Tone Processing and the Acquisition of Tone in Mandarin- and English-Speaking Typically Developing Children and Children with Autism Spectrum Disorder

Ya-fang Lu

A dissertation submitted in partial fulfilment
Of the degree requirements for the degree of
Doctor of Philosophy
Of
University College London

Department of Linguistics
University College London

November, 2015

I, Ya-fang Lu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Ya-fang Lu 16/11/2015

Abstract

Autism Spectrum Disorder (ASD) is characterised by pervasive social difficulties, which partly manifest themselves in inappropriate pragmatics. It has also been hypothesised that individuals with ASD, or at least those on the lower-functioning end of the autism spectrum, may also have atypical pitch and musical perception. This thesis investigates pitch perception in autism in a domain where pitch is directly represented in the grammar: tones. Tone perception was investigated in a series of four experiments with high-functioning English and Mandarin ASD participants with and without language problems and their corresponding TD groups. The first experiment involved a tone comprehension task (only for the Mandarin participants) using picture-matching. The second experiment involved a psychoacoustic tone discrimination task using the Mandarin Tone 1-4 continuum. The third experiment was a categorical perception task involving two tasks: a naming task and a two-step identification task. The results of the experiments indicated subtle but persistent issues with the grammatical representation of tones for Mandarin ASD speakers, especially for those with language problems. Although ASD participants' tone comprehension and tone discrimination abilities are essentially in line with their typical peers, they have different error patterns in comprehension of Tone 2-3 distinctions and they treat nonce word stimuli more like pure tone stimuli in identification, suggesting a weaker representation of abstract tones. In addition, the categorical perception task revealed that although the performance of Mandarin ASD participants in the naming task was not distinguishable from their typically developing peers, the two-step identification task revealed a less strongly categorical perception of the Tone 1-4 continuum. In addition, the performance of the ASD SLP groups was also overall worse. These results altogether constitute a significant discovery of a grammatical impairment of people living with ASD. This population might have prosodic impairments relating their pitch perception, and their ability to categorise pitch contours in a grammatical fashion, in addition to their pragmatic difficulties.

Acknowledgments

I would like to express my deepest appreciation to my supervisor, Kriszta Szendroi, whose unparalleled experience and patience during my graduate studies enabled me to develop an understanding of this profound subject. I have been very fortunate to have an inspiring and supportive supervisor who provided me the opportunity to conduct the research on my own, and at the same time the guidance and encouragement when I encountered difficulties. She is willing to spend time discussing every single detail and giving me useful suggestions on the draft of this thesis. Without her guidance and persistent help this thesis would have been possible. I hope that one day I would become as great and generous an advisor to my students as she has been to me.

Also, I am heartily grateful to all the staff at University College London for their various kinds of support during the years. They are all remarkable teachers and have enlightened my linguistic study. I aim to become a linguist as outstanding as they are. I am thankful to my secondary supervisors, Andrea Santi, Alexandra Perovic, and Andrew Nevins, for their insightful comments and encouragement. My sincere thanks also go to Nausicaa Pouscoulous, for her enlightening comments and questions for the upgrade report. I would also like to thank Ad Neeleman, the head of research department, and Robyn Carston, my graduate tutor, for the various practical help during the PhD years. Last but not least, I am thankful to Richard Breheny, my programme director and personal tutor in MA Linguistics at UCL. It is he who has greatly broadened my horizon and leads me to pursue a higher degree in linguistics.

I would also like to thank the MA students at UCL, Hsiao-Chien Zheng, Eleanor Dolan, Laural Foreman, and Hanis Ramdzan for collecting the data from the Mandarin kindergarten children as well as English participants. In addition, I thank my fellow PhD students, Kevin Tang and Nick Neasom, for their stimulating discussions as well as enlightening comments. Many thanks to Felicity Deamer, Katharine Mair, Lichin Wang, and Tse Wang for recording the experimental stimuli for me.

Also, I am thankful to Judit Gervain from Paris Descartes University for the Mammouth software she developed and the great help for statistics she provided. I also need to thank René Kager, Liquan Liu, and Ao Chen from At Utrecht University for the help with the technical acoustic support with PRAAT.

I owe many thanks to schools and institutions where I collected the research data, Xiao-ma Autism Centre, Xin-xin Kindergarten, Kaohsiung Municipal Qi-shan Elementary School, Kaohsiung Municipal Xi-zhou Elementary School, Kaohsiung Municipal Min-zhu Elementary School, Kaohsiung Municipal Jhong-jheng Elementary School, Kaohsiung Municipal Qi-shan Junior High School, Kaohsiung Municipal Qi-mei Senior High School, and Kaohsiung Municipal Nan-zih Special School. I really appreciate their help and assistance in recruiting the participants and providing the silent environment for conducting the experiments.

Last but not least, I owe my heartfelt gratitude to my dear family and friends. Their love and care helped me overcome the setbacks and complete this dissertation. I would like to give my special thanks to my parents and sister for encouraging and supporting me to pursue the PhD degree. Most importantly, I would like to express my deepest gratitude to my husband, Tse Wang, by saying thank you and I love you.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1 Autism Spectrum Disorder: definition and core features

Autism spectrum disorder (ASD) is a complex and pervasive neurodevelopmental disorder characterised by a triad of impairments in reciprocal social interaction, communication and imagination, which includes a restricted repertoire of activity and interests (American Psychiatric Association, 1994; American Psychiatric Association, 2013; ICD-10, World Health Organisation, 1993). The symptoms of ASD are evident before 36 months of age (APA, 2000). Children with ASD are known to have impaired social ability, sometimes language delay and disorder, and rigid and repetitive behaviour. It is also important to emphasise that ASD is a spectrum disorder, ranging from severe autism with associated learning difficulties, to high-function autism (HFA) with normal non-verbal ability, but language delay and Asperger's Syndrome (AS) with no clinically significant language delay or disorder. Thus, linguistic ability within the ASD population is extremely heterogeneous. It is worth noting that the formal diagnoses of ASD by the American Psychiatric Association (2013) underwent a major change in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders. The separate diagnostic labels of Autistic Disorders, Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) as well as AS are now replaced by one umbrella term, "Autism Spectrum Disorder", for several reasons. Firstly, the previous way was insufficiently precise for a diagnosis. Different clinicians may have diagnosed the same person with different disorders. In addition, since ASD is defined by a common set of behaviours, it should be characterised by a single name with further distinctions made according to the levels of severity.

Social interaction skills are indispensable to perceive mental and emotional states, establish joint attention between conversation partners, and potentially even understand that others' mental state may be distinctive from one's own, the so-called theory of mind (ToM) skills (Baron-Cohen, Leslie, & Frith, 1985; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Charman, 2003; Morton, Haith, & Gibson, 1976). As a consequence, these social interaction skills are essential to properly learn and process the semantic and pragmatic aspects of language. This is also why social deficit is frequently seen as being the primary factor that causes the language problems in ASD. Among all language problems, children with ASD are particularly well known to have pragmatic problems; for instance, they tend to be literal in their interpretation of language and find it difficult to orientate appropriately to conversational situations (Tager-Flusberg, 1999). They also find it difficult to process irony, metaphors, and metonymy (Pexman et al., 2011; Rundblad & Annaz, 2010).

2 Autism and prosody[1]

When Kanner (1943) first delineated the autistic syndrome, he identified abnormal prosody as one of its core features. Prosody is a term that refers to the suprasegmental features of speech, including variations in pitch, duration, intensity, stress, rhythm, rate, pause, intonation, etc. This is an intrinsic determinant of the form of spoken language and carries lexical, morphosyntactic and pragmatic information in all languages of the world. Therefore, prosody can be utilised to help to recognise spoken words, compute syntactic structure, as well as process the structure of discourse (Cutler, Oahan, & van Donselaar, 1997).

Nonetheless, disordered expressive prosody is widely reported to occur in the speech of people with autism (for example, Baltaxe, 1984; Fine,

[1] This subsection draws on the exposition of the project aims by Szendroi et al. (2013).

[2] Of course, the material will not be directly matched across languages in the experiment, but rather it will be ensured that each set of stimuli adheres to the strongest possible set of experimental criteria within the respective languages (frequency, syllable frequency, onset frequenccy, imageability etc.

[3] Potentially, a schemata linking tunes to information structure content or other pragmatic meaning is available to English speakers, but it would not be activated by the stimuli in (5),

Bartoluccim, Ginsberg, and Szatmari, 1991; Frith, 1989; Happe, 1999; Shriberg, Paul, McSweeny, Klin, Cohen, and Volkmar, 2001). It is often noted that individuals with ASD have monotonic or machine-like intonation, deficit in the use of pitch and control of volume, deficiencies in vocal quality, and the use of aberrant stress patterns (Ghaziuddin & Gerstein, 1996; Shriberg et al., 2001). All of these characteristics lead to an unusual way of speaking, or an exotic accent in ASD; however, these prosodic deficits do not exist universally in ASD. Simmons and Baltaxe (1975) found that four of the seven adolescents with autism they studied had notable suprasegmental differences in their speech, and Paul et al. (2005) observed abnormal prosody in 47% of 30 speakers with ASD.

Peppe (2007) examined the receptive and expressive prosodic abilities in children with HFA and found that the clinical children performed significantly worse than the matched controls in the Affect subtests (both reception and expression) in which the distinction between liking and disliking a food item was used. The names of the food items were (generally) said with a rise-fall tune for “like” and a fall-rise for “dislike”. The child was then required to produce this distinction. Since the use of the affect tunes was for pragmatic purposes, it was hard to tell if the children with HFA found it difficult to detect the prosodic patterns, or if they could actually hear the subtle differences in prosody, and yet could not associate the prosodic patterns with certain emotions because of their impaired pragmatic skills.

Therefore, it is important to examine grammatical prosody in order to explore the prosodic skills without the interference of pragmatics. Chevallier, Noveck, Happe, and Wilson (2009) investigated the perception of grammatical prosody in adolescents with AS from three aspects, namely, the interpretation of word stress, the determination of grammatical pauses, and the discrimination of the declarative vs question contour. Firstly, they tested the participants’ ability to select the most appropriate stress pattern in a disyllabic word like “He got the best PREsent he could dream of.” vs. “I preSENT the late-night news.” In addition, they assessed the participants’ ability to take rhythm into account in chunking sequences of two or three words such as “Dragonfly and carrot” vs. “Dragon, fly, and carrot.” The last task they employed was the so-called Turn-end task. They examined the participants’ ability to distinguish questions from declaratives on the basis of prosodic cues only. For example, “This is a dog.” vs. “This is a dog?” The clinical participants performed as well as the typically-developing (TD) controls in all the three tasks; thus, the scholars concluded that grammatical prosody is spared in AS. Since the grammatical prosody was intact while the pragmatic prosody was impaired, they reasoned that there was actually no prosodic problem in AS. Instead, it was the pragmatic problem that led to the difficulty in understanding the pragmatic aspects of prosody.

Despite the fact that the findings in Chevallier et al. (2009) appear to provide a nice clean picture, there are some lose ends. Firstly, the clinical participants in Chevallier et al. were all diagnosed with AS, which has no clinically significant language delay or disorder. Since the clinical participants did not have a language problem, it is not surprising that they could perform as well as the TD controls in all the grammatical prosody tasks. In contrast, Peppe, Cleland, Gibbon, O’Hare, and Castilla (2011) found that HFA children performed significantly worse than their controls, who were matched on chronological age (CA), not only on pragmatic prosody tasks like Affect and Contrastive stress, but also on grammatical prosody tasks such as Chunking and Turn-end. Moreover, even the AS children performed significantly worse than the controls matched on CA on the Chunking task.

All the studies in the literature provide mixed evidence of the prosody in ASD. Although it is generally agreed that the pragmatic prosody is impaired in ASD, the performance of grammatical prosody is disputed, since some findings have shown that it is intact in ASD, whereas others indicate that it is actually impaired. To push a theory that identifies the core deficit in autism as one of impaired Theory of Mind such as Chevallier et al (2009) further, it may be argued that people with ASD initially have no difficulty in perceiving prosodic patterns; however, since these patterns are mainly used for discourse/pragmatic functions which are difficult for them or meaningless, they gradually pay less attention to prosody in general. This could result in a less than optimal performance, even in grammatical prosody tasks. In other words, the social impairment and pragmatic problem 'spills over' to language prosody.

Alternatively, it may be argued that there is actually a real prosodic problem in ASD, and that such children's language difficulties may, in part, be related to their atypical perception of the pitch, intensity etc. of the speech signal. More recently, it has been hypothesised that individuals with ASD, or at least those on the lower-functioning end of the autism spectrum, may also have enhanced pitch and musical perception (Bonnel et al., 2010; McCann & Peppe, 2003).

Heaton et al. (2008) examined the discrimination of pitch differences between pairs of words, nonce words, and non-speech pitch contour analogues in children with ASD and matched controls and found that ASD participants were more sensitive to pitch height differences than their matched controls across different types of auditory stimuli. The scholars also hypothesised that the enhanced auditory perception may hinder linguistic development; however, their findings were inconclusive. They found that two of the four ASD participants who scored above 90% in their most difficult auditory discrimination condition had very low scores for receptive language tasks. However, the scores of the other two individuals were within the normal range and there was a general tendency for a positive correlation between the language scores and the performance of the auditory discrimination tasks (Heaton et al., 2008). It is believed that it is possible that standardised receptive language tests are just too general to be sufficiently sensitive to identify the potentially negative effect of enhanced auditory processing on language abilities.

Thus, it is unclear overall whether and if, how and in what populations in ASD, the impaired pitch perception abilities may contribute to language problems. Therefore, this project seeks to compare the speech perception -

language acquisition link in typical and atypical development to better understand how auditory mechanisms contribute to language abilities. The ASD population represents a particularly interesting case to address this question. A cross-linguistic perspective is adopted by investigating typically-developing and ASD populations in Taiwan and the UK, to unravel universal and specific aspects of language development in these two trajectories. That is, the project would compare and contrast the participants from different language backgrounds: Mandarin Chinese and English. These two prevalent languages share some universal aspects just as all the other language in the world. Nevertheless, Mandarin Chinese and English have their specific aspects and distinctive features. While Mandarin Chinese, just as other tone languages, uses prosody to encode lexical and grammatical differences, English and other non-tone languages only utilize prosody to encode pragmatic and emotional information. In other words, the function of prosody dissociates in the two languages: in Mandarin it has a lexical and grammatical role, as well as a pragmatic one, while in English it only has the latter. This means that Mandarin is particularly well-suited for studying the understanding of prosody in populations living with ASD, because it allows for testing prosodic abilities independent of pragmatic ones.

In addition, it is proposed to test the hypothesis by exploring the auditory perception and language functions of two populations on the autism spectrum, namely, high-functioning ASD children with significant language problems (HFA-SLP) and ASD children with no significant language problems (HFA-NLP). This facilitates the teasing apart of the effects of autistic impairment in non-linguistic domains (e.g. social cognition, (Baron-Cohen, Leslie, & Frith, 1985), on the one hand, and atypical speech perception, on the other. From this point onward, when referring to 'ASD children', it means both groups.

HFA-SLP speakers perform significantly worse than age-matched and language-matched controls in expressive prosody (O'Connor, 2012; Paul, Augustyn, Klin, & Volkmar, 2005; PEPS-C, S. Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007), while HFA-NLP (a.k.a. AS) children do as well as their language-matched controls (Chevallier et al., 2009; Susan Peppé, Cleland, Gibbon, O'Hare, & Castilla, 2011). Work on the nature of the grammatical deficit in ASD is scarce, but the existing evidence points to several problem areas that distinguish HFA-SLP from HFA-NLP (e.g., Perovic, Modyanova, & Wexler, 2007, 2012). Thus, it seems that HFA-SLP speakers, but not HFA-NLP speakers, have grammatical problems, some of which, it is hypothesised, may have their origin in early problems with prosodic perception.

Despite the fact that a wide range of articles have assessed the prosodic ability in ASD in European languages (McCann & Peppe, 2003), few studies have investigated tone processing in tone languages. There is evidence to suggest that it takes more time and effort for children with ASD to acquire various tone patterns in tone languages than their typical peers. If enhanced pitch perception were to affect language development, this effect is expected to be stronger in a language that employs tones for lexical differences. In addition, acoustically speaking, tones are similar to intonational tunes; however, their function is not pragmatic. In this way, tone languages provide a good opportunity to study the grammatical use of prosody *without* the associated discourse-pragmatic features. Pitch contours that are acoustically similar to intonational tunes are part of the phonological description of lexical items in tone languages like Mandarin Chinese, which means that word meanings are discriminated, in addition to phonemic contrasts, by lexical tones. Given the lexical, and thus fully grammatical function of tones, *any* delay or deviance in their use would indicate linguistic problems, independent of the socio-communicative deficit of ASD.

The current study, to the best of our knowledge, is the first on tone perception in ASD. It serves to examine the development of tone perception in children with ASD. If children with ASD indeed perform differently from their TD counterparts, then it provides an opportunity to explore whether children with ASD display a delayed or a deviant developmental pattern. A delayed pattern would be identified if participants with ASD showed a pattern that is typical of somewhat younger TD controls. A deviant pattern would be identified if participants with ASD showed a pattern that TD participants do not display at any age. Moreover, a delayed developmental pattern indicates that children with ASD actually follow the same developmental path as the TD controls, but there are some specific hindrances to increase the difficulty for them. On the other hand, a deviant developmental pattern implies that children with ASD might perceive the tones in their own unique way and did not follow the same developmental path as the TD participants.

3 Mandarin lexical tones and their grammatical representation

Mandarin Chinese, one of the most prevalent languages in the world, is a tone language. While most European languages like English use prosody to encode pragmatic and emotional information, tone languages such as Mandarin utilise pitch differences mainly to encode lexical and grammatical differences. For example, in English, in the utterance "John.," the falling intonation expresses the declarative meaning , but in "John?," the rising intonation indicates the interrogative meaning. On the other hand, in Chinese, the consonant-vowel sequence [ma] pronounced with a high and level pitch means "mother," but the same sequence pronounced with a high falling pitch means "scold."

Mandarin has four lexical tones for stressed syllables, each of which is primarily recognised by fundamental frequency (F0) changes. The pitch contours of Tone 1 to Tone 4 are (i) flat and high, (ii) rising, (iii) low, and (iv) falling, respectively (Xu et al., 2004), as shown in Figure 1. In fact, Mandarin has a fifth tone in unstressed syllables. This is referred to as a neutral tone and its pitch value depends on its preceding full tone (Shen, 1990). The current study will only focus on the four lexical tones for stressed syllables, since the neutral tone for unstressed syllables does not have an associated pitch contour.

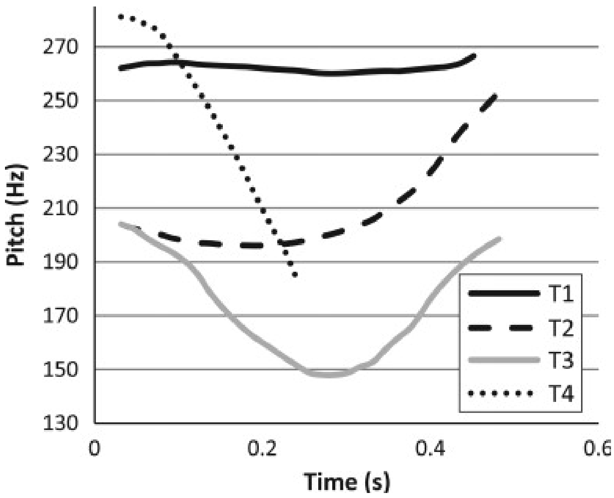

Figure 1. Mandarin tones (Hao 2012)

In addition to the F0 contour, tones in Mandarin differ in vowel duration (Fu & Zeng, 2000), syllable amplitude, and voice quality, such as creaky voice (Garding, Kratochvil, & Svantesson, 1986). Nonetheless, when it comes to the perception of tone, F0 contour is the primary cue and this alone is sufficient for listeners to distinguish various tones. As indicated by Massaro, Cohen, and Tseng (1985), the contribution of other acoustic characteristics is negligible in the presence of F0 information.

