

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy*, *20*(3), 339–351. https://doi.org/10.1111/infa.12077

Shao, J., Wang, L., & Zhang, C. C. (2020). Talker processing in Mandarin-speaking congenital amusics. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1361–1375. https://doi.org/10.1044/2020_JSLHR-19-00209

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., Sen, A., Singhal, N., Erickson, D., & Singh, N. C. (2010). Sounds of melody—Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. https://doi.org/10.1016/j.neulet.2010.04.066

Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., Tan, M., Nadig, A., & Hyde, K. (2018). Music improves social communication and auditory–motor connectivity in children with autism. *Translational Psychiatry*, *8*(1), 1–13. https://doi.org/10.1038/s41398-018-0287-3

Sheinkopf, S. J., Mundy, P., Oller, D. K., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, *30*(4), 345–354. https://doi.org/10.1023/A:1005531501155

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115.

Shriberg, L. D., & Widder, C. J. (1990). Speech and prosody characteristics of adults with mental retardation. *Journal of Speech, Language, and Hearing Research*, *33*(4), 627–653. https://doi.org/10.1044/jshr.3304.627

Smith, T., & Eikeseth, S. (2011). O. Ivar Lovaas: Pioneer of applied behavior analysis and intervention for children with autism. *Journal of Autism and Developmental Disorders*, *41*(3), 375–378. https://doi.org/10.1007/s10803-010-1162-0

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. https://doi.org/10.1017/S0272263114000047

Soulières, I., Mottron, L., Saumier, D., & Larochelle, S. (2007). Atypical categorical perception in autism: Autonomy of discrimination? *Journal of Autism and Developmental Disorders*, *37*(3), 481–490. https://doi.org/10.1007/s10803-006-0172-4

Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, *10*(4), 303–316. https://doi.org/10.1016/S0010-9452(74)80024-9

Speidel, G. E., & Nelson, K. E. (1989). A Fresh Look at Imitation in Language Learning. In G. E. Speidel & K. E. Nelson (Eds.), *The Many Faces of Imitation in Language Learning* (pp. 1–21). Springer. https://doi.org/10.1007/978-1-4612-1011-5_1

Srinivasan, S. M., & Bhat, A. N. (2013). A review of "music and movement" therapies for children with autism: Embodied interventions for multisystem development. *Frontiers in Integrative Neuroscience*, *7*, 22. https://doi.org/10.3389/fnint.2013.00022

Stadnick, N. A., Stahmer, A., & Brookman-Frazee, L. (2015). Preliminary effectiveness of project ImPACT: A parent-mediated intervention for children with autism spectrum disorder delivered in a community program. *Journal of Autism and Developmental Disorders*, *45*(7), 2092–2104. https://doi.org/10.1007/s10803-015-2376-y

Stewart, M. E., & Ota, M. (2008). Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*, *109*(1), 157–162. https://doi.org/10.1016/j.cognition.2008.07.010

Stewart, M. E., Petrou, A. M., & Ota, M. (2018). Categorical speech perception in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *48*(1), 72–82. https://doi.org/10.1007/s10803-017-3284-0

Sulzer-Azaroff, B., Hoffman, A. O., Horton, C. B., Bondy, A., & Frost, L. (2009). The Picture Exchange Communication System (PECS): What do the data say? *Focus on Autism and Other Developmental Disabilities*, *24*(2), 89–103. https://doi.org/10.1177/1088357609332743

Sun, X., Allison, C., Wei, L., Matthews, F. E., Auyeung, B., Wu, Y. Y., Griffiths, S., Zhang, J., Baron-Cohen, S., & Brayne, C. (2019). Autism prevalence in China is comparable to Western prevalence. *Molecular Autism*, *10*(1), 7. https://doi.org/10.1186/s13229-018-0246-0

Swaminathan, J., Krishnan, A., Gandour, J. T., & Xu, Y. (2008). Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Transactions on Biomedical Engineering*, *55*(1), 281–287. https://doi.org/10.1109/TBME.2007.896592

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., Rees, L., & Wheelwright, S. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry*, *39*(5), 747–753. https://doi.org/10.1111/1469-7610.00373

Szelag, E., Kowalska, J., Galkowski, T., & Pöppel, E. (2004). Temporal processing deficits in high-functioning children with autism. *British Journal of Psychology*, *95*(3), 269–282. https://doi.org/10.1348/0007126041528167

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, *11*(1), 45–56. https://doi.org/10.1007/BF01531340

Tager-Flusberg, H., & Finch, K. H. (2020). Chapter 19—Early development of speech and language. In J. Rubenstein, P. Rakic, B. Chen, & K. Y. Kwan (Eds.), *Neural Circuit and Cognitive Development (Second Edition)* (pp. 413–434). Academic Press. https://doi.org/10.1016/B978-0-12-814411-4.00019-6

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1309–1325. https://doi.org/10.1044/2018_JSLHR-S-18-0304

Tang, W., Xiong, W., Zhang, Y., Dong, Q., & Nan, Y. (2016). Musical experience facilitates lexical tone processing among Mandarin speakers: Behavioral and neural evidence. *Neuropsychologia*, *91*, 247–253. https://doi.org/10.1016/j.neuropsychologia.2016.08.003

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Toth, K., Munson, J., N. Meltzoff, A., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders*, *36*(8), 993–1005. https://doi.org/10.1007/s10803-006-0137-7

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development*, *75*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Tsiouri, I., & Greer, R. D. (2003). Inducing vocal verbal behavior in children with severe language delays through rapid motor imitation responding. *Journal of Behavioral Education*, *12*(3), 185–206. https://doi.org/10.1023/A:1025508311022

T'sou, B., Lee, T., Tung, P., Chan, A., Man, Y., & To, C. K. S. (2006). *Hong Kong Cantonese Oral Language Assessment Scale*. Hong Kong: Language Information Sciences Research Centre.

Turner, L. M., Stone, W. L., Pozdol, S. L., & Coonrod, E. E. (2006). Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism*, *10*(3), 243–265. https://doi.org/10.1177/1362361306063296

Virues-Ortega, J., Julio, F. M., & Pastor-Barriuso, R. (2013). The TEACCH program for children and adults with autism: A meta-analysis of intervention studies. *Clinical Psychology Review*, *33*(8), 940–953. https://doi.org/10.1016/j.cpr.2013.07.005

Volkmar, F. R., Paul, R., Klin, A., & Cohen, D. J. (2005). *Handbook of autism and pervasive developmental disorders, diagnosis, development, neurobiology, and behavior*. John Wiley & Sons.

Walenski, M., Tager-Flusberg, H., & Ullman, M. T. (2006). Language in Autism. In S. O. Moldin & J. L. R. Rubenstein (Eds.), *Understanding autism: From basic neuroscience to treatment* (pp. 175–203). CRC Press/Routledge/Taylor & Francis Group. https://doi.org/10.1201/9781420004205.ch9

Wan, C Y, Demaine, K., Zipse, L., Norton, A., & Schlaug, G. (2010). From music making to speaking: Engaging the mirror neuron system in autism. *Brain Research Bulletin*, *82*(3), 161–168. https://doi.org/10.1016/j.brainresbull.2010.04.010

Wan, Catherine Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., Norton, A., & Schlaug, G. (2011). Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. *PLoS ONE*, *6*(9), e25505.

Wan, Catherine Y., Rüüber, T., Hohmann, A., & Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Perception: An Interdisciplinary Journal*, *27*(4), 287–295. https://doi.org/10.1525/mp.2010.27.4.287

Wang, W. S. Y. (1973). The Chinese language. *Scientific American*, *228*(2), 50–60.

Wang, W. S. Y. (1978). The three scales of diachrony. In B. B. Kachru (Ed.), *Linguistics in the Seventies: Directions and Prospects* (pp. 63–75). University of Illinois: Urbana, IL.

Wang, W. S.-Y. (1976). Language change. In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech* (Vol. 280, pp. 61–72). New York: New York Academy of Sciences.

Wang, X., & Peng, G. (2014). Phonological processing in Mandarin speakers with congenital amusia. *The Journal of the Acoustical Society of America*, *136*(6), 3360–3370. https://doi.org/10.1121/1.4900559

Wang, Xiaoyue, Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, *7*, 43254. https://doi.org/10.1038/srep43254

Watson, L. R., Roberts, J. E., Baranek, G. T., Mandulak, K. C., & Dalton, J. C. (2012). Behavioral and physiological responses to child-directed speech of children with autism spectrum disorders or typical development. *Journal of Autism and Developmental Disorders*, *42*(8), 1616–1629. https://doi.org/10.1007/s10803-011-1401-z

Watson Linda R., Baranek Grace T., Roberts Jane E., David Fabian J., & Perryman Twyla Y. (2010). Behavioral and Physiological Responses to Child-Directed Speech as Predictors of Communication Outcomes in Children With Autism Spectrum Disorders. *Journal of Speech, Language, and Hearing Research*, *53*(4), 1052–1064. https://doi.org/10.1044/1092-4388(2009/09-0096)

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *43*(2), 230–246. https://doi.org/10.1037/h0084224

Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders*, *30*(2), 157–162.

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *13*(1), 1–7. https://doi.org/10.1162/089892901564108

White, S., Frith, U., Milne, E., Rosen, S., Swettenham, J., & Ramus, F. (2006). A double dissociation between sensorimotor impairments and reading disability: A comparison of autistic and dyslexic children. *Cognitive Neuropsychology, 23*(5), 748–761. https://doi.org/10.1080/02643290500438607

Whitehouse, A. J. O., & Bishop, D. V. M. (2008). Do children with autism 'switch off' to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524. https://doi.org/10.1111/j.1467-7687.2008.00697.x

Whitehurst, G. J., & Vasta, R. (1975). Is language acquired through imitation? *Journal of Psycholinguistic Research*, *4*(1), 37–59. https://doi.org/10.1007/BF01066989

Wolk, L., & Brennan, C. (2013). Phonological investigation of speech sound errors in children with autism spectrum disorders. *Speech, Language and Hearing, 16*(4), 239–246. https://doi.org/10.1179/2050572813Y.0000000020

Wolk, L., Edwards, M. L., & Brennan, C. (2016). Phonological difficulties in children with autism: An overview. *Speech, Language and Hearing*, *19*(2), 121–129.