Let us turn to a more detailed explanation of how Mandarin tones are represented in the grammar of native speakers of the language. As already mentioned, in Mandarin, each lexical item is associated with one of the four tones. This carries crucial information: two words may differ *only* in their tones. One such minimal foursome is given in (1).

(1) a. shu1: book
    b. shu2: uncle
    c. shu3: mouse
    d. shu4: tree

Therefore, in this language, tone contours are considered to be phonemic in the sense that they constitute distinctive features. The tone it has is part of the description of a lexical item in the mental lexicon, see (2).

(2) a. [shu] – Tone 1

b. [shu] – Tone 4

Since native speakers can distinguish and produce nonce words with various tones, it can be concluded that they also store abstract schema such as the ones in (3):

(3) a. [ ]$_{word}$ – Tone 1

b. [ ]$_{word}$ – Tone 4

In English, lexical items cannot be distinguished on the basis of tone contours. Therefore, tone contours are not associated with lexical items in the lexicon. There is no reason to think that schemas such as (3) would be stored by English speakers, and of course they have no access to schemas like in (2).

Finally, it can be assumed that speakers of both languages would make use of general, low-level non-linguistic auditory abilities rather than linguistic knowledge of tones to distinguish pitch differences between pure tone contours.

4 Tone perception by native and non-native hearers

Heaton et al. (2008) examined the discrimination of pitch differences between pairs of words, nonce words, and non-speech pitch contour analogues in heterogeneous ASD children and matched controls and found that ASD participants were more sensitive to pitch height differences than their matched controls across different types of auditory stimuli. Both ASD and TD groups performed worse in discriminating pitch in speech stimuli than in non-speech stimuli (i.e. pure tones). There were no significant differences between real word and nonce word stimuli. Therefore, the following hypothesis of modularity can be formulated;

(4) Linguistic stimuli (both words and nonce words) activate specialised language systems. Pure tones activate general auditory processing systems.

The way in which this hypothesis would apply in the case of a specific pair of stimuli is as (5);

(5) a. shu1 'book'
b. shu4 'tree'

When a Mandarin speaker hears the stimuli in (5a) and (5b), a real word pair, this will activate the schemata in (2a) and (2b), respectively. In addition, the abstract schemata in (3) will be activated: (5a) would activate (3a), and (5b) would activate (3b). Given that tone is very much part of linguistic knowledge in Mandarin, it can be assumed that the activation of the linguistic schemata (2) and (3) would hinder any non-linguistic acoustic analysis of the stimuli. In other words, based on the hypothesis in (4), it is assumed that tone information, which is strongly linguistically-relevant in Mandarin, would be analysed primarily by the linguistic subsystems. Particularly since stimulus (5a) would activate (2a) and (3a), while stimulus (5b) would activate (2b) and (3b), there is sufficient linguistic information available to Mandarin speakers to distinguish these stimuli.

However, the situation would be quite different for an English speaker. Since the sample stimuli in (5) consist of syllables that do not violate the phonotactic rules of English, they can simply be used for expository purposes.[2] Since tones are not associated with lexical items in English, there are no concrete schemata of the type in (2) or abstract schemata of the type in (3) available to English speakers.[3] Thus, English speakers would find it

[2] Of course, the material will not be directly matched across languages in the experiment, but rather it will be ensured that each set of stimuli adheres to the strongest possible set of experimental criteria within the respective languages (frequency, syllable frequency, onset frequenccy, imageability etc.

[3] Potentially, a schemata linking tunes to information structure content or other pragmatic meaning is available to English speakers, but it would not be activated by the stimuli in (5), since these are single-syllable items.

very difficult to distinguish the stimuli based on their linguistic knowledge. Thus, their ability to distinguish the stimuli would necessarily invoke their non-linguistic general acoustic discriminatory abilities. However, since the stimuli in (5) involve linguistic material, the invocation of general acoustic mechanisms would be hindered by the hypothesis in (4).

As far as Mandarin ASD participants are concerned, their behaviour should depend on the strength of the grammatical representation of tones in their minds. We can distinguish three scenarios as in (6).

(6) a. Scenario1: NO DIFFERENCE
Mandarin ASD speakers, including Mandarin ASD speakers with significant language problems have identical grammatical representation of tones compared to age-matched typically developing children.

b. Scenario 2: SOME DIFFERENCE
Mandarin ASD speakers, or at least Mandarin ASD speakers with significant language problems have less strong grammatical representation of tones compared to age-matched typically developing children.

c. Scenario 3: FULL BREAKDOWN
Mandarin ASD speakers, including Mandarin ASD speakers with no language problems have weak grammatical representation of tones compared to age-matched typically developing children.

In Scenario 1, Mandarin ASD children are expected to pattern with Mandarin typically developing children in the current experiments. In Scenario 3, they are expected to pattern with English native speakers in the experiments. The interesting scenario is Scenario 2. One possibility for instance is if Mandarin ASD children, or at least Mandarin ASD children with language problems, have access to schemata like (2), but not to schemata like (3). In such a scenario, the prediction is that they would pattern with Mandarin typically developing children in their performance on stimuli involving real words, but for nonce words, they would rely on their general acoustic capacity, just like English children. As a result, their performance on nonce words would pattern with their performance on pure tone stimuli.

Let us now turn to a summary of the experiments I designed and performed to investigate the perception, comprehension and grammatical representation of tones in Mandarin and English individuals with or without ASD.

5 Plan of thesis

The aim of this doctoral thesis is to investigate auditory processing in ASD by taking a cross-linguistic perspective and comparing data from two distinctive languages, namely, English and Mandarin Chinese. If enhanced pitch perception were to affect language development, this effect is expected to be stronger in a language that employs tones for lexical differences. Thus, a potential impairment of prosody in Mandarin Chinese would potentially result in a more severe language breakdown in clinically equivalent populations. The results are expected to have important theoretical implications for the understanding of the interpretation of prosody in autism. Since prosodic studies in autism are scarce, the results will make a significant contribution to the literature.

Therefore, this project aims to chart the full territory between low-level pitch perception and higher level linguistic functions of prosody, beginning with the former and gradually increasing the level of abstractness. The aim of the project is to explore the interaction between speech perception and language acquisition in typical and atypical English-speaking and Mandarin-speaking populations. A series of experiments will be conducted in order to examine tone processing and the acquisition of tone in children with ASD.

The tone perception task in Chapter 2 will explore the perception of lexical tones by means of a picture-matching task, adopting the similar methodology in Wong et al. with some modifications to test the perception of lexical tones by Mandarin-speaking children with autism and their controls. The participants will also be presented with four pictures in each trial, as shown in Figure 2. However, the target word (e.g. hu3; tiger) will have a minimal pair that will only differ in tone (e.g. hu2; fox) in every trial. Besides, there is one phonetically similar foil which has the same tone as the target word (e.g. shu3; mouse), while the other foil is semantically related to the target word (e.g. shi1; lion). The main purpose of this perception test is to examine whether the children with ASD are able to match the auditory information to the actual linguistic items. Also, this study can facilitate a further investigation of the errors in tones, phonetic segments, and semantics in picture-matching tasks by children with ASD.

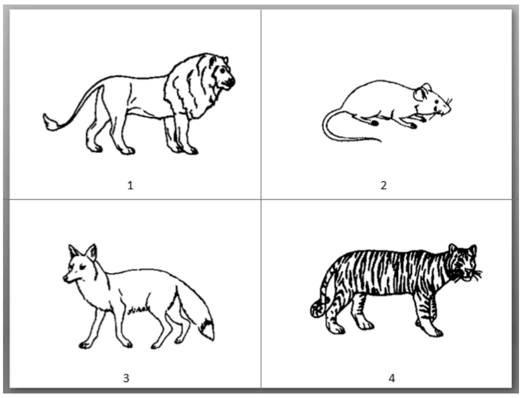

Figure 2. Illustration of tone perception task

The psychoacoustic tone discrimination task in Chapter 3 is designed to investigate whether speakers with ASD have an enhanced auditory perception of pitch contours, as well as to explore the potential differences between linguistic (real words and nonce words) and non-linguistic (pure tone) stimuli. The experiment will be performed in two languages, English and Mandarin, in order to reveal any language-particular differences. The participants will also undertake standardised language tests (PPVT, BPVS, and TROG). This allows for an investigation of a potential correlation between auditory perception abilities and general language abilities.

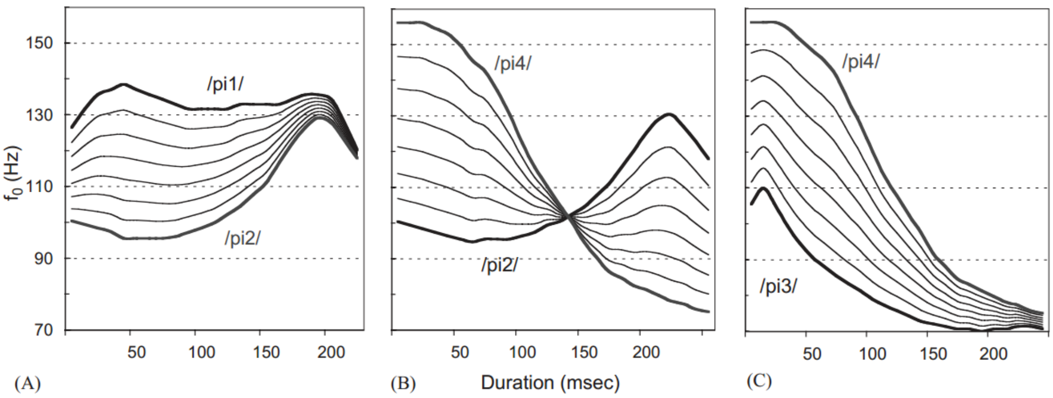

Figure 3. 2. (A-C) Tone contours for /pi/ in the three continua: (A) in Tone1 vs. Tone 2, (B) in Tone 2 vs. Tone 4, and (C) in Tone 3 vs. Tone 4

Chapter 4 will turn to explore the categorical perception of tones with a forced-choice identification task, as well as a two-step discrimination. This pair of experiments investigated whether individuals treat in-between items of Mandarin tones in a categorical way or in a psychophysical way. The methodology in Halle, Chang, and Best (2004) will be adopted to explore the identification of Mandarin tones by Mandarin-speaking children with autism, English-speaking children with autism, and their controls. It is expected that the English-speaking participants will process the tone information in a psychophysical way. On the other hand, it will be interesting to see if the Mandarin-speaking children with autism have a poorer performance than the TD Mandarin-speaking children due to the interference from the linguistic information and whether or not they will show the categorical perception of lexical tones. In addition, it is possible that the real word-nonce word distinction influences the results of the Mandarin-speaking ASD participants. This is because lexical top-down support is available for real words but not for nonce words. Finally, pure tone stimuli are potentially not treated as linguistic by at least some of our Mandarin participants and thus would not show categorical perception behaviour but a psychophysical one.

In this way, the following three chapters were designed to explore the tone processing and the acquisition of tone step by step. The tone comprehension task in Chapter 2 establishes our baseline in terms of how well participants are able to distinguish tonal minimal pairs in their comprehension. The psychoacoustic tone discrimination task in Chapter 3 then increases the perceptual difficulty and explores the auditory perception of pitch contours along the tone continuum. This task also uses nonce words and pure tones pitch contours to tap deeper into participants' auditory perception of pitch. The categorical perception task in Chapter 4 examines whether (and to what extent) lexical tones are represented in a categorical fashion by speakers.

6 Contribution of others

I would like to acknowledge here some contribution of other researchers to the studies reported in this work. For the general research idea and some of the interpretation of the results I partially relied on the project proposal of Szendroi et al. (2013). I invented the design and created the experimental materials for all four experiments. The experiments reported in Chapters 3 and 4 ran on software developed by Judit Gervain (Paris Decartes/ CNRS). I collected the data and Mandarin ASD NLP, Mandarin ASD SLP, Mandarin YTD and Mandarin OTD children for all experiments. The Mandarin KTD data for the experiments reported in Chapters 2-4 were collected by Hsiao-Chien Zheng and analysed in her MA dissertation Zheng (2014). The English data were collected by Eleanor Dolan, Laural Foreman, and Hanis Ramdzan and partially analysed in their MSc dissertations (Dolan, 2014; Foreman, 2014; Ramdzan, 2014), respectively. All the data analysis reported in this thesis is my own work, unless a citation is added in the text.

Chapter 2: Tone comprehension task

1 Introduction

In order to utilise the appropriate tone in tone languages, it is firstly essential to recognise the fact that different tones have different lexical meanings. Secondly, a certain auditory processing ability is required to perceive the different tones and distinguish their pitch contours. Having developed the ability to discriminate various lexical tones in the tone language, it is necessary to map and link the tonal categories to certain lexical items in order to comprehend the tones. Finally, some articulation skill is required to produce the intended tones. Therefore, this rationale is adopted in the following literature review.

Qin and Mok (2012) tested the speech and non-speech tonal perception of Cantonese of Mandarin-, English- and French-speaking adults with an AX discrimination paradigm. Two syllables, /jau/ and /se/, with six various lexical tones in Cantonese were utilised as speech stimuli. The non-speech pure tone stimuli were resynthesized from the six Cantonese tones and had an F0 contour similar to the corresponding speech stimuli. The results indicated that each group performed the non-speech task much better than the speech task. While all the groups performed equally well in the less demanding non-speech task, Mandarin-speaking adults were significantly more accomplished than the English and French groups in the speech task. The researchers concluded that native language experience had a huge influence on the perception of non-native speech tones.

Lee, Vakoch, and Wurm (1996) investigated the perception of four Mandarin lexical tones of Cantonese, Mandarin, and English native speakers (mean age: 23) with an AX task. The stimuli were 18 pairs of Mandarin words with exactly the same phoneme and tone, as well as 36 pairs with the same phoneme but a different tone. 18 of the 36 "different" pairs had both tones corresponding to Mandarin real words, while one of the tones of the other 18 corresponded to Mandarin real words and the other had no corresponding real word. The results showed that the Mandarin-speaking participants performed significantly better than the Cantonese- and English-speaking participants for both lexical and non-lexical tones, while the Cantonese-speaking participants performed better than the English-speaking participants for both Mandarin real words and nonce words. The researchers concluded that native speakers could better discriminate tones from their own language across the lexical status of tones.

Hume and Johnson (2003) explored the four lexical tones of Mandarin in perceptual spaces of Mandarin-speaking and English-speaking adults. Compared to the English native speakers, the overall perceptual space for Mandarin native speakers was expanded due to the fact that lexical tones are contrastive in Mandarin. Nevertheless, it is vital to note that the space between Tones 2 and 3 actually merges rather than expands for Mandarin native speakers and both Mandarin- and English-speaking participants found it extremely difficult to differentiate Tones 2 and 3. The researchers then utilised the sinewave analogues of four Mandarin tones in an AX discrimination paradigm and the results suggested that Tones 2 and 3 were better separated for the non-speech sinewave analogue stimuli in the perceptual space for both groups. This study demonstrated that, although both native and non-native speakers may find some pairs of tones (like Tone 2 vs. Tone 3) in Mandarin particularly hard to discriminate, it may be that non-speech stimuli with tonal information could make it less difficult for them due to the expanded space between Tones 2 and 3 for the non-speech sinewave analogue.

Huang (2007) also investigated the perception of Mandarin tones by Mandarin- and English-speaking adults with an AX discrimination paradigm. The participants were presented with 140 pairs of /bao1/, /bao2/, /bao3/, or /bao4/. The tones in a pair were either identical (e.g. bao1bao1) or different (e.g.bao1bao2), and the participants were required to judge whether the tones in the pair were identical or different by pressing corresponding buttons. The results indicated that both the Mandarin- and English-speaking adults found the Tone 2 versus Tone 3 pair the most confusing. They further revealed that the English-speaking adults greatly relied on the pitch onsets and offsets as phonetic cues in order to differentiate the tones; thus, they also found it difficult to discriminate the Tone 2-Tone 1 pair, as well as the Tone 2-Tone 4 pair. On the other hand, the Mandarin-speaking adults were able to detect the f0 contour on a monosyllable and found Tone 2-Tone 4 one of the easiest tone pairs to discriminate.

According to Zhu & Dodd (2000), tone has the highest saliency in Mandarin since it satisfies three important criteria in the notion of phonological saliency. Firstly, unlike some optional phonological units, such as syllable-initial consonants or syllable-final consonants, tone is compulsory for every syllable. Secondly, a change in tone distinguishes lexical information and affects lexical meaning. Thirdly, the component of Mandarin tones contains a comparatively small number of permissible choices. Unlike syllable-initial consonants, which provide twenty-one choices in Mandarin, tone is much more phonologically salient because it only has four alternative contrasts. As a result, TD Mandarin-speaking children acquire tone much earlier than syllable-initial consonants, syllables-final consonants, and vowels.

Despite lexical tones being widely regarded as the most distinctive characteristic and essential phonetic feature of Mandarin Chinese (Zhu & Dodd, 2006; Tsao, 2008), there are relatively few studies related to children's acquisition of Mandarin tones, especially the perception of tones. In terms of the production of lexical tones in Mandarin, a few longitudinal case studies based on observation before 1990 found a discrepancy in the participants' age at the onset of lexical tones. While Jeng (1979) observed that two of his participants could produce lexical tones at around the age of 1;6, the participants of Clumeck (1980) did not completely acquire tones until they were three years old. Besides, two of the children studied by Shiu (1990) had still not acquired Tones 2 and 3 when they were 2;4 and 3;0 years old.

More recently, Zhu and Dodd (2006) conducted a one-year longitudinal study of four Mandarin-speaking infants in Beijing, who were around one year old at the beginning of the data collection and two years old at the end. The data was collected every fifteen days and the children were recorded for about an hour while playing with their mothers. The speech samples were then transcribed using the International Phonetic Alphabet (IPA), and the results indicated that the four tones emerged by the age of 1;7; moreover, the four tones stabilised (accurate use of the tones in the speech sample was higher than two thirds of opportunities for the tones) by the age of 1.10. Although these longitudinal studies indicated that tonal acquisition starts early, they are not able to establish a reliable age when children master all tones.

A large cross-sectional study was also conducted to investigate 129 Mandarin-speaking children (age 1;6-4;6) in Beijing using a picture-naming task (Zhu & Dodd, 2006). The participants were asked to name 44 pictures with words of one to three syllables. The list of words included the most common nouns and daily phrases, such as “thank you” and “bye bye”, to ensure that even the youngest children knew and were able to utilise them. The results of the picture-naming task were then transcribed by Mandarin-speaking phoneticians, and since there were only two tonal errors, even in the youngest group, the researchers concluded that the youngest group (age 1;6-2;0) could properly produce the four lexical tones in various contexts. However, it is worth noting that the frequency and imageability of the words were not properly balanced. Besides, since these were words with one to three syllables, it is highly likely that the transcript of the tone production was automatically rectified by the extra information provided by the adjacent syllables.

Nevertheless, the studies by Wong, Schwartz, and Jenkins (2005) and Wong (2012a; 2012b; 2013) provided a different and insightful view of the production of lexical tones in Mandarin-speaking children. They conducted a series of picture-naming tasks on pre-school three-to-five year-old Mandarin-speaking children growing up in the United States and Taiwan. The pre-school children and the adults both produced the same set of monosyllabic Mandarin words, and all the productions were low-pass filtered to eliminate lexical information while retaining tonal data. These low-pass filtered monosyllabic sounds were then judged by ten Mandarin-speaking adults by identifying the name of the tone category (Tone 1, Tone 2, Tone 3, and Tone 4). Wong, Schwartz, and Jenkins (2005) found that the adult productions were more accurately identified than those of the three year-old children, and that most of the errors made by the children (10 out of 13) related to the production of Tone 3.

Wong (2012a) explored the same data as in Wong (2005) using seven acoustic characteristics and further emphasised that three year-olds did not produce adult-like tones in isolated monosyllabic words. Even when the production of the children was correctly categorised by adult judges, it was still phonetically different from the production of the adults. The most adult-like tone was Tone 4, Tone 1 was less adult-like, Tone 2 even less, and Tone 3 was found to be the least adult-like tone. Wong (2012b) further compared the Mandarin-speaking three year-old children growing up in the United States with the three year-olds growing up in Taiwan. Interestingly, the children growing up in Taiwan were found to make more errors in Tones 2 and 4 than the Mandarin-speaking children growing up in the United States, but again, none of the four tones produced by these two groups had adult-like accuracy. Wong (2013) further compared the monosyllabic Mandarin tones produced by 4- and 5 year-old Mandarin speaking children growing up in Taiwan with the production of three year-olds growing up in Taiwan in order to track the development of tone production in pre-school Mandarin-speaking children. The results suggested that none of the tones produced by the three age groups achieved adult-like perceived accuracy; in addition, there was no significant difference between the tone accuracy of the three age groups. Little development was observed in pre-school children's accuracy of the four tones in monosyllabic Mandarin words. However, despite the apparent ease and simplicity of the picture-naming tasks, the three year-olds expended a great deal of extra effort to process and retrieve the corresponding meanings and mapped sounds. It is also crucial to note that this series of picture-naming tasks tested the variability of pronunciation and this tended to blur whether the child was attempting the wrong tone or attempting the right tone, but articulating it wrongly.

In terms of Mandarin tones, Shi (2010) examined 18 monolingual Mandarin-learning infants (four-to-six months in the younger age group and eight-to-eleven months in the older age group) with a minimal tonal pair /mi2/ and /mi4/ with a visual fixation procedure. In the Familiarisation phase, the infants were presented with trials containing the target tone /mi2/ or /mi4/ in citation syllabic forms for thirty seconds. Then, the Test phase began and the infants were presented with disyllabic utterances containing /mi2/ or /mi4/. The test was designed to determine if the Mandarin-learning young infants were able to recognise the target tone (in the Familiarisation phase) in variable tonal contexts (in the Test phase). The results indicated that, while the younger age group looked at target and non-target sequences (indicating listening times) for a similar length of time, the older age group looked at the target sequences significantly longer. As a consequence, Shi (2010) concluded that Mandarin-learning infants aged eight-to-eleven months began to categorise Tone 2 and Tone 4 in Mandarin in variable tonal contexts, but not those aged four-to-six months.

Gao, Shi, and Li (2010) examined 20 monolingual Mandarin-learning infants aged four-to-thirteen months with a minimal tonal pair /tsan2/ and /tsan3/ with a visual fixation procedure. In the Familiarisation phase, the infants were presented with trials containing seven tokens of the target tone. Then the Test phase began and the infants were either presented with a

“Same” type or “Different” type trial. Six novel tokens of the same tonal category were presented in the “Same” type trial, whereas six novel tokens of the contrasting tone were presented in the “Different” type trial. The results showed that, while there was no significant difference between the time spent looking in the “Same” test trial and the last trial of the Familiarisation phase, the looking time in the “Different” test trial was significantly longer than in the last trial of the Familiarisation phase. This suggested that even Mandarin-learning infants aged four-to-thirteen months were able to perceive the difference between Tone 2 and Tone 3 in Mandarin. These infant perception studies are very informative in the sense that they indicated that infants are sensitive to the relevant tonal differences in linguistic stimuli at a very young age. However, perception is just the first step toward acquisition; the children also need to be able to form the tonal categories, store them as long-term memory units, and associate tonal information with individual lexical items.

Some studies comparing infants exposed to Mandarin to those who have not been exposed to a tone language indicated that they have begun to form tonal categories in the long-term memory. Mattock and Burnham (2006) examined the speech and non-speech tone discrimination in Mandarin- and English-learning infants (6 month-olds and 9 month-olds) using a head-turn procedure and two speech minimal tonal pairs: contour-contour contrast /ba2/ and /ba4/, and level-contour contrast /ba3/ and /ba4/. The non-speech tonal pairs had corresponding speech pairs: rising versus falling violin sounds, and rising versus low violin sounds. In the Familiarisation phase, the infants were repeatedly presented with one stimulus of the pair. Then the Test phase began and they were presented with a change (the other stimulus of the pair) or no-change (still the same stimulus as in the familiarisation phase) trial. The results indicated that Mandarin-learning 6 month-olds and 9 month-olds performed equally well in discriminating both speech and non-speech tones, whereas English-learning 9 month-olds’ speech tone discrimination declined compared to their Mandarin-learning counterparts.

The authors believed that this was evidence of perceptual reorganisation in the first year in a non-tone language environment.

Liu & Kager (2011) also conducted discrimination task of Mandarin lexical tones in 122 5-6 and 11-12-month-old Dutch-learning infants with a head-turn procedure. The Mandarin syllables /ta/ with four lexical tones was recorded for the stimuli. The stimuli were continuum along Tone 1 and Tone 4, and each continuum proceeded through eight steps from one endpoint (step 1) to the other (step 8). Following a statistical learning paradigm (Maye, Weiss, & Aslin, 2008), there were three conditions for the Familiarisation phase (FAM) (unimodal frequency distribution, bimodal frequency distribution, and no familiarisation phase). Then infants were repeatedly exposed to stimulus with Step 6 in the habituation phase (HAB). After the infants were habituated on the stimulus with step 6, they went to the dishabituation phase (DIS) and heard two tokens with Step 3 on the tone continuum.

The results indicated that Dutch-learning 5-6-month-olds had significantly longer looking time in the two DIS tokens than in the last two HAB trials under unimodal and as well as bimodal conditions, suggesting that they could discriminate the non-native lexical tones, and this tonal sensitivity outweighed the effects of statistical learning. On the other hand, the 11-12-month-olds could only differentiate the non-native lexical tones in the bimodal condition, showing that statistical learning influenced the perception of non-native lexical tones. The authors believed just as the study by Mattock and Burnham (2006), this was evidence of perceptual reorganisation in the first year in a non-tone language environment. While the Dutch-learning 5-6-month-olds demonstrated the early tonal perception, the effects of the tonal perception were partly reversed by statistical learning for Dutch-learning 11-12-month-olds.

The next study indicated that proper long-term tonal category formation occurs significantly later than infancy, at least for Tones 2 and 3. Liu, Tsao, Chang, and Hsu (2013) tested 150 Mandarin-speaking children (aged 4-8) with an AX discrimination task on a minimal tonal pair /i2/ and /i3/. The participants would hear two stimuli in each trial and there were four possibilities for these two stimuli: /i2i2/ or /i2i3/ or /i3i2/ or /i3i3/. They would have to decide if these two stimuli were the same or not by pressing buttons. The results illustrated that the percentage of correct responses of 4 year-olds (64%) was significantly lower than of 6 year-olds (78%), 7 year-olds (80%) and 8 year-olds (88%); furthermore, the percentage of correct responses of 5 year-olds (71%) was significantly lower than that of 8 year-olds. The researchers concluded that the perception of lexical tones is gradually developed between the ages of four and eight.

With regard to tone comprehension, Clumeck (1977) examined two children who were born and raised in monolingual, Mandarin-speaking families in the United States, and found that they were unable to differentiate Tone 2 from Tone 3 in an object identification task at ages 3;4 and 2;9, respectively. In addition, Wong (2005) implemented a picture-matching task with thirteen three year-old Mandarin-speaking children, who were raised in monolingual Mandarin-speaking families in the United States. The twenty-four monosyllabic Mandarin words stimuli were chosen from the top 250 most frequently used words in the spontaneous speech of American pre-school children (Hall, Nagy, & Linn, 1984). Half of them were single, like hua1 "flower" and ma3 "horse", and the other half were six minimal pairs only different in tone, such as shu1 "book" versus shu4 "tree" and yu2 "fish" versus yu3 "rain". The six possible tone contrasts in Mandarin were all included in the task (Tone 1 vs. Tone 2, Tone 1 vs. Tone 3, Tone 1 vs. Tone 4, Tone 2 vs. Tone 3, Tone 2 vs. Tone 4, and Tone 3 vs. Tone 4). Twenty-four of the total thirty-six trials were utilised to test the perception of the twenty-four monosyllabic Mandarin words without a minimal pair counterpart among the foils, while the other twelve trials were designed to retest the six minimal pairs with a minimal pair counterpart among the foils. The examiner presented four pictures to the participants in each trial and asked them, “Which one is …?”. At least one picture represented the same tone as the target word.

The results indicated that the three year-olds were able to accurately perceive the four tones in the monosyllabic words with segmental support; in other words, when there was no minimal pair counterpart among the foils, even the three year-olds could perceive the four tones equally well based solely on the segmental information. However, when there were insufficient segmental cues to complete the picture-matching task, i.e. with minimal pairs that had the same segmental structure and only differed in tone, the three year-olds could only discriminate Tones 1, 2, and 4 with relatively high accuracy (90%, 87%, and 90%, respectively), and when the target tone was Tone 3, the responses were only correct 69% of the time. Tone 2 was misidentified as the target tone 21% of the time, and Tone 4 was mislabelled as the target tone 8% of the time.

In fact, when Wong (2005) conducted a correlational analysis of the tone production task and the tone perception task, intriguingly, she found that there was no correlation between children’s production and perception of tone. It is possible that the six minimal pairs were inadequate for demonstrating the corresponding links. In addition, the author mentioned that the picture-matching task for the tone perception test was much easier than the picture-naming task for the tone production test; thus, these two tasks involved considerably different task demands, especially for young children. It would have been better if the stimuli in the picture-naming task had been consistent recordings rather than the varied pronunciation uttered by the examiners because this would have ensured that all participants had received identical standard input. Finally, much care should be taken in terms of the frequency and imageability of the stimuli. The researcher chose Mandarin monosyllabic words solely based on the frequent word list of American children; yet, cultural diversity and the nature of the language may result in huge differences in frequently used words, even for young children. Therefore, is worth making an effort to check the frequency of these words in the Mandarin corpora. Imageability is also an important factor of a picture-matching task. The minimal pair, jiao3 “foot” and jiao4 “shout” utilised by Wong is an example of imbalanced imageability. While jiao3 “foot” was frequent daily vocabulary with high imageability, jiao4 “shout” was a verb with much less imageability. Consequently, extra care and attention should be paid when choosing stimuli in the picture-matching task to ensure that they are used with similar frequency and comparable imageability.

This brief overview of the existing literature related to the production, perception and comprehension of the acquisition of Mandarin tones now leads to more specific findings regarding the order in which different tones are acquired. Although some researchers, such as Clumeck (1977), found that Tone 2 was acquired first when the participant began to use words at 1;10, most findings demonstrate that Tone 1 and Tone 4 are acquired before Tone 2 and Tone 3. (Li & Thompson, 1997; Shiu, 1990; Zhu & Dodd, 2006). Li and Thompson (1977) conducted a cross-sectional study of children’s production of Mandarin tones with seventeen Mandarin-speaking children (age 1;6-3;0) by means of a picture-naming task, and proposed that Tone 1 and Tone 4 were initially predominant at an early stage; then, the four tones could all be produced at the next stage, but there was consistent confusion between Tone 2 and Tone 3, and Tone 2 and Tone 3 could finally be produced distinctly without confusion in the later stage. The one-year long longitudinal study of four Mandarin-learning infants by Zhu and Dodd (2006) also suggested that Tone 1 and Tone 4 emerged first (around the age of 1;2), followed by Tone 2 (around the age of 1;3). Tone 3 was the last to emerge (around the age of 1;5). A similar pattern was also found with the age of stabilisation using a criterion of 66.7%: Tone 1 was stabilised first, followed by Tone 4, while Tones 2 and 3 were the last to be acquired.

In fact, Tones 2 and 3 are claimed to be the most difficult contrast for Mandarin-acquiring infants, second language learners, and even for Mandarin-speaking adults (Gandour, 1978; Huang, 2004; Li & Thompson, 1977), and the confusion may be due to the physical similarity of the two tones (Blicher et al., 1990; Shen & Lin, 1991). Because they both start at about the same pitch level (Gandour, 1978) and have a dynamic F0 shape, Tone 2 and Tone 3 exhibit a similar F0 contour, which leads to perceptual difficulties. Moreover, the phonological connection between Tones 2 and 3 also plays a role in the confusion. Lexical tones sometimes undergo modification when they occur in combination, and this kind of tonal change is called tone sandhi. One of the sandhi rules in Mandarin is that, if a word with Tone 3 is followed by another word with Tone 3, the first word must be changed to be with Tone 2. As a result of this implicit and untaught phonological rule, the input of Tones 2 and 3 varies and neutralises the distinction of Tones 2 and 3 to some extent (Gandour, 1978). Wang, Jongman, and Sereno (2003) indicated that it is perceptually effective to tell the difference between Tones 2 and 3 by recognising the "turning point," which is a point in time at which the pitch contour in Tone 3 changes from falling to rising. However, by the age of 3, children have not learnt to recognise and/or utilise this characteristic to properly differentiate Tone 2 from Tone 3.

There are few studies related to the acquisition of lexical tones of Mandarin-speaking children with SLI. Zhu & Dodd (2000) explored the phonological production of thirty-three Mandarin-speaking children (aged 2;8-7;6) with atypical speech development. These children suffered from functional phonological disorder, but they all met the criteria, such as normal hearing, normal oral structure, normal language comprehension, and no hearing or behavioural problems. Each child was required to complete the picture-naming task in single words, as well as the picture-describing task in connected speech. The results indicated that even the most severely disordered children had no problem with tones and seldom made tonal errors. When errors did occur, the patterns were as follows: Tone 4 was the most frequent substitute for Tone 1, Tone 1 was the most frequent substitute for Tone 4, Tone 2 and Tone 3, and Tone 2 was also used as a substitute for Tone 3.

Similar patterns were preliminarily observed in Mandarin-speaking children with autism based on personal experience. Despite the fact that they could pronounce Tone 1 fairly well, they sometimes uttered Tone 1 when they attempted Tone 2 or Tone 4; in addition, they often mispronounced Tone 3 as Tone 2. These intriguing phenomena imply that it is vital to explore tone processing and the acquisition of lexical tones of Mandarin-speaking children with autism in order to further assess the development of language and focus in ASD.

To sum up, some researchers believe that three year-olds have completely acquired tones and could properly produce the four lexical tones in various contexts (Clumeck, 1980; Zhu and Dodd, 2006). Even children with SLI have stabilised perception and comprehension of lexical tones at this age (Zhu & Dodd, 2000). On the other hand, other scholars suggest three-year-olds have still not acquired all the lexical tones (Shiu 1990) and are unable to accurately perceive the tones in the monosyllabic words without segmental support (Wong, 2005). In terms of order of acquisition of lexical tones, most findings demonstrate that Tone 1 and Tone 4 are acquired before Tone 2 and Tone 3. (Li & Thompson, 1997; Shiu, 1990; Zhu & Dodd, 2006). Liu, Tsao, Chang, and Hsu (2013) had an AX discrimination task on Tone 2 versus Tone 3, and it turned out that the percentage of correct responses of six year-olds was around 78%. Therefore, it was decided to choose six year-olds as our YTD, since they have completely acquired lexical tones and should be able to perform the tasks in our project.

2 Research questions

The research questions are as follows;

Q1: Can typically developing children and ASD children comprehend Tones 1, 2, 3, and 4 in Mandarin? According to the literature, typically developing children acquire Tones 1 and 4 before they acquire Tones 2 and 3. Thus, the specific question to be asked is as follows;

Q1a: Do typically developing children comprehend Tones 1 and 4 better than Tones 2 and 3?
Q1b: Does the pattern of acquisition found in typically developing children extend to children with ASD (including those with or without language problems (ASD-SLP, ASD-NLP)), or do children with ASD display a deviant (not just delayed) developmental pattern?
Q1c: Do children with ASD (including children with or without language problems (ASD-SLP, ASD-NLP)) display the same rate of acquisition of tones in their comprehension as typically developing children?

Q2a: When tonal errors occur in comprehension, what is the pattern of those errors for typically developing children?
Q2b: Does the pattern of tonal errors found in typically developing children extend to those with ASD (including children with or without language problems (ASD-SLP, ASD-NLP)), or do children with ASD display a different pattern?

Q3a: Do typically developing children make semantic or phonetic errors in their comprehension of freestanding lexical items?
Q3b: Is the rate of ASD children's semantic and phonetic errors similar to that of typically developing children?

3 Method

The task was a comprehension task of freestanding single morpheme words using picture-matching.

3.1 Participants

The participants were twenty-two monolingual Mandarin-speaking children with ASD (ASD) (6;6-18;11), eighteen typically developing 6 year-old kindergarten children (KTD), ten typically developing 9 year-old young children (YTD), and nine typically developing 17 year-old adolescents (OTD). One ASD child was excluded from the analysis of all the experiments because of the inability to properly complete any of the tasks. Thirteen of the children with ASD were recruited from an after-school programme for children with Autism in Kaohsiung City in Taiwan, whereas nine of the participants with ASD were recruited from schools in Taipei City in Taiwan. The ten typically developing nine year-old young children (YTD) and nine typically developing 17 year-old adolescents (OTD) were recruited from an elementary school, a middle school, and a high school in Kaohsiung City in Taiwan. Eighteen typically developing 6 year-old children were recruited from a kindergarten in Taoyuan, Taiwan.

All the participants were required to have (i) no hearing problems[4]; (ii) no major physical disability or structural abnormality of the vocal tract; (iii) non-verbal ability within the normal range (i.e. IQ:>70); (iv) Mandarin as a first language and the main language used at home. All the clinical children had undergone a multi-disciplinary assessment of their communication disorder and been diagnosed by a paediatrician as living with autism with normal cognitive ability and early delay in speech/language development, according to ICD-10 (WHO, 1993). One participant was diagnosed as having AS; however, according to the latest version of DSM-V published in 2013, Asperger's Syndrome has recently been eliminated as a separate diagnosis and is now included under ASD. It was not physically possible to conduct a clinical assessment to reconfirm the diagnosis.

In addition, the Wechsler Abbreviated Scale of Intelligence II (WASI-II; Wechsler, 2011) was utilised to test the non-verbal general cognitive abilities (non-verbal intelligence quotient; NV-IQ) of children with ASD. The WASI was developed to meet the demand for a short and reliable measure of intelligence in clinical settings for participants aged between six and ninety. It contains 4 sub-tests: Vocabulary, Block Design, Similarities, and Matrix Reasoning, yet only Block Design and Matrix Reasoning are utilised to examine the NV-IQ. The Block Design sub-test consists of a set of 13 printed two-dimensional geometric patterns that the examinee is

[4] Although this characteristic was simply assessed by means of a discussion with carers this time, a proper hearing assessment test may be considered for application in the future.

required to replicate in a specific time utilising two-colour cubes. The Matrix Reasoning sub-test is a series of 35 incomplete gridded patterns that the examinee completes by pointing to or stating the number of the correct response from five possible choices. The raw scores from the sub-test can be converted to t-scores and NV-IQ using the norms provided in the manual.