Wolk, L., & Giesen, J. (2000). A phonological investigation of four siblings with childhood autism. *Journal of Communication Disorders*, *33*(5), 371–389.

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434–443. https://doi.org/10.1121/1.4768883

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421. https://doi.org/10.1044/1092-4388(2003/034)

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420–422. https://doi.org/10.1038/nn1872

Wu, H., Lu, F., Yu, B., & Liu, Q. (2020). Phonological acquisition and development in Putonghua-speaking children with Autism Spectrum Disorders. *Clinical Linguistics & Phonetics,* 34(9), 844-860.

Xi, J., Jiang, W., Zhang, L., & Shu, H. (2009). Categorical Perception of VOT and Lexical Tones in Chinese and the Developmental Course. *Acta Psychologica Sinica*, *41*(7), 572–579.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223–231. https://doi.org/10.1016/j.neuroscience.2010.06.077

Xu, Yi. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, 7–10.

Xu, Yi, & Prom-on, S. (2019). Economy of effort or maximum rate of information? Exploring basic principles of articulatory dynamics. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02469

Xu, Yisheng, Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

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

You, R. S., Serniclaes, W., Rider, D., & Chabane, N. (2017). On the nature of the speech perception deficits in children with autism spectrum disorders. *Research in Developmental Disabilities*, *61*, 158–171.

Young, R. L., & Nettelbeck, T. (1995). The abilities of a musical savant and his family. *Journal of Autism and Developmental Disorders*, *25*(3), 231–248. https://doi.org/10.1007/BF02179286

Yu, K., Wang, R., Li, L., & Li, P. (2014). Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00729

Yu, K., Zhou, Y., Li, L., Su, J., Wang, R., & Li, P. (2017). The interaction between phonological information and pitch type at pre-attentive stage: An ERP study of lexical tones. *Language, Cognition and Neuroscience*, *32*(9), 1164–1175. https://doi.org/10.1080/23273798.2017.1310909

Yu, L. (2018). *An electrophysiological investigation of linguistic pitch processing in tonal-language-speaking children with autism (Unpublished doctoral dissertation)* [Doctoral thesis]. Department of Speech, Language, and Hearing Sciences, University of Minnesota, USA.

Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S., & Zhang, Y. (2015). Pitch processing in tonal-language-speaking children with autism: An event-related potential study. *Journal of Autism and Developmental Disorders*, *45*(11), 3656–3667. https://doi.org/10.1007/s10803-015-2510-x

Yu, L., Stronach, S., & Harrison, A. J. (2020). Public knowledge and stigma of autism spectrum disorder: Comparing China with the United States. *Autism*, 136236131990083. https://doi.org/10.1177/1362361319900839

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *130*(6), 4063–4069. https://doi.org/10.1121/1.3651818

Zaehle, T., Wüstenberg, T., Meyer, M., & Jäncke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fMRI study. *European Journal of Neuroscience*, *20*(9), 2447–2456. https://doi.org/10.1111/j.1460-9568.2004.03687.x

Zhang, C., Peng, G., & Wang, W. S.-Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *The Journal of the Acoustical Society of America*, *132*(2), 1088–1099. https://doi.org/10.1121/1.4731470

Zhang, C., Peng, G., & Wang, W. S.-Y. (2013). Achieving constancy in spoken word identification: Time course of talker normalization. *Brain and Language*, *126*(2), 193–202. https://doi.org/10.1016/j.bandl.2013.05.010

Zhang, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S. Y. (2018). The effect of speech variability on tonal language speakers' second language lexical tone learning. *Frontiers in Psychology*, *9*(OCT), 1–13. https://doi.org/10.3389/fpsyg.2018.01982

Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLOS ONE*, *6*(6), e20963. https://doi.org/10.1371/journal.pone.0020963

Zhang, Y. (2016). Categorical Perception. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese Language and Linguistics*. Leiden, Netherlands: Brill. https://doi.org/10.1163/2210-7363_ecll_COM_000071

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Tusler, L., & Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science*, *14*(3), 566–581. https://doi.org/10.1111/j.1467-7687.2010.01004.x

Zheng, H.-Y., Minett, J. W., Peng, G., & Wang, W. S.-Y. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes*, *27*(2), 184–209. https://doi.org/10.1080/01690965.2010.520493

Zheng, L., Grove, R., & Eapen, V. (2019). Spectrum or subtypes? A latent profile analysis of restricted and repetitive behaviours in autism. *Research in Autism Spectrum Disorders*, *57*, 46–54. https://doi.org/10.1016/j.rasd.2018.10.003