In terms of the participants' linguistic abilities, an attempt was made to follow earlier studies' assessment practice (Chevallier et al., 2009; Heaton et al., 2008; Peppé et al., 2007; Peppé et al., 2011); unfortunately, this attempt was futile, since there is no reliable standardised grammatical test to examine children's grammatical ability in Taiwan. The standardised receptive vocabulary test (Mandarin version of Peabody Picture Vocabulary Test-Revised; PPVT-R; Lu & Liu; 1998) and a translated version of a receptive grammar test (Test for Reception of Grammar-II; TROG-II; Bishop, 2003; Lin and Chi, 2007) were used in a pilot study. These were regarded as being suitable tools to investigate vocabulary in children, even those with cognitive impairment, as they both involved simple picture-matching tasks. Since a Mandarin version of PPVT is widely used in Taiwan (Wang & Lin, 2008) for research on children's education as well as special education, the PPVT Verbal Mental Age equivalent (PPVT VMA) was found to be more reliable and closer to CA than the TROG Verbal Mental Age equivalent (TROG VMA). On the other hand, the TROG VMA was consistently significantly lower than CA, even for TD children; thus, it was decided to abandon it for the main study and only utilise the standardised PPVT-R to explore the participants' receptive vocabulary. Although vocabulary tests do not directly test grammatical abilities, the scores of vocabulary tests tend to strongly correlate with those of grammatical tests (Fisher et al., 2005).

In addition, an attempt was made to establish the history of language onset by means of a parental questionnaire (occurrence of first words, and productive vocabulary at 24 months, e.g. Girolametto et al., 2001; Mirak &

Rescorla, 1998) supplemented by information from the child's health records; however, this was unsuccessful. The Mandarin ASD group turned out to show a bimodal age distribution, so we decided to split the group into two. The five children under ten-years old were identified as young participants with ASD, while the remaining sixteen children, who were at least thirteen years old, as older participants with ASD. The psychometric data of children in sub-groups is shown in Table 1.

		Mandarin KTD (*n = 18*)	Mandarin YTD (*n = 10*)	Mandarin OTD (*n = 9*)	Mandarin Younger ASD (*n = 5*)	Mandarin Older ASD (*n = 16*)
CA (months)	Mean	74	104	200	103	200
	SD	5	28	17	15	17
	Range	64-80	81-146	172-215	78-117	157-227

Table 1. Psychometric data of the chronological age (CA) of children in sub-groups

A one-way ANOVA was performed to explore the chronological age (CA) of the between subject factor, Group (Mandarin KTD, Mandarin YTD, Mandarin OTD, Mandarin Younger ASD, Mandarin Older ASD). Group ($F(4, 53) = 162.298$, $p < .001$) has a main effect. Bonferroni post-hoc tests indicated that these five groups were significantly different from each other (all *p*s < .015) except for two conditions. One is that the CA of Mandarin Younger ASD was not significantly different from Mandarin YTD (p = 1.000), and the other is that the CA of Mandarin Older ASD was not significantly different from Mandarin OTD (p = 1.000). Therefore, Mandarin Younger ASD could be considered to have similar age as Mandarin YTD, while Mandarin Older ASD had similar age as Mandarin OTD.

Because of the great heterogeneity in the linguistic abilities of the ASD population, it was proposed that the hypothesis should be tested by exploring the auditory perception and language functions of two populations on the autism spectrum: high-functioning ASD children with significant language problems (HFA-SLP) and ASD children with no significant language problems (HFA-NLP). This would facilitate the separation of the effects of autistic impairment in non-linguistic domains (e.g. social cognition, (Baron-Cohen, Leslie, & Frith, 1985), on the one hand and atypical speech perception on the other. Henceforth, 'ASD children' will refer to both groups. (this is reflected by the distinction between 'autism' and 'Asperger's Syndrome' in DSM-IV: individuals with ASD who have significant language problems are labelled autistic, whereas those without delayed or impaired language development are diagnosed as having Asperger's Syndrome. In line with the impending DSM-V, this study refrains from using such labels.) The performance of HFA-SLP speakers in expressive prosody was significantly worse than that of the age-matched and language-matched controls (O'Connor, 2012; Paul, Augustyn, Klin, & Volkmar, 2005; PEPS-C, Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007), whereas HFA-NLP (a.k.a. Asperger's) children performed as well as the language-matched controls (Chevallier et al., 2009; Peppé, Cleland, Gibbon, O'Hare, & Castilla, 2011).

As a consequence, our older ASD participants were split into two groups, Mandarin SLP and Mandarin NLP, based on their language ability. The younger group of ASD participants were comprised of individuals who had no linguistic delays. Consequently, this group is labelled as Mandarin YNLP. Like the TD participants, the Mandarin NLP and Mandarin YNLP had age-appropriate receptive vocabulary ability, but the receptive vocabulary ability of the Mandarin SLP was lower than age-appropriate, and a delay in language onset was reported. Delay was understood as (i) the age-equivalent score on PPVT-R (PPVT-VMA) at least one year lower than CA or (ii) standard scores not in the normal range, i.e. at least 1 standard deviation lower than 100. It is worth noting that the Mandarin version of PPVT-R only applies to young children aged 3 to 12 years; thus, the highest equivalent age it provides is 12+ years. Since all of the Mandarin NLP participants had extremely high raw scores and thus an age equivalent to 12+ years, the PPVT VMA greatly underestimated their real linguistic ability.

The data of the Mandarin NLP was subjected to a paired samples T test with CA and PPVT VMA as the paired variables. The analysis showed no significant difference between the paired variables, CA and PPVT VMA, in Mandarin NLP ($t(5) = .65$, $p = .554$). On the other hand, a paired samples T test on the data of the Mandarin SLP with CA and PPVT VMA as the paired variables indicated a significant difference ($t(5) = 5.62$, $p = .002$). It can be concluded from the mean of the CA and PPVT VMA and the direction of the t-value, that the PPVT VMA was significantly lower than the CA in the Mandarin SLP, suggesting that the Mandarin SLP participants experienced language delay. A paired samples T test on the data of the Mandarin YNLP with CA and PPVT VMA as the paired variables indicated no significant differences between the paired variables, CA and PPVT VMA, in Mandarin YNLP ($t(4) = -1.257$, $p = .277$). Just as the TD participants as well as Mandarin NLP, the Mandarin YNLP had age-appropriate receptive vocabulary ability. The psychometric data for the sub-groups of the retained twenty-one ASD children is shown in Table 2.

		ASD ($n = 21$)	Mandarin NLP ($n = 7$)	Mandarin SLP ($n = 9$)	Mandarin YNLP ($n = 5$)
CA (months)	Mean	177	214	188	103
	SD	45	9	14	16
	Range	78-227	197-227	156-203	78-120
PPVT VMA (months)	Mean	130	144+	122	123
	SD	25	0	33	25
	Range	70-144+	144+	70-144	96-144
NV-IQ	Mean	90	91	88	92
	SD	13	9	18	8
	Range	70-113	83-103	70-113	82-102

Table 2. Psychometric data for chronological age (CA), PPVT verbal mental age (PPVT VMA), and non-verbal IQ (NV-IQ) of sub-groups of ASD children

3.2 Task

The tone comprehension was tested with a picture-matching task. This methodology seemed to be appropriate for the population because of its relative simplicity. It is essentially the same format as the PPVT test.

3.3 Stimuli

The stimuli were 32 monosyllabic real Mandarin words and 32 sets of 4 pictures. An example is provided in Figure 4. One of the pictures matched the target word (e.g. hu3; tiger), while another showed a minimal pair which only differed in tone (e.g. hu2; fox). In addition, one picture contained an object with a phonetically similar name to the target word with an identical tone to that of the target word (e.g. shu3; mouse). Finally, there was one semantically-related picture (e.g. shi1; lion). The semantic distractor was always both phonetically different from the target word and had a different tone.

The words used were selected according to stringent criteria, the first of which was that they should all be meaningful with all four tones; for example, hu1 "to exhale", hu2 "fox", hu3 "tiger", and hu4 "to protect" are all lexical items in Mandarin. Secondly, the target words, as well as all the distractors, had to be comparable in terms of imageability. Thirdly, since it is well-known that frequency influences lexical retrieval (Gardner, Rothkopf, Lapan, & Lafferty, 1987) (Martin, Weisberg, & Saffran, 1989), great care was taken to make sure that the target word and the distractors have comparable frequencies. Eight of the target words had Tone 1, Tone 2, Tone 3 and Tone 4, respectively. The tonal distractors were equally distributed across the three possibilities as much as possible so that all possible tone-pair combinations were tested.

The audio stimuli were recorded by me, a female native Mandarin speaker. The recordings were phonetically analysed to ensure that the correct tone was used. All the recordings were pre-tested for comprehension and tone identification by native Mandarin-speaking adults.

In terms of the visual stimuli, the target and distractor pictures were counter-balanced by left-right and up-down dimensions. Some of the pictures were taken from a set of 260 standard pictures of Snodgrass (1980), while others were taken from the Internet. However, they were all adjusted to make them black and white with a consistent style and the pattern.

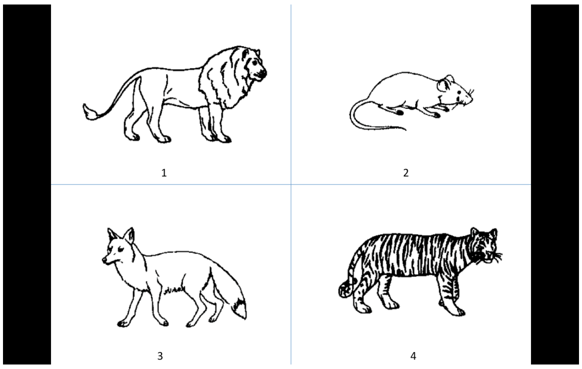

Figure 4. Example of visual stimulus for Experiment 2 with target word hu3 “tiger” in option four, tonal distractor hu2 “fox” in option three, phonetic distractor shu3 “mouse” in option two, and semantic distractor shi1 “lion” in option one

3.4 Procedure

Ethical approval for the study was gained from the UCL Research Ethics Committee (UCL Ethics Project ID Number: 0987/002). The information and consent sheets were given to the participants’ parents to understand the nature of the study in advance. In addtion, all the participants agreed to attend the study after the tasks were explained to them in age appropriate language. They were encourged to complete all the tasks, but they were also informed that they were free to leave the study at any time.

In this task, the participants heard one stimulus per trial, and they were asked to match the stimulus to the picture on the screen. One stimulus was played as each trial was presented. Then, the examiner asked: ‘Which one is …?’ and played the recorded stimulus once. The participants’ task was to point to the corresponding picture. If a participant did not know the answer, the examiner encouraged him/her to make a guess. The stimuli were organised into two opposite orders of presentation to avoid an order effect. The participants were asked to complete the two opposite orders of presentation in-between other tasks. Half of the participants completed the first order first and the other half finished the second order first.

4 Results

4.1 Preparation of data for analysis

Three items were removed because of their low imageability and subsequent high percentage of misunderstanding (ya3 'elegant' , ya4 'surprised', jie4 'to borrow'). Therefore, the number of items in each tone was unequal; this meant that there were seven items with Tone 1, eight items with Tone 2, eight items with Tone 3, and six items with Tone 4. As a consequence, a percentage of correct responses were used instead of the actual number of correct responses when analysing the data.

4.2 Data analysis

4.2.1 Breakdown according to response type

On average, all the participants got 81% of the answers correct in this task. As for the types of error, the percentage of tonal errors (14%) was much higher than that of phonetic (2%) and semantic errors (2%). (See Table 3.) Mandarin OTD had the highest percentage of correct responses (87%) among all the groups, whereas the percentages of correct responses of Mandarin YNLP (78%), Mandarin KTD (76%) and Mandarin YTD (77%) were lower than 80%. This may be due to the fact that the Mandarin OTD had fewer tonal errors (11%) than others, and that Mandarin YNLP, Mandarin KTD and Mandarin YTD had more tonal errors (17%, 17%, and 15%, respectively), semantic errors (2%, 3%, and 4%, respectively), and phonetic errors (all 3%). In addition, Mandarin NLP as well as Mandarin SLP had similar results as Mandarin OTD, and made fewer tonal, semantic, and phonetic errors than the three younger groups.

	Correct response	Tonal error	Semantic error	Phonetic error
Mandarin ASD (*n* = 21)	82%	14%	1%	2%
Mandarin NLP (*n* = 7)	86%	12%	1%	1%
Mandarin SLP (*n* = 9)	84%	13%	1%	2%
Mandarin YNLP (*n* = 5)	78%	17%	2%	3%
Mandarin KTD (*n* = 18)	76%	17%	3%	3%

Mandarin YTD (n = 10)	77%	15%	4%	3%
Mandarin OTD (n = 9)	87%	11%	0%	2%
Overall (n = 58)	81%	14%	2%	2%

Table 3. Reponses to tone comprehension task

All the data was subjected to a mixed ANOVA with one within-subject factor, Tone (Tone 1, Tone 2, Tone 3, Tone 4), and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, Mandarin OTD) with the dependent variable, the percentage of correct responses. Group ($F(5, 52) = 2.372$, $p = .052$, $\eta^2 = .186$) had a marginal effect. (The accepted significance level is $p < .050$.) However, Bonferroni post-hoc tests did not indicate any significant difference between the six groups. On the other hand, Tone ($F(3, 156) = 24.691$, $p < .001$, $\eta^2 = .322$) did have a major effect. Bonferroni post-hoc tests showed that the percentage of correct responses in Tone 2 was significantly lower than in Tone 1 ($p < .001$), Tone 3 ($p = .037$), and Tone 4 ($p < .001$), and that the percentage of correct responses in Tone 3 was significantly lower than in Tone 1 ($p = .002$) and Tone 4 ($p < .001$). There was no statistically significant interaction between the within-subject factor Tone and the one between-subject factor Group ($F(15, 156) = 1.131$, $p = .333$, $\eta^2 = .098$).

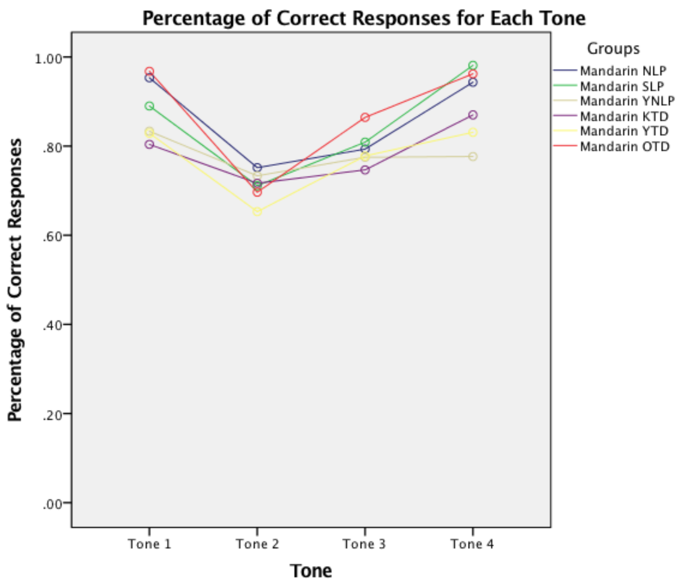

Figure 5. Percentage of correct responses for each tone for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, and Mandarin OTD

All the data was subjected to a mixed ANOVA with one within-subject factor, Tone (Tone 1, Tone 2, Tone 3, Tone 4) and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, Mandarin OTD) with the dependent variable, percentage of tonal errors. Group ($F(5, 52) = 1.600$ $p = .177$, $\eta^2 = .133$) had no major effect, but Tone ($F(3, 156) = 42.967$, $p < .001$, $\eta^2 = .452$) did have a major effect. Bonferroni post-hoc tests showed that the percentage of tonal errors in Tone 2 was significantly higher than in Tone 1 ($p < .001$) and Tone 4 ($p < .001$), and was marginally higher than in Tone 3 ($p = .055$). In addition, the percentage of tonal errors in Tone 3 was significantly higher than in Tone 1 ($p < .001$) and Tone 4 ($p < .001$). There was no statistically significant interaction between the within-subject factor Tone and the one between-subject factor Group ($F(15, 156) = 1.206$, $p = .273$, $\eta^2 = .104$).

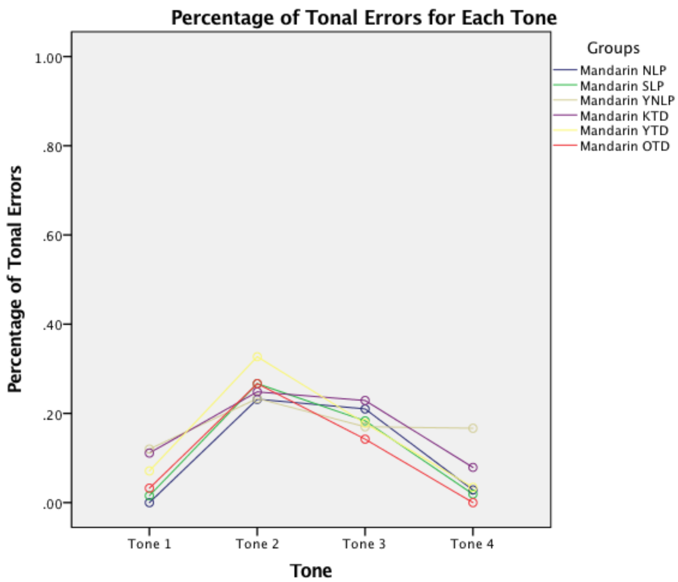

Figure 6 Percentage of tonal errors for each tone for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, and Mandarin OTD

All the data was subjected to a mixed ANOVA with one within-subject factor, Tone (Tone 1, Tone 2, Tone 3, Tone 4) and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, Mandarin OTD) with the dependent variable, percentage of semantic errors. Group ($F(5, 52) = 1.710$ $p =. 149$, $\eta^2 = .141$)

had no major effect. On the other hand, there was a main effect of Tone F(3, 156) = 3.116, p = .028, η^2 = .057. Bonferroni post-hoc tests indicated that the percentage of semantic errors in Tone 2 was significantly lower than in Tone 4 (p = .042), and was marginally lower than in Tone 3 (p = .072). There was no statistically significant interaction between the within-subject factor Tone and the between-subject factor Group (F(15, 156) = 1.260, p = .234, η^2 = .108).

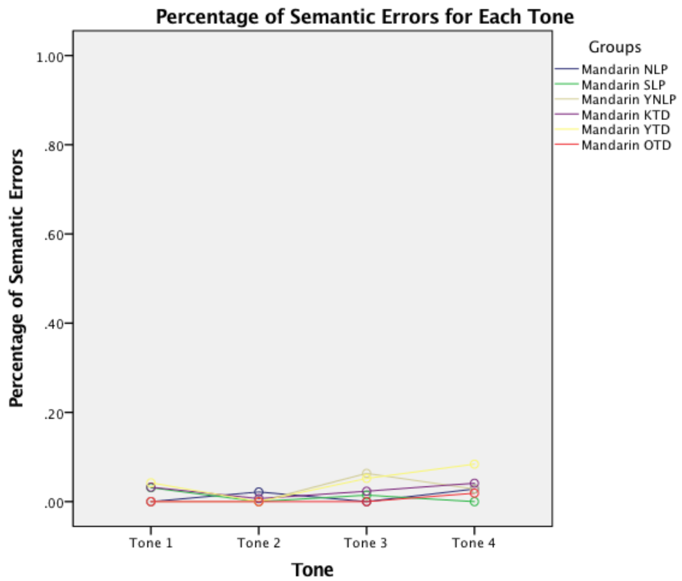

Figure 7. Percentage of semantic errors for each tone for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, and Mandarin OTD

All the data was subjected to a mixed ANOVA with one within-subject factor, Tone (Tone 1, Tone 2, Tone 3, Tone 4) and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, Mandarin OTD) with the dependent variable, percentage of phonetic errors. Group (F(5, 52) = .703, p =.624, η^2 = .063) had no major effect, but Tone (F(3, 156) = 6.054, p = .001, η^2 = .104) did have a major effect. Bonferroni post-hoc tests for Tone showed that the percentage of phonetic errors in Tone 3 was significantly lower than in Tone 1 (p = .001) and Tone 2 (p = .002), and marginally lower than in Tone 4 (p = .098). There was no statistically significant interaction between the within-subject factor Tone and the between-subject factor Group (F(15, 156) = .886, p = .581, η^2 = .078).

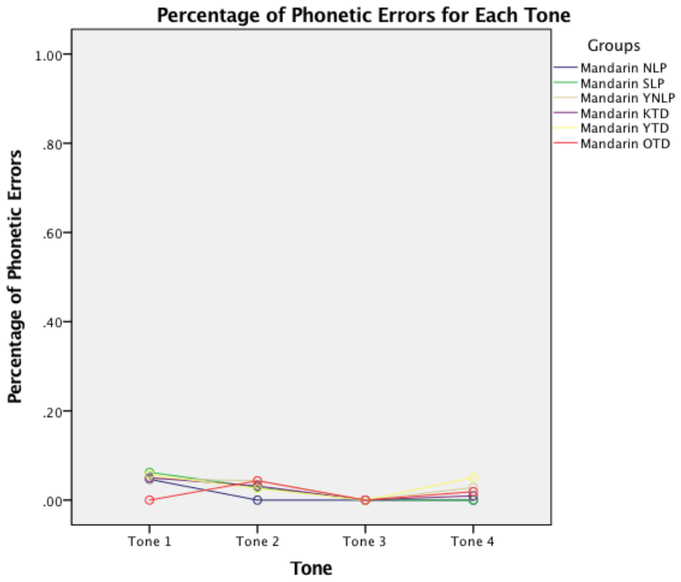

Figure 8. Percentage of phonetic errors for each tone for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, and Mandarin OTD

4.2.2 Tonal error patterns

The following tables indicate the results for each target tone across groups, as well as for each group respectively.

		Correct response	Tonal error	Semantic error	Phonetic error
Mandarin ASD	Tone 1	89%	4%	1%	5%
	Tone 2	73%	25%	1%	2%
	Tone 3	79%	19%	2%	0%
	Tone 4	91%	6%	2%	1%
Mandarin NLP	Tone 1	95%	0%	0%	5%
	Tone 2	75%	23%	2%	0%
	Tone 3	79%	21%	0%	0%
	Tone 4	94%	3%	3%	0%
Mandarin SLP	Tone 1	89%	2%	3%	6%
	Tone 2	71%	27%	0%	3%
	Tone 3	81%	18%	1%	0%
	Tone 4	98%	2%	0%	0%
Mandarin YNLP	Tone 1	83%	12%	0%	5%
	Tone 2	73%	23%	0%	4%
	Tone 3	78%	17%	6%	0%
	Tone 4	78%	17%	3%	3%
Mandarin KTD	Tone 1	80%	11%	3%	5%
	Tone 2	72%	25%	1%	3%
	Tone 3	75%	23%	2%	0%
	Tone 4	87%	8%	4%	1%
Mandarin YTD	Tone 1	83%	7%	4%	6%
	Tone 2	65%	33%	0%	3%
	Tone 3	78%	18%	5%	0%

	Tone 4	83%	3%	8%	5%
Mandarin OTD	Tone 1	97%	3%	0%	0%
	Tone 2	70%	27%	0%	4%
	Tone 3	86%	14%	0%	0%
	Tone 4	96%	0%	2%	2%
Overall	Tone 1	87%	7%	2%	4%
	Tone 2	71%	26%	0%	3%
	Tone 3	79%	19%	3%	0%
	Tone 4	89%	5%	4%	2%

Table 4. Results for each target tone

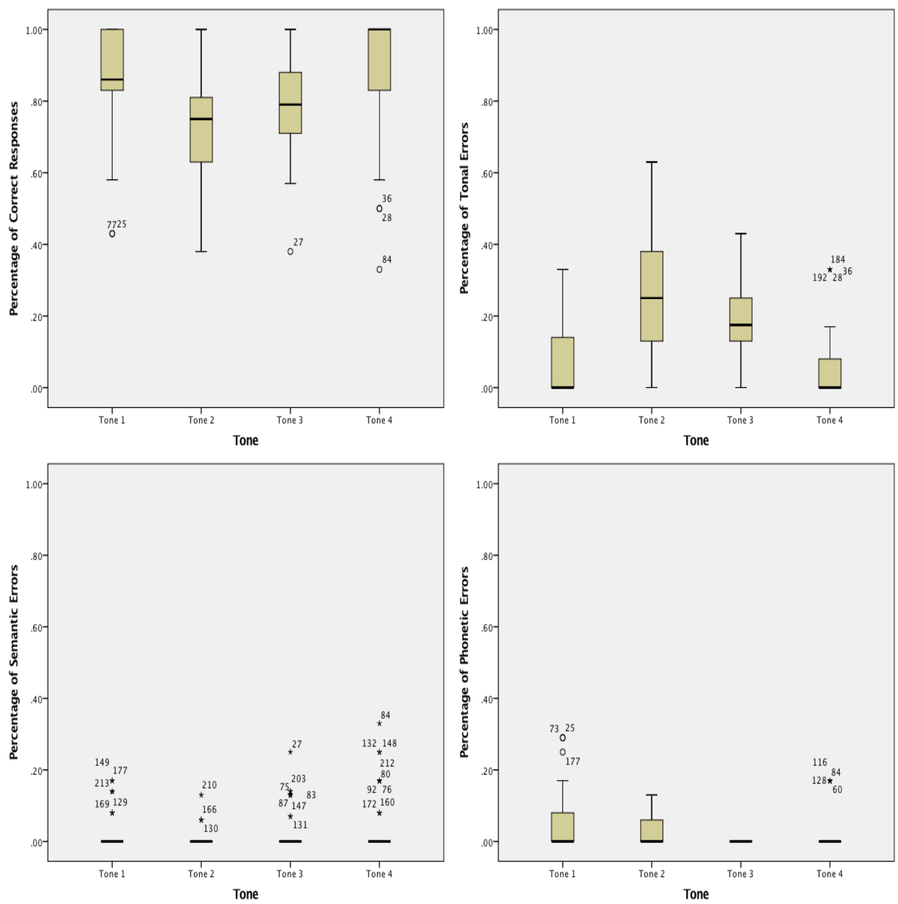

Figure 9. Percentage of correct responses, tonal errors, semantic errors, and phonetic errors across six groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin KTD, Mandarin YTD, and Mandarin OTD)

Each trial contained a minimal pair, where the target tone and its tonal distractor shared the same speech segment and only differed in tones. The percentage of tonal errors participants made when encountering different tonal distractors is illustrated in the tables below. In the tables, we considered the total number of items that had a particular target-distractor tone pairing. For example, if the target is Tone 1 and the distractor is Tone 2, participants made tonal errors only 3% of the time, meaning that from the total number of trials with Tone 1 target and Tone 2 distractor, only 3% of the trials participants misidentified the lexical tones.

When the target was Tone 1, the participants made just a few mistakes, whether the distractors were Tone 2 (3%), Tone 3 (4%) or Tone 4 (6%). This also happened in Tone 4, where the error rates for the three kinds of distractors (Tone 1, Tone 2, and Tone 3) were all below 5%. Since the pitch contour of Tone 4 is falling in Mandarin, it appears that the all participants were able to identify it relatively easily and correctly, especially when the distractor had the opposite pitch contour, (i.e. the rising pitch contour of Tone 2 in Mandarin), since no errors were found. On the other hand, they seemed to have difficulty in perceiving Tone 2, which they identified as Tone 1 with a 40% error rate, the highest in the table, and they mislabelled Tone 2 as Tone 3 with a 31% error rate, the second highest in the table. In addition, the participants found it difficult to perceive Tone 3, mislabelling it as Tone 2 with a 26% error rate and they perceived Tone 3 as Tone 4 with a 16% error rate.

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	3%	4%	6%
Target T2	40%	-	31%	6%
Target T3	3%	26%	-	16%
Target T4	5%	0%	5%	-

Table 5. Confusion matrix for tonal error type across groups

The ASD group only made a few errors when the target tone was Tone 1 or Tone 4; however, they found it difficult to perceive Tone 2 and Tone 3. When the target tone was Tone 2, they perceived Tone 2 as Tone 1 with a 43% error rate, the highest in the table, and they labelled Tone 2 as Tone 3 with a 25% error rate. Moreover, the participants identified Tone 3 as Tone 2 with a 32% error rate, the second highest in the table, and they categorised Tone 3 as Tone 4 with a 12% error rate.

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	2%	2%	5%
Target T2	43%	-	25%	5%
Target T3	2%	32%	-	12%

Target T4	7%	0%	7%	-

Table 6. Confusion matrix for tonal error type for ASD

Again, the ASD group was divided into Mandarin NLP, Mandarin SLP, and Mandarin YNLP. Mandarin NLP did not make any error when the target tone was Tone 1. However, they identified Tone 2 as Tone 1 with a 42% error rate, the highest in the table, and they categorised Tone 2 as Tone 3 with a 25% error rate. It is important to point out that the 42% error rate in misidentifying Tone 2 as Tone 1 was higher than all the groups, including Mandarin YNLP (33%), Mandarin YTD (35%), Mandarin OTD (39%), and Mandarin KTD (14%), except Mandarin SLP (50%). Also, the 25% error rate in misidentifying Tone 2 as Tone 3 was lower than all the TD groups, including Mandarin YTD (40%), Mandarin OTD (33%), and Mandarin KTD (35%). What is more, Mandarin NLP perceived Tone 3 as Tone 2 with a 29% error rate, the second highest in the table. It is worth noting that Mandarin NLP labelled Tone 4 as Tone 1 with an 8% error rate, which was lower than Mandarin YNLP (17%), but higher than all other four groups, including Mandarin SLP (0%), Mandarin YTD (5%), Mandarin OTD (0%), and Mandarin KTD (4%).

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	0%	0%	0%
Target T2	42%	-	25%	0%
Target T3	0%	29%	-	17%
Target T4	8%	0%	0%	-

Table 7. Confusion matrix for tonal error type for Mandarin NLP

As for Mandarin SLP, it is interesting to note that the results were surprisingly neat and clear; in fact, they only mislabelled Tone 1 as Tone 4, but not as Tone 2 or Tone 3, with a 4% error rate. Besides, they only identified Tone 4 as Tone 3, but not as Tone 1 or Tone 2, with a 6% error rate. Furthermore, Table 8 indicates that Mandarin SLP found it difficult to perceive Tone 2 when the distractor was Tone 1. The 50% error rate was the highest across the tables. At times they also identified Tone 2 as Tone 3 with a 25% error rate. Lastly, they labelled Tone 3 as Tone 2 with a 33% error rate, the second highest in the table, and they categorised Tone 3 as

Tone 4 with a 6% error rate. These results may illustrate that Mandarin SLP participants have their own unique and peculiar system in perceiving the various 4 tones in Mandarin; in other words, when Mandarin SLP encountered a stimulus with Tone 2, they identified it as Tone 1 half of the time. When they heard a sound with Tone 3, they may have perceived it as Tone 2 around one third of the time. When they were presented with a stimulus with Tone 4, it is possible that they may have identified it as Tone 3 from time to time, and so they mislabelled Tone 1 as Tone 4. Given that Tone 1 and Tone 4 are relatively easy to acquire and perceive, even for young children, it is not surprising that the error rates were just around 5%. These altogether demonstrated a chain of sound changes, just as those that may be found in diachronic linguistics.

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	0%	0%	4%
Target T2	50%	-	25%	6%
Target T3	0%	33%	-	6%
Target T4	0%	0%	6%	-

Table 8. Confusion matrix for tonal error type for Mandarin SLP

As for Mandarin YNLP, they misidentified Tone 2 as Tone 3 with a 33% error rate, and they also misperceived Tone 3 as Tone 2 with a 33% error rate. While Mandarin YNLP miscategorised Tone 1 as Tone 2 with a 8% error rate, most of the groups seldom made such errors: Mandarin NLP (0%), Mandarin SLP (0%), Mandarin YTD (0%), Mandarin OTD (6%), and Mandarin KTD (0%). In addition, the 17% error rate in misidentifying Tone 4 as Tone 1 was much higher than all the other five groups: Mandarin NLP (8%), Mandarin SLP (0%), Mandarin YTD (5%), Mandarin OTD (0%), and Mandarin KTD (4%). Lastly, it is important to point out that Mandarin YNLP made errors in every condition except that they could perfectly distinguish the target Tone 4 (falling contour) from the distractor Tone 2 (rising contour).

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	8%	8%	11%
Target T2	33%	-	25%	8%
Target T3	8%	33%	-	17%

Target T4	17%	0%	17%	-

Table 9. Confusion matrix for tonal error type for Mandarin YNLP

The Mandarin YTD just made a few errors when the target tone was Tone 1 or Tone 4. They misidentified Tone 1 as Tone 3 and Tone 4, both with a 10% error rate, and they misperceived Tone 4 as Tone 1 and Tone 3 with a 5% error rate. On the other hand, they made many more errors when the target tone was Tone 2 or Tone 3. The participants perceived Tone 2 as Tone 1 with a 35% error rate, the second highest in the table, and they labelled Tone 2 as Tone 3 with a 40% error rate, which was not only the highest in the table, but also the highest across all the groups. (Mandarin NLP: 25%, Mandarin SLP: 25%, Mandarin YNLP: 25%, Mandarin OTD: 33%, and Mandarin KTD: 35%). Furthermore, the participants identified Tone 3 as Tone 2 with a 23% error rate, the second highest in the table, and they categorised Tone 3 as Tone 4 with a 20% error rate.

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	0%	10%	10%
Target T2	35%	-	40%	15%
Target T3	5%	23%	-	20%
Target T4	5%	0%	5%	-

Table 10. Confusion matrix for tonal error type for Mandarin YTD

The Mandarin OTD just made a few errors when the target tone was Tone 1. Moreover, they had no difficulty in perceiving Tone 4 and made no errors when the target tone was Tone 4. Nevertheless, they identified Tone 2 as Tone 1 with a 39% error rate, which was higher than that of Mandarin YNLP (33%), Mandarin YTD (35%), and Mandarin KTD (14%). In addition, they labelled Tone 2 as Tone 3 with a 33% error rate, which was higher than all the ASD groups: Mandarin NLP (25%), Mandarin SLP (25%), as well as of Mandarin YNLP (25%). Like Mandarin NLP and Mandarin SLP, the Mandarin OTD made no errors when the target tone was Tone 3 and the distractor was Tone 1. Besides, Mandarin OTD categorised Tone 3 as Tone 2 with a 17% error rate, which was lower than the other five groups, including Mandarin NLP (29%), Mandarin SLP (33%), Mandarin YNLP (33%), Mandarin YTD (23%), and Mandarin KTD (29%). Finally, Mandarin OTD perceived Tone 3 as Tone 4 with a 22% error rate, which was higher than Mandarin NLP (17%), Mandarin YNLP (17%), Mandarin YTD (20%), and Mandarin KTD (27%).

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	6%	0%	4%
Target T2	39%	-	33%	0%
Target T3	0%	17%	-	22%
Target T4	0%	0%	0%	-

Table 11. Confusion matrix for tonal error type for Mandarin OTD

Last, but not least, when the target was Tone 1, the Mandarin KTD sometimes mislabelled it as Tone 4 (22%), and yet they made no mistake with the distractor of Tone 2 and Tone 3. On the other hand, the Mandarin KTD identified Tone 2 as Tone 3 with a 35% error rate, and perceived Tone 3 as Tone 2 with a 29% error rate. In addition, the 14% error rate in misidentifying Tone 2 as Tone 1 was lower than all the other five groups, including Mandarin NLP (42%), Mandarin SLP (50%), Mandarin YNLP (33%), Mandarin YTD (35%), as well as the Mandarin OTD (39%). Finally, it is worth noting that Kindergarten labelled Tone 4 as Tone 3 with a 19% error rate, which was much higher than the other four groups, including Mandarin NLP (0%), Mandarin SLP (6%), Mandarin YNLP (17%), Mandarin YTD (5%), and Mandarin OTD (0%).

	Distractor T1	Distractor T2	Distractor T3	Distractor T4
Target T1	-	0%	0%	22%
Target T2	14%	-	35%	15%
Target T3	11%	29%	-	27%
Target T4	4%	3%	19%	-

Table 12. Confusion matrix for tonal error type for Mandarin KTD

4.2.3 Semantic errors

Unlike the numerous tonal errors in 1682 trials, there were just 40 semantic errors (2%), which were made in three trials in which the participants made a relatively high number of semantic errors. These were /hu3/ “tiger” versus /shi1/ “lion” in the first trial (contributing 20% of semantic errors), /ya1/ “duck” versus /e2/ “swan” in the ninth trial (17% of semantic errors), and

/jao3/ “to bite” versus /tian3/ “to lick” in the twenty-eighth trial (15% of semantic errors). 18 of these semantic errors were made by 18 Mandarin KTD, 12 by 10 Mandarin YTD, 4 by 5 Mandarin YNLP, 3 by 9 Mandarin SLP participants, 2 by 7 Mandarin NLP participants, and only 1 by 9 Mandarin OTD.

Given that most of the semantic errors were made by the Mandarin KTD, the 7 year-olds in Mandarin YTD, and a 5 year-old and an 8 year-old in Mandarin YNLP, it was suspected that age may have been correlated with the number of semantic errors and a Pearson product-moment correlation test was conducted in SPSS to determine their relationship. The results showed that the age of the participants and the number of semantic errors were moderately negatively correlated (r = -.486, p < .001). This indicated that the younger the participant, the more likely he or she would be to make semantic errors.

4.2.4 Phonetic errors

There were just 41 phonetic errors in the 1682 trials (2%), all of which were made in two trials in which the participants made a relatively high number of phonetic errors. These were /jie1/ “street” versus /jian1/ “sharp” in the eighteenth trial (contributing 15% of phonetic errors) and /maio2/ “sprout” versus /yao2/ “to shake” in the twenty-ninth trial (15%). Intriguingly, these frequently occurred phonetic errors were all related to the nasal sounds. That is, when the target word and the phonetic distractor differed in the existence of a nasal sound (and a slight difference in the vowel in the eighteen trial), the participants tended to make more phonetic errors since the nasal sounds were less salient and harder to perceive and process.

Besides, 2 of the 41 phonetic errors were made by 7 Mandarin NLP participants, 6 were made by 9 Mandarin SLP, 5 were made by 5 Mandarin YNLP, 9 were made by 10 Mandarin YTD, 4 were made by 9 Mandarin OTD, and 15 were made by 18 Mandarin KTD. Since every group had 1-3% of phonetic errors, it was deemed to be more enlightening to explore the relationship between age and phonetic errors. The results of the Pearson product-moment correlation demonstrated that there was a marginal weak negative correlation between the participants' ages and the number of phonetic errors (r = -.255, p = .077). These results altogether indicated that unlike the number of semantic errors, the number of phonetic errors was not closely related to the age. Every group in the current study had about 2-3% of phonetic errors and did not demonstrate the great development along the age from six to nineteen years old.

4.2.5 Correlations with age and language ability

Since the results of the Pearson product-moment correlation test indicated a moderate correlation between the age of the participants and the number of semantic errors, whereas there was only a marginal weak correlation between the age of the participants and the number of phonetic errors, it was deemed to be interesting to explore the relationship between age and all kinds of response errors in the data. In order to explore the progress of the typical development along the age, the three TD groups (Mandarin KTD, Mandarin YTD, and Mandarin OTD) were firstly examined based on the suggestion by Evans (1996). If the absolute value of r is between .00 and .19, then it indicates a very weak correlation. If the absolute value of r is between .20 and .39, then it indicates a weak correlation. If the absolute value of r is between .40 and .59, then it indicates a moderate correlation. If the absolute value of r is between .60 and .79, then it indicates a strong correlation. If the absolute value of r is between .80 and 1.00, then it indicates a very strong correlation.

Table 13 shows that the age of the TD participants is moderately positively correlated with the number of correct responses (r = .493, p = .002), manifesting the overall development along the age from six to nineteen years old. In addition, there was a significant weak negative correlation between the age of the participants and the number of tonal errors (r = -.371, p = .024), illustrating that the younger the participant was, it was likely that he or she would make more tonal errors. Nonetheless, it is worth noticing that although the age was slightly correlated with the number of tonal errors, the comprehension of lexical tones may only subtly improve along the age. The age of the participants and the number of semantic errors were moderately negatively correlated (r = -.401, p = .014), indicating that the lexical meanings would become more stable and easier to retrieve along the age. On the contrary, there was no significant correlation between the age of the participants and the number of phonetic errors (r = -.233, p = .185). Just as discussed above, the percentage of phonetic errors was similar in every group in the current study, and thus there was no significant development of phonetic perception along the age. Lastly, it makes sense that the number of correct responses was significantly correlated with the number of tonal errors, the number of semantic errors, and the number of phonetic errors, respectively (all ps < .003).

		Age	Correct response	Tonal error	Semantic error	Phonetic error
Age	Pearson Correlation	1	.493**	-.371*	-.401*	-.223
	Sig. (2-tailed)	-	.002	.024	.014	.185
Correct response	Pearson Correlation	.493**	1	-.837**	-.632**	-.469**
	Sig. (2-tailed)	.002	-	.000	.000	.003
Tonal error	Pearson Correlation	-.371*	-.837**	1	.243	.061
	Sig. (2-tailed)	.024	.000	-	.147	.720
Semantic error	Pearson Correlation	-.401*	-.632**	.243	1	.246
	Sig. (2-tailed)	.014	.000	.147	-	.142
Phonetic error	Pearson Correlation	-.223	-.469**	.061	.246	1
	Sig. (2-tailed)	.185	.003	.720	.142	-

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Table 13. Correlation between the age of the participants, the number of correct responses, the number of tonal errors, the number of semantic errors, and the number of phonetic errors in three TD groups (Mandarin KTD, Mandarin YTD, and Mandarin OTD)

As for the ASD participants, the PPVT score was also taken into consideration in order to explore the correlation between the linguistic abilities and other variables. Table 14 shows that the age of the participants was marginally positively correlated with the PPVT score (r = .561, p = .058), pointing out that while the linguistic abilities in ASD children would develop as they grow up, the great diversity within the ASD participants may blur the age effects on the linguistic abilities. In addition, the age of the participants is strongly positively correlated with the number of correct responses (r = .602, p = .038), manifesting the overall development along the age even in the ASD children. On the other hand, there was no significant correlation between the age of the participants and the number of tonal errors (r = -.493, p = .104), once again illustrating the great diversity within the ASD children may hugely affect their linguistic abilities such as the comprehension of lexical tones. The age of the participants and the number of semantic errors were strongly negatively correlated (r = -.652, p = .021), indicating that unlike the comprehension of lexical tones, the processing of the lexical meanings could be greatly improved along the age in the ASD children. There was a marginal moderate negative correlation between the age of the participants and the number of phonetic errors (r = -.510, p = .091), suggesting that unlike the TD groups, ASD children may make progress in the phonetic perception along the age. This is possibly due to the extra phonetic and pronunciation training they received in the after-school programme. Finally, Table 14 also indicates that the PPVT score of the ASD participants was significantly correlated with the number of correct responses, the number of tonal errors, the number of semantic errors, and the number of phonetic errors (all ps < .035), suggesting that the linguistic abilities may be more reliable in predicting the performance in the tone comprehension task. The better linguistic abilities the participant with ASD had, it was likely that he or she could perform better in the tone comprehension task.

		Age	PPVT score	Correct response	Tonal error	Semantic error	Phonetic error
Age	Pearson Correlation	1	.561	.602*	-.493	-.652*	-.510
	Sig. (2-tailed)	-	.058	.038	.104	.021	.091
PPVT score	Pearson Correlation	.561	1	.712**	-.628*	-.640*	-.612*
	Sig. (2-	.058	-	.009	.029	.025	.035

	tailed)						
Correct response	Pearson Correlation	.602*	.712**	1	-.952**	-.741**	-.084**
	Sig. (2-tailed)	.038	.009	-	.000	.006	.002
Tonal error	Pearson Correlation	-.493	-.628*	-.952**	1	.527	.653*
	Sig. (2-tailed)	.104	.029	.000	-	.078	.021
Semantic error	Pearson Correlation	-.652*	-.640*	-.741**	.527	1	.659*
	Sig. (2-tailed)	.021	.025	.006	.078	-	.020
Phonetic error	Pearson Correlation	-.510	-.612*	-.084**	.653*	.659*	1
	Sig. (2-tailed)	.091	.035	.002	.021	.020	-

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Table 14. Correlation between the age of the participants, the PPVT score, the number of correct responses, the number of tonal errors, the number of semantic errors, and the number of phonetic errors in ASD participants

5 Discussion

Q1: Can typically developing children and ASD children comprehend Tones 1, 2, 3, and 4 in Mandarin?

The present study examined the perception of typically-developing (TD) children, as well as ASD children, of Mandarin tones in monosyllabic words by the accuracy of identifying the corresponding pictures. The TD children and ASD children were both able to perceive the four Mandarin tones at word level and differentiate lexical items with moderate accuracy based on the pitch contours, with around 15% of tonal errors in the task.

Although the percentage of correct responses or tonal errors of the three Mandarin TD groups (Mandarin KTD, Mandarin YTD, Mandarin OTD) were not significantly different (Figure 5 and Figure 6), there was a developmental trend that the percentage of correct responses became slightly higher and higher from Mandarin KTD, Mandarin YTD, to Mandarin OTD (Table 4). On the other hand, the percentage of tonal errors became slightly lower and lower from Mandarin KTD, Mandarin YTD, to Mandarin OTD. The correlation test indicated that the age of the TD participants was moderately positively correlated with the number of correct responses, illustrating an overall development among six to nineteen year-olds (Table 13). In addition, there was a significantly weak negative correlation between the age of the TD participants and the number of tonal errors, illustrating that the younger the participants, the more likely they were to make more tonal errors. Nonetheless, it is worth noting that, although the age was slightly correlated with the number of tonal errors, the comprehension of lexical tones may only have subtly improved with age.

In addition, the percentage of correct responses and tonal errors for Mandarin NLP and Mandarin SLP were both somewhere between the performance of Mandarin YTD and OTD, which would be appropriate for their ages. The correlation tests indicated that the age of the ASD participants was strongly positively correlated with the number of correct responses, illustrating an overall development with age, even in the ASD children (Table 14). On the other hand, there was no significant correlation between the age of the participants and the number of tonal errors, once again illustrating that the great diversity within ASD children may hugely affect their linguistic abilities, such as the comprehension of lexical tones.

Finally, the correlation tests indicated that the linguistic abilities of the ASD participants was significantly correlated with the number of correct responses, and the number of tonal errors (all *p*s $< .035$), suggesting that linguistic abilities may be more reliable in predicting the performance in the tone comprehension task. The better the linguistic abilities of participants with ASD, the more likely they were to perform better in the tone comprehension task.

Q1a: Do typically developing children comprehend Tones 1 and Tone 4 better than Tones 2 and Tone 3?

Just as indicated in the literature (Li & Thompson, 1997; Shiu, 1990; Zhu & Dodd, 2006), Tone 1 and Tone 4 are acquired before Tone 2 and Tone 3, and it is easier to perceive Tone 1 and Tone 4 than Tone 2 and Tone 3. The

TD children in the present study perceived Tone 1 and Tone 4 with the highest accuracy, Tone 3 with lower accuracy, and Tone 2 with the lowest accuracy (Figure 6). While Wong (2005) found that 3-year-olds had no difficulty in perceiving Tone 2, just as Tone 1 and Tone 4, Tone 2 and Tone 3 were significantly harder to accurately perceive in the present study than Tone 1 and Tone 4, thus leading to a lower percentage of correct responses, as well as a higher percentage of tonal errors in Tone 2 and Tone 3. These results were consistent with the claim in the literature that it is much easier to perceive Tone 1 and Tone 4 than Tone 2 and Tone 3.

Q1b: Does the pattern of acquisition found in typically developing children extend to children with ASD (including those with or without language problems (Mandarin SLP, Mandarin NLP)), or do children with ASD display a deviant (not just delayed) developmental pattern?

Just as the TD children, the ASD children comprehended Tone 1 and Tone 4 better than Tone 2 and Tone 3. However, after separating the ASD group according to their linguistic abilities, Mandarin NLP were found to have the lowest percentage of correct responses and the highest percentage (73%) of tonal errors (21%) in Tone 3, whereas Mandarin SLP had the lowest percentage of correct responses and the highest percentage (69%) of tonal errors (30%) in Tone 2, although this difference was not statistically significant. Since the TD children and the ASD children both found Tone 2 and Tone 3 much harder to perceive and process, it made sense to utilise Tones 1 and 4 as the tone pair for the subsequent tone discrimination and naming tests.

Q1c: Do children with ASD (including children with or without language problems (Mandarin SLP, Mandarin NLP)) display the same rate of acquisition of tones in their comprehension as typically developing children?

The mean chronological ages (CA) of Mandarin NLP and SLP was somewhere between the mean CA of Mandarin YTD and that of Mandarin OTD. Since Mandarin NLP and SLP did not perceive the lexical tones significantly differently from Mandarin YTD or OTD, they displayed a similar rate of acquisition of tones in their comprehension without a significant delay in development.

Q2a: When tonal errors occur in comprehension, what is the pattern of those errors for typically developing children?

The TD children were able to identify Tone 1 and Tone 4 easily and correctly; thus, they seldom made tonal errors when the target was Tone 1 or Tone 4 (Table 10). This could be attributed to the acoustical dissimilarity of Tone 1 and Tone 4 from other tones. On the other hand, the TD children seemed to have difficulty in perceiving Tone 2 and Tone 3. They tended to misidentify Tone 2 as Tone 3 and Tone 3 as Tone 2, which supports the claim that discrimination among Tone 2 and Tone 3 poses the greatest challenge due to the phonetically-physical similarity of these two lexical tones (Gandour, 1978; Huang, 2004; Huang, 2007; Hume & Johnson, 2003; Li & Thompson, 1977). Because Tone 2 and Tone 3 both start at about the same pitch level and have a dynamic F0 shape, both tones exhibit a similar F0 contour, which makes perception difficult. It is worth noting that there was a directional asymmetry between Tone 2 and Tone 3, namely, that when the target was Tone 2 and the distractor was Tone 3, perception may have been more difficult than when the target was Tone 3 and the distractor was Tone 2.

Q2b: Does the pattern of tonal errors found in typically developing children extend to those with ASD (including children with or without language problems (ASD-SLP, ASD-NLP)), or do children with ASD display a different pattern?

Just as the TD children, the Mandarin SLP and NLP seldom made tonal errors when the target was Tone 1 or Tone 4 (Table 6, 7, and 8). They also found it difficult to perceive Tone 2 and Tone 3, and tended to confuse these two lexical tones. In addition, the ASD children sometimes mislabelled Tone 2 as Tone 1, and misidentified Tone 3 as Tone 4. It is important to point out that there was also a directional asymmetry between Tone 2 and

Tone 3 in the ASD NLP and SLP, but in the opposite direction. Therefore, when the target was Tone 3 and the distractor was Tone 2, it actually caused more tonal errors than when the target was Tone 2 and the distractor was Tone 3. Thus, Mandarin NLP and SLP both made fewer tonal errors when the target was Tone 2 and the distractor was Tone 3 than the TD children. However, when the target was Tone 3 and the distractor was Tone 2, Mandarin NLP made more tonal errors than Mandarin YTD and OTD, and Mandarin SLP made even more tonal errors than all the Mandarin TD groups, including Mandarin KTD. This may suggest a slight tendency or strategy by the ASD children. It may be possible that they tended to perceive Tone 3 as Tone 2 in auditory processing, but not the other way round, or they may have simply developed a strategy to prefer Tone 2 when they were presented with words that had ambiguous pitch contours of Tone 2 or Tone 3.

Q3a: Do typically developing children make semantic or phonetic errors in their comprehension of free-standing lexical items?

In addition to the numerous tonal errors, the TD children made just a few semantic and phonetic errors (around 3% respectively) in their comprehension of free-standing lexical items (Figure 7 and Figure 8). This suggests that, compared to phonemes or semantic meanings, lexical tones made it much more difficult and easier to misunderstand when perceiving and processing lexical items. Mandarin KTD and YTD made significantly more semantic errors than Mandarin SLP and OTD. In addition, as indicated by the correlation test, the age of the participants and the number of semantic errors were moderately negatively correlated, suggesting that the younger the participants, the more likely they were to make semantic errors.

On the other hand, every TD group made similarly few phonetic errors, and when they did occur, they were usually related to nasal sounds, which were less salient and harder to perceive and process. Unlike semantic errors, there was no significant correlation between the age of the participants and the number of phonetic errors, and there was no prominent development shown with age in six to eighteen year-olds.

Q3b: Is the rate of ASD children's semantic and phonetic errors similar to that of typically developing children?

The ASD participants did not perform significantly differently from the Mandarin OTD in terms of semantic and phonetic errors. This suggests that children with autism could actually perceive lexical items properly and were not deficient in semantic or phonetic discrimination. The correlation tests showed that the age of the ASD participants and the number of semantic errors was strongly negatively correlated, indicating that unlike the comprehension of lexical tones, the processing of lexical meanings could be greatly improved with age in the ASD children.

Furthermore, there was a marginally moderate negative correlation between the age of the ASD participants and the number of phonetic errors (r = -.510, p = .091), which suggests that, unlike the TD groups, ASD children may make progress in phonetic perception with age. This could possibly be attributed to the extensive phonetic and pronunciation training they received in the after-school programme.

Finally, the correlation tests indicated that the linguistic abilities of the ASD participants significantly correlated with the number of semantic errors and the number of phonetic errors (all ps < .035). This suggests that their linguistic abilities may be more reliable in predicting their performance in the tone comprehension task. The better the linguistic abilities of the participants with ASD, the more likely they were to perform better in the tone comprehension task.

Chapter 3: Psychoacoustic tone discrimination task

1 Introduction

1.1 Aim of current work

Some children with autism spectrum disorder exhibit delayed language development, as well as atypical auditory perception, which are sometimes enhanced. Some recent evidence suggests that there may be a link between these two elements with the atypically perceived speech signal providing sub-optimal input for language learning. Since the acoustic, prosodic properties of speech carry highly relevant lexical and grammatical information and have been increasingly recognised to play a crucial role in typical language acquisition, it is plausible that atypical auditory perception may be the cause of at least some aspects of language problems in children with autism. The current project seeks to investigate this causal link by exploring the auditory perception and language development of high-functioning children with autism by uncovering the specific mechanisms through which the former may impact the latter.

The properties of the mammalian auditory system have been increasingly recognised to contribute to language development in fundamental ways (Gervain & Mehler, 2010; Trainor & Desjardins, 2002; Vouloumanos & Werker, 2007a). The speech signal encoded by the auditory system serves as input for language learning. Importantly, auditory processing transforms this signal by organising it into different representational patterns and these transformations have a direct impact on the input for learning a language (e.g. Endress et al., 2009; Fitch, 2000, Morgan & Demuth, 1996). The precise nature of the relevant processing mechanisms and the ways in which they contribute to language acquisition are only now beginning to be understood.

Four groups will be investigated in this thesis: typically-developing children and children with Autism Spectrum Disorder (ASD), who are native speakers of English and Mandarin. ASD is a neuro-developmental disorder, which is characterised by social and communication impairment, as well as inflexible thought and behaviour (DSM-IV, American Psychiatric Association, 2000); furthermore, people with ASD experience significant perceptual abnormalities (Rogers & Ozonoff, 2005). The language competence of learners with ASD is considerably varied and, more recently, it has been hypothesised that individuals with ASD, or at least those on the lower-functioning end of the spectrum, may also demonstrate atypical pitch and musical perception (reviewed in McCann & Peppé, 2003; Bonnel et al., 2010; O'Connor, 2012; Lepistö et al., 2008; Jones et al., 2009). These studies revealed the enhanced discrimination of simple tones (Jones et al., 2009) and of pitch in linguistic stimuli (Heaton et al., 2008). Therefore, the language delay of individuals with ASD may not only relate to their deficient social cognition, but also their atypical low-level perceptual ability (Reynolds, Newsom, & Lovaas, 1974). This is supported by the apparent absence of atypical pitch and musical perception in ASD individuals with milder language problems (Chevallier, Noveck, Happé, & Wilson, 2009; Jones et al., 2009). However, the hypothesis of a negative association between enhanced perception and specific language skills has not yet been systematically tested (Heaton et al., 2008; Jones et al., 2009).

1.2 Previous work

Ceponiene et al. (2003) explored the pitch and duration perception of nine high-functioning Finnish children with ASD (mean age: 8.9 yrs.) and ten controls (mean age: 8.4 yrs.) with cortical event-related brain potentials (ERPs), which were extracted from the electroencephalogram (EEG) study. They found that four of the children with ASD had low language ability and needed pictures to communicate, while the other five could read and communicate using sentences. They used three kinds of stimuli: simple tones, complex tones, and Finnish vowels /ö/, and each type was presented in a separate block. Each block contained 400 stimuli of the same class.

86% of the stimuli were standard items, 7% were frequency deviants with 10% higher frequency than the standard items, and 7% were duration deviants. The participants were presented with at least three blocks of each type of stimulus, and three types of electrophysiological measurements were recorded: auditory sensory ERPs, the mismatch negativity component (MMN), and the P3a component. The ERPs could reflect sound frequency or intensity, the MMN showed the detection of infrequent frequency deviants from the standard items, and the P3a demonstrated the involuntary attention switch to salient events in the environments.

The results indicated that there was no significant difference between the children with ASD and the typically-developing children in the sensory domain; however, while the P3a was significant in all three types of stimulus in typically-developing children, no P3a was elicited by the changes in vowel in participants with ASD. The authors concluded that the children with ASD could actually perceive the changes in pitch just as their matched controls, but there might be impairments in their attention orientation to speech sound such as vowel changes.

Lepisto also ran a series of electrophysiological studies in order to further explore the auditory perception in children with autism. Lepisto et al. (2005) tested fifteen Finnish children with ASD and fifteen matched controls (mean age 9.4 years) by recording auditory ERPs. Two of the children with ASD were also diagnosed with co-morbid attention-deficit/hyperactivity disorder (ADHD), whereas the thirteen participants with ASD were free of any other diagnoses. There were two kinds of stimulus types: speech as well as non-speech counterparts, which were created as a composition of four sinusoidals. Each block randomly presented 400 stimuli of the same class, 76% of the stimuli were the standard items, 8% of which were the frequency deviants with 10% higher or lower frequency than the standard items, 8% of which were the duration deviants, and 8% of which were the vowel (/a/ versus /o/ for speech condition or the non-speech vowel counterparts for the non-speech condition) deviants. The participants were presented with the blocks while watching silent videos and ignoring the stimuli. Just as the previous study, the three types of electrophysiological measurements, ERPs, MMN, and P3a were recorded.

The results suggested that the MMN responses were enlarged in children with ASD for both speech and non-speech pitch changes, providing positive evidence for the auditory hypersensitivity and enhanced pitch perception in ASD. In addition, the P3a was diminished in the children with ASD for pitch changes in speech stimuli, but not for the pitch changes in non-speech stimuli. These corresponded with the study by Ceponiene et al. (2003) in the way that while children with ASD could sensitively perceive the pitch changes, there were deficits in the involuntary orientation for the speech stimuli.

Lepisto et al. (2006) tested ten Finnish children with AS (i.e. ASD-NLP) (mean age 8.11) and ten matched controls (mean age 8.10 years) by recording auditory ERPs. All of these AS children had normal cognitive and language development, and none of them met the diagnostic criteria for ASD. Although two of them were also diagnosed with ADHD, in both cases AS was considered to be the primary diagnosis. The same stimuli and procedure were adopted from Lepisto et al. (2005). The results suggested that just as the ASD children, AS children had an enhanced MMN for both speech and non-speech pitch changes, providing positive evidence for the auditory hypersensitivity and enhanced pitch perception in AS. In addition, the P3a was diminished in the AS children for pitch changes in speech stimuli, but not for the pitch changes in non-speech stimuli. In spite of the fact that the AS children in this study and the ASD children in Lepisto et al. (2005) differed remarkably in the language development, these two clinical groups had similar results and both showed the enhanced pitch perception as well as impairments in the involuntary orientation.

In order to test whether this pattern of auditory processing could also be observed in adulthood, Lepisto et al. (2007) tested nine Finnish adults with

AS (mean age: 27) and nine matched controls (mean age: 30 years) by recording auditory ERPs. All of these AS adult had normal cognitive and language development, and none of them met the diagnostic criteria for ASD. The same paradigm was adopted from Lepisto et al. (2005). In addition, there was a behavioural sound-identification AX test utilizing the same stimuli as in the ERP study. 50% of the pairs were the identical sounds, whereas 50% of the pairs were the different sounds. The task for the participants was to determine whether the two sounds were the same or different by pressing corresponding buttons. The results suggested that just like the ASD children as well as the AS children, AS adults had enhanced MMN amplitudes for both speech and non-speech pitch changes, providing positive evidence for the auditory hypersensitivity and enhanced pitch perception even in AS adults. In addition, the P3a was diminished in the AS adults for pitch changes in speech stimuli, whereas the P3a was enhanced for pitch changes in non-speech sounds. As for the results of the sound-identification task, there was no significant group difference in hit rates, and yet AS adults had significantly longer reaction times than their matched peers for the non-speech pitch differences. Together with the results in Ceponiene et al. (2003), Lepisto et al. (2005), and Lepisto et al. (2006), the researchers revealed that ASD children, AS children, as well as AS adults all have enhanced cortical pitch perception and atypical involuntary orientation to speech sound changes.

There are two mainstream theories to explain the atypical pattern of attention to perceptual details in speech processing, one is weak central coherence (WCC; Happe, 1999; Happe & Frith, 2006), and the other is enhanced perceptual functioning (EPF; Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). WCC theory (Happe, 1999) originally emphasized on a core deficit in central processing leading to malfunction in extracting global form/meaning. However, after reviewing numerous empirical studies of coherence, Happe and Frith (2006) concluded that people with ASD indeed showed superior performance on tasks requiring detail-focused processing, and yet it was not clear whether the superiority of local processing actually cost the normal global processing. Since individuals with ASD could properly process the global meaning when explicitly required, the researchers speculated that it was possible that there was a processing bias for local over global levels of information for people with ASD. On the other hand, Mottron et al. (2006) proposed an EPF model arguing that perception played a different and superior role in autistic cognition, resulting enhanced visual as well as auditory perception in tasks with lower-order.

In order to explore the local and global auditory perception in ASD, Foxton et al. (2003) tested thirteen ASD participants (mean age: 18.1 years) and fifteen matched controls (mean age: 17.7 years) with same-different decision auditory tests. While all thirteen ASD participants shared similar clinical symptoms and could be described under the umbrella term ASD (Gillberg, 2002), eleven of the ASD participants satisfied DSM-IV (Diagnostic and Statistical Manual of Mental Disorders-Edition IV) criteria for AS since they did not have delay in the use of language for social communication. The stimuli were pairs of five-note pitch sequences. The notes were all pure tones of 250ms duration, and the possible lowest pitch in the sequences was 250 Hz. In Test 1, the "same" sequences were exactly the same, whereas for the "different" sequences, one of the notes (not the first or the last notes) in the second sequence was altered by a magnitude of two "notes" so that the patterns of rises and falls in pitch were violated. In Test 2 with local pitch interference, the second sequence was always transposed up in pitch by half an octave. Therefore, for the "same" sequences, the second sequence was actually consistently higher than the first sequence by half an octave, and yet remained the same patterns in pitch as the first sequence. On the other hand, for the "different" sequences, the second sequence was not only higher than the first sequence, but also had one of the altered notes (not the first or the last notes) to violate the pitch patterns. In Test 3 with local pitch and timing interference, the pitch patterns were either a rise followed by a fall or a fall followed by a rise. In addition, the second sequence was always transposed up in pitch by half an octave. For the "same" sequences, the pair of sequences would have the same pitch pattern but differed in the exact points of the rises and falls. For the "different" sequences, the pair of sequences would have the different contour patterns with various points of the rises and falls. Test 4 was utilized to compare the first test. It was identical to the first test, but only differed in the fact that the note changes would not violate the patterns of rises and falls in pitch. For all these tests, the task for the participants was to determine whether the pair of sequences was the "same" or "different" by pressing corresponding buttons.

The results indicated that Test 3 scores were significantly lower than Test 2 scores, and that Test 2 scores were significantly lower than Test 1 scores for the control group. Nonetheless, despite the fact that the ASD group also had a trend in this direction, the main effect of Task was not significant. The authors suggested that while the typically developing individuals had the interference from the global structure, the interference from the auditory coherent whole was absent or weak in the participants with ASD. In addition, the results in Test 4 did not show any significant group difference and failed to demonstrate the enhanced pitch perception in ASD. This is inconsistent with the Enhanced Perceptual Processing Model found in many other studies discussed above. However, the researchers believed that it might be possible that the hypersensitivity in the pitch perception was only evident in fine-grained pitch differences, but not in their study, which utilized relatively large pitch differences between the five-note pitch sequences. There was another fundamental problem in this study. That is, for all these four same-different decision auditory tests, the examiners did not provide the baseline for the participants to demonstrate what is the "same" and what is the "different". Furthermore, it is important to point out that in the Test 2 and Test 3, the "same" sequences were not exactly the same. There were also differences in the pitch height or the exact time of turning points in pitch patterns. All of these could be confounding for the participants since the "same" sequences were actually "different" from the standard sequences.

Jarvinen-Pasley, Pasley, and Heaton (2008) explored the linguistic content of speech as well as the non-speech perceptual features in twenty-eight children with ASD (mean age: 12.2 years) and twenty-eight matched controls (mean age: 12 years) with a quasi-open-format paradigm. 57% of the clinical children were formally diagnosed with autistic disorder (ASD), whereas 43% of the children were noted as AS without significant delay in language development. The match controls were children with moderate learning difficulties (MLD) so that the chronological age, the verbal mental IQ (tested by British Picture Vocabulary Scale; BPVS; Dunn, Whetton, & Pintilie, 1997), as well as the non-verbal IQ (tested by Raven Standard Progressive Matrices; RSPM; Raven, Court, & Raven, 1992) were all matched with the participants with ASD. Both groups had about 75 for verbal mental IQ and 80 for non-verbal IQ. There were twenty-four sentences as the experimental stimuli. Each of these sentences was paired with one corresponding picture, and yet the sentences just referred to "situations" instead of directly naming the objects in the pictures. These sentences were read in one of four pitch contours: ascending, descending, low-high-low, or high-low-high, and then the visual symbols were created based on the pitch contours of the sentences. Every participant underwent two kinds of trainings, one was the perceptual training, and the other was the linguistic training. In the perception training, the participants were firstly shown a visual display depicting the four possible pitch contours of the sentences, and then they were asked to point out the shape that matched the sentence in the trials. On the other hand, in the linguistic training, the participants were shown with pictures depicting stories, and then they were asked to identify the picture that matched the sentence in the trials. Once the training phase was completed, the participants were informed that they were free to match the sentence with a shape or a picture in the following main test. In each trial of the main test, the participants would hear one sentence and were presented with two pictures and two shapes. For the two pictures, one was the corresponding picture to the sentence, and the other was the distractor picture. For the two shapes, one was the corresponding visual symbol of the pitch contour of the sentence, and the other was the distractor visual symbol of pitch contour. The task for the participants was to choose one item that best matched the sentence.

The results indicated that while both groups preferred to choose the linguistic semantic content (94% of the trials in the control group, and 65% in the clinical group) to the perceptual intonation content, children with ASD produced significant more perceptual interpretation than their matched peers. Moreover, autistic children not only provided equally accurate linguistic responses, but also made significantly more accurate perceptual judgments than the matched counterparts. The authors suggested that these demonstrated the enhanced perceptual processing, weakened linguistic bias, as well as unimpaired linguistic accuracy in participants with ASD, and these findings were actually largely consistent with WCC as well as EPF theories. That is, although WCC theory would predict a dominant featural/surface-biased information processing styles in participants with ASD, study by Snowling and Frith (1986) showed that this tendency might disappear when the participants with ASD were instructed to focus on the semantic level. Moreover, despite the fact that EPF theory would argue for the locally oriented and enhanced perceptual functioning as the preferred processing style for individuals with ASD, higher-level processing abilities actually remained intact.

Bonnel et al. (2010) used a four-interval, two-alternative forced-choice task (4I-2AFC) to examine the pitch discrimination of fifteen people with autism, fourteen with Asperger's syndrome (AS), and fifteen typically-developing individuals. The mean age of these three groups was around twenty-three, and their mean IQs were all about 105. There were two kinds of stimuli: pure tones and complex tones. In each trial, the participants heard four sounds, "AB-AA", and they were required to determine which pair contained a different sound. The standard stimulus had a 500Hz frequency, and the deviant stimuli were adjusted according to a 3-down/1-up adaptive procedure (Levitt, 1971) based on the current performance of the participant in order to calculate the threshold of the discrimination. The results suggested that the participants with ASD performed significantly better than the typically-developing individuals and those with AS in terms of discriminating the pure tones. The authors stated that it was the participants with ASD with linguistic difficulties who displayed an enhanced pitch perception, as had been demonstrated by the study of Mottron et al. (2006). This led to the question of whether the atypical pitch perception of individuals with ASD may be responsible for their language impairment to some degree.

Jones et al. (2009) used the PEST (Parameter Setting by Sequential Estimation, Findlay, 1978) in a psychoacoustic task to investigate the frequency discrimination of seventy-two adolescents with ASD and forty-eight IQ and age-matched controls. The mean age of these two groups was fifteen and a half, and the mean of their IQ was around ninety. Thirty-nine of the seventy-two adolescents with ASD were diagnosed with childhood autism, and thirty-three participants met the ICD-10 criteria for “other ASDs”. (Three of them had “atypical autism”, twenty-eight had “other pervasive developmental disorders”, and two had “pervasive development disorder unspecified) Meanwhile, twenty-six of the matched control participants were typically-developing children and the other twenty-two had special educational needs (non-ASD).

In this task, the participants were presented with two dinosaurs on the screen. In the first trial, one dinosaur produced a standard stimulus with 600Hz frequency and the other uttered a starting probe stimulus with a frequency of 982Hz. The participants were required to point to the dinosaur with the loudest sound, and if they succeeded, they were given a harder trial; conversely, if they were unsuccessful, they were given an easier trial. The task ended after forty trials or six reversals (change in direction), at which point, the participants’ discriminatory threshold could be determined. The level of difficulty of the subsequent trials was calculated using the PEST procedure, which has the characteristic that big differences are reduced to small ones relatively fast, while small differences are rigorously tested to determine the exact threshold.

The results indicated that there was no significant difference in the discrimination of pitch between the adolescents with ASD and their matched controls at the group level; however, a subgroup of 20% of the participants with ASD showed enhanced frequency discrimination. The scholars concluded that this subgroup with exceptional frequency perception may share particular defining features and represent a specific phenotype in ASD. This study was intriguing in that it indicated that the performance of the perception of pitch in ASD was diverse along the spectrum with some participants with ASD demonstrating enhanced frequency perception.

Heaton et al. (2008) examined the ability to discriminate pitch differences of fourteen children with ASD (mean age: 10;6) and fourteen matched controls (mean age: 10;6) with moderate learning difficulties or typical development. In addition to the matched chronological age (CA), these two groups had comparable BPVS verbal mental age-equivalence (8;1 and 7;4, respectively) as well as TROG verbal mental age-equivalence scores (5;5 and 5;4 respectively), although it was worth noticing that both BPVS and TROG verbal mental ages were much lower than CA for both groups.

There were three kinds of stimulus types: real word, nonce word, and pure tone. Five monosyllabic English real words and five monosyllabic nonce words were recorded for the speech stimuli in the AX discrimination task. Four pairs were created by PRAAT for each real word and nonce word. The first stimulus of each pair was the originally recorded word. In the "same" stimulus pairs, the second stimulus was identical to the first stimulus. In the "small", "medium", or "large" different stimulus pairs, the pitch contours of second stimulus were 2, 3, or 6 semitones away from that of first stimulus. Then the non-speech pure tone stimuli were created based on the segmental information of the speech stimuli. Forty trials were selected pseudo-randomly from the bank of pairs for each type of stimulus. Ten of these were "same" stimulus pairs, while thirty were "different" (ten with 2-semitone differences, ten with 3-semitone differences, and ten with 6-semitone differences). The different types of stimulus were presented in different blocks to avoid confusion, and the order of the three blocks was counterbalanced across participants. In each of the total 120 trials, the participants had to decide if the pair was "the same" or "not the same" by pressing the relevant button.

The results indicated that the participants with ASD were more sensitive than the controls to the change in pitch across the different types of auditory stimuli, which supported the EPF model. In addition, while the findings did not show a significant difference between real words and nonce words for ASD group and their matched controls, the discrimination of pure tone was significantly better than the speech stimuli for both groups. The scholars also hypothesised that enhanced auditory perception may hinder linguistic development; however, their findings were inconclusive. They found that two of the four ASD participants who scored more than 90% in their most difficult auditory discrimination condition had very low scores for the receptive language tasks. However, the scores of the other two individuals were within the normal range and there generally tended to be a positive correlation between the language scores and the performance of the auditory discrimination tasks (Heaton et al., 2008).

However, it is possible that standardised receptive language tests are too general to be sufficiently sensitive to identify the potentially negative effect of enhanced auditory processing on the ability to learn a language. There were also several flaws in this study; for example, most of the matched controls had moderate learning difficulties. Although the researchers claimed that no child with ASD was included in the control group, it was possible that children with moderate learning difficulties performed differently from TD children in pitch discrimination. It would have been better to use two control groups, one matched on language ability, i.e. a younger group, and one matched on age. Besides, the percentage of "same" pairs (25%) and "different" pairs (75%) in this study was imbalanced, and this may have affected the participants' responses; therefore, the percentage of "same" pairs and "different" pairs should be around the chance level (50%) to avoid potential bias. Finally, there was another fundamental problem in this study. That is, for the same-different decision auditory test, the examiners did not provide the baseline for the participants to demonstrate what is the "same" and what is the "different". This could be confounding especially for the participants with ASD since they tended to have hypersensitive auditory processing.

1.3 Motivation, hypothesis and research questions

It was firstly decided to test tones because they tap into pitch perception and they are also linguistic entities, at least for Mandarin speakers. Since the target to be tested was the potential interaction between pitch perception and language abilities, this appeared to be a relevant area to begin with. Secondly, there was a need to test real words and nonce words, as well as pure tones. The latter is purely about pitch perception. The real words benefit from top-down help, at least for Mandarin speakers, and the nonce words should be harder, especially if the categorical perception of the tones is affected.

Therefore, the aim of this project is to examine the link between speech perception and language acquisition in typical and atypical development to better understand the contribution of auditory mechanisms to language abilities. Since the ASD population are a particularly interesting group to test this link, a cross-linguistic perspective is adopted to investigate typically-developing and ASD individuals in the UK and Taiwan, in order to unravel the universal and specific aspects of language development in these two trajectories. The research questions are as follows:

Enhanced pitch perception

Q1: What is the psychoacoustic profile of pitch of ASD children and their typically-developing peers? Can typically-developing (TD) children and children with autism spectrum disorder (ASD) discriminate the pitch contours for real words, nonce words, and pure tones?

According to the literature, participants with ASD with language difficulties may show atypical pitch perception; thus, the following specific questions need to be answered;

Q1a: Do ASD children as a holistic group show atypical pitch perception? If so, do they demonstrate more enhanced pitch perception than their controls?

Q1b: Do ASD children with significant language problems (ASD-SLP) perceive pitch significantly better than ASD children without language problems (ASD-NLP) or their typically-developing peers?

Q1c: Do Mandarin-speaking ASD children and English-speaking ASD children both exhibit atypical pitch perception?

Role of native language

Q2: Do native languages change the perception of auditory pitch? Can Mandarin-speaking participants (tone language speakers) better discriminate pitch contours at a group level than English-speaking individuals (non-tone language speakers) at a group level?

Q2a: Can Mandarin-speaking TD (OTD/ YTD) discriminate pitch contours better than English-speaking TD (OTD/ YTD)?

Q2b: Can Mandarin-speaking ASD (SLP/ NLP) discriminate pitch contours better than English-speaking ASD (SLP/ NLP)?

Role of stimulus type (real word, nonce word, pure tone)

Q3: Do ASD children and their controls differently discriminate the pitch contours of different kinds of stimuli (real words, nonce words, and pure tones)? Is there any interaction between the group and the stimulus?

Negative correlation with general language abilities

Since the literature suggests that the ability to process auditory sounds may be the key factor of language problems;

Q4: Does the ability to perceive pitch correlate with linguistic ability?

2 Method

2.1 Participants

The same Mandarin-speaking participants in the previous experiment (twenty-one ASD participants, ten Mandarin YTD, and nine Mandarin OTD) participated in the current task. However, Mandarin KTD did not take part in this psychoacoustic tone discrimination task due to the time issue. In addition, seventeen English-speaking ASD (with a mean age of nine years), twelve English-speaking YTD (with a mean age of nine years), and five English-speaking OTD (with a mean age of nineteen years) took part in this current task. The English-speaking ASD were recruited through The Great Ormond Street Autism Clinic in London, UK, and the English-speaking YTD were recruited from a school in the research assistant's locality. Two of the participants with ASD had also been diagnosed with ADHD and Turner's syndrome respectively, and the other fifteen participants with ASD had no mental or neurological disorders. These children also completed standardised receptive vocabulary (BPVS: British Picture Vocabulary Scale) and receptive grammar (TROG-2: Test for Reception of Grammar-2) tests. In both of these tests, the children were provided with groups of four pictures and were asked to choose the associated single words (BPVS) or sentences (TROG) which were read out by the experimenter. Since the design of these two tests is similar and relatively simple, they are regarded as being suitable tools to investigate vocabulary and syntax in young children, even those who are cognitively impaired. The English-speaking ASD were further split into English NLP as well as English SLP based on the official clinical diagnosis on their language abilities. In spite of the fact that the mean CA and the mean BPVS-VMA were nearly the same in English SLP, it is decided to only focus on the TROG VMA scores to divide the sub-groups, for TROG does not only tap into the vocabulary knowledge, but also examine the grammar and syntactic development in the participants. In addition, the English-speaking OTD were first-year students at the UCL and there were awarded a credit for participating in the experiment.

Group		CA	BPVS	TROG
English NLP (n = 9)	Mean	116	114	117
	Range	79-155	93-132	109-128
English SLP (n = 8)	Mean	102	103	95
	Range	91-117	63-124	62-142
English YTD (n = 12)	Mean	110	115	110
	Range	76-145	93-139	97-134
English OTD (n = 5)	Mean	230	-	-
	Range	221-237	-	-

Table 15. CA, BPVS VMA, and TROG VMA in months for each English group

2.2 Task

This task was designed to test the discrimination of pitch contours derived from Mandarin Tones 1 and 4, using linguistic (real words, nonce words) and non-linguistic (pitch contours) stimuli in an AAx/AxA task using a 'two-up one-down' adaptive procedure. Either the second or third stimulus was the same as the first, while the fourth one was different. The participants' task was to indicate which stimulus was different from the other two. This methodology was chosen to provide the participants with a basis of comparison, unlike in an AX task (cf. Heaton et al., 2008).

2.3 Stimuli

Following Heaton et al. (2008), the pitch contours were tested in real words, nonce words and pure tone contours using the human voice. The contours were based on the contrast between Tone 1 and Tone 4.

A one-syllable Mandarin word /jie/ was chosen as the real word stimulus. The word /jie/ has a comparable imageability with all the four tones (jie1: street, jie2: knot, jie3: sister, jie4: to borrow). The frequency, syllable frequency, and onset frequency were also considered. The word /da/ had been utilised in the previous pilot tests; however, in spite of the fact that the word /da/ has a comparable imageability with all the four tones (da1: to build, da2: answer, da3:hit, da4: big) and has been used in studies that explored the lexical tones in Mandarin (Chen & Kager, 2011; Chen, 2013; Liu & Kager, 2011, 2013), there is a huge difference in the frequency of usage between da1 and da4. In other words, while da4 is used extremely frequently in Mandarin Chinese, da1 is comparatively rare; thus, it was decided to use /jie/ instead of /da/ to ensure that the frequency of usage was balanced across the word with all the four tones in order to avoid any effect of the frequency of usage.

It was known that the word /jie/ was recognised and actively used by even the least linguistically-able participants. /jie/ with four lexical tones were also utilized as stimuli in Experiment 1. Following the phonotactic rules of Mandarin, a one-syllable nonce word /chei/, was created, being a nonce word with the four lexical tones in Mandarin.

The third set of stimuli, the pure tone stimuli, was created by removing segmented information from the real word stimuli by PRAAT with the assistance of professional phoneticians. Firstly, all the sound files were converted into mono sounds of the same length (400ms), and then, the "Analyse periodicity" function was utilised to extract the pitch on its own, thereby obtaining a set of pure tone contours in a human voice.

The audio stimuli were recorded by me, as a female native Mandarin speaker. All the items were produced with the tested tones (Tones 1 to 4). The recordings were phonetically analysed to ensure that the correct tone was used. Each stimulus item was then manipulated in PRAAT to obtain 16 pitch contours falling between Tone 1 and Tone 4, as illustrated in Figure

10. Only 8 pitch contours were initially created along the /jie1-jie4/ continuum, respectively. However, the pre-test indicated that the Mandarin-adults were able to discriminate these 8 steps with comparative ease; therefore, instead of 8 steps, 16 intermediate tones were created along Tones 1 to 4 to increase the difficulty and avoid the ceiling effect.

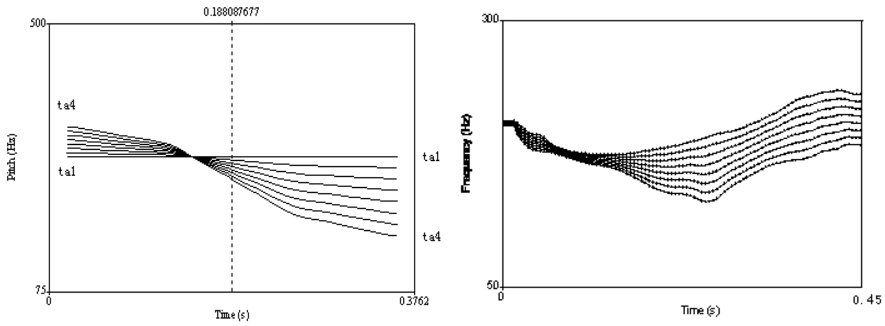

Figure 10. (a) Eight pitch contours along a /ta1-ta4/ continuum in Liu and Kager (2011); (b) /ma2-ma3/ continuum in Chen & Kager (2011)

2.4 Procedure

The aim of the task was to determine the smallest difference the participants could reliably perceive between Tone 1 and the intermediate tones on the Tone 1-Tone 4 continuum. The first sound presented was always Tone 1. The test stimulus was the second or third sound. Since the test stimulus was chosen from the intermediate tones, it constituted a 1- to 15-step difference from Tone 1. The remaining sound was again Tone 1.

The stimuli were presented using software entitled Mammoth Task developed by Judit Gervain (2014), who kindly provided it (see Figure 11 for illustration). In this task, the participants saw three dinosaurs each making a sound on the computer screen. The yellow dinosaur in the middle always produced the first stimulus, followed by the red dinosaur in the bottom left corner, and then the blue one in the bottom right. The participants' task was to indicate whether it was the red dinosaur or the blue dinosaur that was producing a different sound from the other two. Keyboard keys Q and P were used to correspond with the red dinosaur on the left of the screen and the blue dinosaur on the right of the screen, respectively.

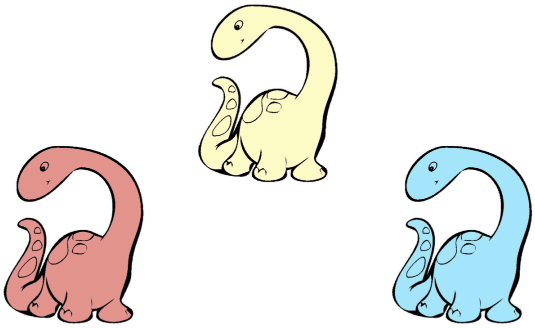

Figure 11. Illustration of the Mammouth Task

The participants' threshold was obtained by the so-called 'two-up one-down' adaptive procedure. This meant that, after the initial test stimulus, any subsequent stimulus was determined by the participants' performance. If the participant was able to correctly discriminate the initial stimulus twice consecutively, a harder stimulus was presented, i.e. one that was closer to Tone 1 on the Tone 1-4 continuum. If participants continued to correctly discriminate the new test stimulus twice consecutively, an even harder stimulus would be presented to them, and so on, until they reached the 50 items presented and the performance ceiling or until 6 reversals took place. In terms of reversals, if the participants made a mistake at any point, the next stimulus given would be easier than the one they perceived incorrectly. Thus, each mistake counted for what was called a 'reversal' and the stimulus following a 'reversal' was easier than the previous one. A 'reversal' was obtained in the opposite direction if the participant gave two consecutive correct responses after giving an incorrect response. In this case, the next stimulus would be harder than the previous two stimuli. 'Reversals' are illustrated in Figure 12 below. The test ended after 6 reversals or 50 trials, whichever was reached first.

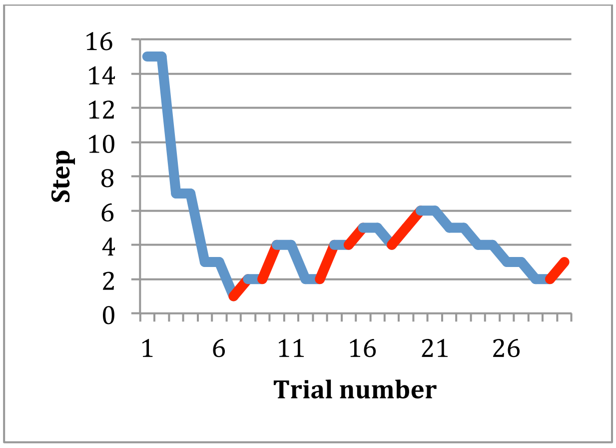

Figure 12. The steps undergone by subject 45 along the whole procedure for the nonce word /chei/. The red lines indicate "reversals".

Following the adaptive PEST (parameter estimation by sequential testing) procedure (Taylor & Creelman, 1967), this software initially uses large differences to increase the level of difficulty. While this makes the procedure much faster than using smaller differences between subsequent stimuli from the beginning, it only allows thresholds that are relatively far apart. For this reason, later in the task, smaller differences were used after 3 reversals to increase the level of difficulty (or decrease it after an error was made), which enabled the participant's threshold to be fine-tuned. This combined strategy appeared to be optimal, since it facilitated the acquisition of a relatively precise threshold within a realistic time and effort window. More specifically, during the first three reversals in the task, the test stimuli were obtained by dividing (or multiplying) the previous test stimulus by 2; therefore, the initial 15-step difference was reduced to 7 (due to rounding) after two consecutive correct responses, then to 3, then to 1, and so forth. After three reversals, the test items were obtained by dividing (or multiplying) the previous test item by 1.3, so that a 3-step difference was reduced to a 2-step difference (after rounding), and vice versa.

Since Heaton et al. (2008) indicated that participants became confused when the real word, nonce word, and pure contour stimuli were presented randomly together, it seemed that participants utilized different strategies for these three kinds of stimuli. As a consequence, the items in the three conditions were presented in a block design to prevent confusion.

As already indicated, the items were presented in an AXA or AAX format. Half of them were presented in an AXA format, while the other half were presented in an AAX format. Pseudo-randomised lists were created to counter-balance for an ordering effect. Individual lists were randomly assigned to each participant by their participant number.

Practice trials were carried out at the beginning of each block with feedback provided. The examiner initiated each trial when the child was attentive. In each trial, the participants saw three dinosaurs each making a sound on a computer screen, and then they were asked to indicate which dinosaur was producing a different sound from the other two by pressing key Q (for the red dinosaur on the left) or key P (for the blue dinosaur on the right). The main trials were designed following the practice trials, but without any feedback. Since there were 3 kinds of stimuli (real words, nonce words, and pitch contours), there were 3 conditions in the main trials and it took around 15 minutes to complete them.

3 Results

3.1 Mandarin and English results for the three stimulus types

The responses of each participant in the computerised pitch discrimination task were imported into the SPSS programme for an analysis. Means, standard deviations, and ranges for thresholds across experimental conditions for the three groups are shown in Table 16. The mean threshold was calculated by the means of stimulus levels of the last three reversals. As mentioned above, the adaptive procedure in Mammoth enabled the acquisition of a discriminatory threshold as the result of each condition. The better the discrimination ability an individual had, the smaller the number of the threshold he or she would get. (The minimum threshold was 1, and the maximum was 16.)

As mentioned earlier, eleven of the twenty-one Mandarin-speaking ASD children had no significant language problems (Mandarin NLP), while the other ten had significant language problems (Mandarin SLP). Therefore, the results are also analysed according to their language ability.

Overall speaking, the threshold of nonce words (M = 2.02) was higher than the threshold of real words (M = 1.43) and pure tones (M = 1.69). In addition, the threshold across stimulus types was higher in English overall (M = 1.77) than in Mandarin overall (M = 1.68). The threshold across stimulus types was higher in English ASD (M = 1.97) than in English TD (M = 1.55). Similarly, the threshold across stimulus types was higher in Mandarin ASD (M = 1.96) than in Mandarin TD (M = 1.36). When four English sub-groups were compared together, it turned out that the threshold across stimulus types was highest in English NLP (M = 2.21) and lowest in English OTD (M = 1.29). When five Mandarin sub-groups were compared together, it turned out that the threshold across stimulus types was highest in Mandarin SLP (M = 2.76) and lowest in Mandarin YTD (M = 1.30). The highest threshold across the table was the threshold of nonce words in Mandarin SLP (M = 3.96), whereas the lowest threshold across the table was the threshold of pure tones in Mandarin OTD (M = 1.11). It is worth noticing that there was an enormous range in the threshold of nonce words in Mandarin SLP (1.00-12.71), and the 95% confidence interval was from 2.182 to 4.540. (The range in the threshold of real words in Mandarin SLP was 1.00-2.60, and the 95% confidence interval was from 1.204 to 2.033. The range in the threshold of pure tones in Mandarin SLP was 1.00-5.40, and the 95% confidence interval was from 0.498 to 2.891.)

			Real word	Nonce word	Pure tone	Total
Mandarin ASD (n = 21)		Mean	1.43	2.70	1.65	1.96
		SD	0.55	3.12	1.29	2.06
		Range	1.00-2.69	1.00-12.71	1.00-5.40	1.00-12.71
	Mandarin NLP (n = 7)	Mean	1.38	1.52	1.27	1.39
		SD	0.59	0.30	0.43	0.27
		Range	1.00-2.69	1.11-2.00	1.00-2.21	1.00-2.69
	Mandarin SLP (n = 9)	Mean	1.62	3.96	2.13	2.76
		SD	0.68	4.55	1.95	2.73
		Range	1.00-2.60	1.00-12.71	1.00-5.40	1.00-12.71

	Mandarin YNLP (*n* = 5)	Mean	1.25	2.09	1.42	1.58
		SD	0.18	0.67	0.35	0.27
		Range	1.13-1.57	1.33-2.80	1.00-1.80	1.00-2.80
Mandarin TD (*n* = 19)		Mean	1.37	1.64	1.12	1.36
		SD	0.41	0.75	0.14	0.36
		Range	1.00-2.67	1.00-4.08	1.00-1.53	1.00-4.08
	Mandarin YTD (*n* =10)	Mean	1.27	1.52	1.13	1.30
		SD	0.19	0.47	0.16	0.33
		Range	1.00-1.57	1.14-2.56	1.00-1.53	1.00-2.56
	Mandarin OTD (*n* = 9)	Mean	1.48	1.77	1.11	1.46
		SD	0.56	0.99	0.10	0.69
		Range	1.08-2.67	1.00-4.08	1.00-1.25	1.00-4.08
Mandarin Overall (*n* = 40)		Mean	1.40	2.23	1.40	1.68
		SD	0.48	2.40	0.96	1.56
		Range	1.00-2.69	1.00-12.71	1.00-5.40	1.00-12.71
English ASD (*n* = 17)		Mean	1.54	1.89	2.50	1.97
		SD	0.70	1.49	2.73	1.83
		Range	1.00-3.50	1.06-7.13	1.00-11.36	1.00-11.36
	English NLP (*n* = 9)	Mean	1.37	2.25	3.10	2.21
		SD	0.63	1.94	3.60	2.34
		Range	1.00-3.04	1.06-7.13	1.07-11.36	1.00-11.36
	English SLP (*n* = 8)	Mean	1.74	1.43	1.90	1.70
		SD	0.76	0.23	1.50	0.98
		Range	1.13-3.50	1.09-1.75	1.00-5.40	1.00-5.40
English TD (*n* = 17)		Mean	1.32	1.62	1.55	1.55
		SD	0.19	0.56	0.78	0.53
		Range	1.05-1.67	1.13-3.08	1.07-4.08	1.05-4.08
	English YTD (*n* = 12)	Mean	1.45	1.72	1.71	1.66
		SD	0.15	0.64	0.94	0.70
		Range	1.33-1.67	1.13-3.08	1.11-4.08	1.11-4.08
	English OTD (*n* = 5)	Mean	1.18	1.43	1.26	1.29
		SD	0.12	0.36	0.16	0.25
		Range	1.05-1.29	1.17-2.06	1.07-1.50	1.05-2.06
English Overall (*n* = 34)		Mean	1.46	1.76	2.06	1.77
		SD	0.57	1.14	2.09	1.44
		Range	1.00-3.50	1.06-7.13	1.00-11.36	1.00-11.36
Overall (*n* = 74)		Mean	1.43	2.02	1.69	1.72
		SD	0.51	1.95	1.58	1.50
		Range	1.00-3.50	1.00-12.71	1.00-11.36	1.00-12.71

Table 16. Mean thresholds, standard deviation (SD), and ranges of pitch discrimination

A mixed ANOVA was performed with one within-subject factor, Stimulus (Real word, Nonce word, Pure tone), and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD, English NLP, English SLP, English YTD, English OTD). There was no main effect of Group $F(8, 50) = .858$, $p = .558$, $\eta^2 = .121$, and there was no main effect of Stimulus $F(2, 100) = 1.605$, $p = .206$, $\eta^2 = .031$. Bonferroni post-hoc tests indicated that the threshold of three kinds of stimulus types did not significantly differ from each other. The threshold of nonce words was not significantly higher than that of real word (p = .155) nor pure tones (p = .962), and the threshold of pure tones was not significantly higher than that of real words (p = 1.000). There was no interaction between the within-subject factor, Stimulus, and the between-subject factor, Group F(16,100) = 1.407, p = .154, η^2 = .184. However, Bonferroni post-hoc tests indicated that the threshold of nonce words in Mandarin SLP was significantly higher than that of real words (p = .006) and pure tones (p = .038). In addition, the threshold of pure tones in English NLP was significantly higher than that of real word (p = .015), and marginally higher than that of nonce words (p = .059).

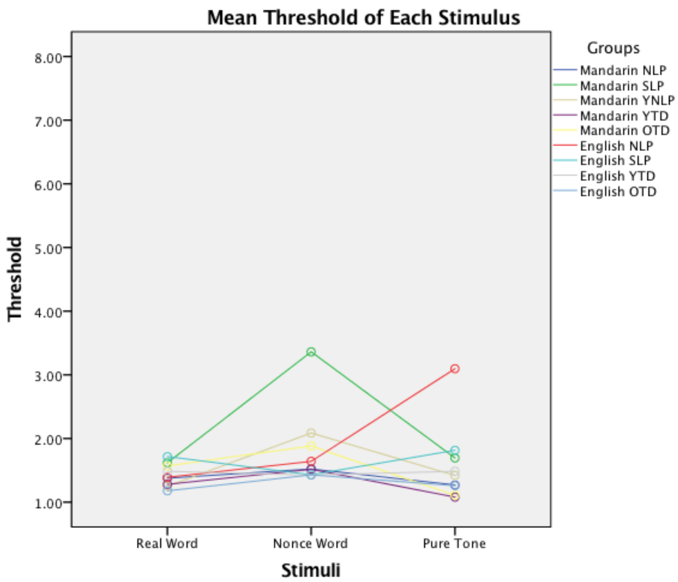

Figure 13. Mean threshold of each stimulus for each group

A mixed ANOVA was performed with one within-subject factor, Stimulus (Real word, Nonce word, Pure tone) and one between-subject factor, Group (Mandarin ASD, Mandarin TD). There was no main effect of Group F(1, 33) = 1.134, p = .295, η^2 = .033. There was a main effect of Stimulus F(2, 66) = 5.864, p = .005, η^2 = .151. Bonferroni post-hoc tests indicated that the threshold of nonce words was significantly higher than that of pure tones (p = .008), but not significantly higher than that of real words (p = .166). The threshold of real words was not significantly higher than that of pure tones (p = .758). There was no interaction between the within-subject factor, Stimulus, and the between-subject factor, Group F(2, 66) = 1.009, p = .370, η^2 = .030.

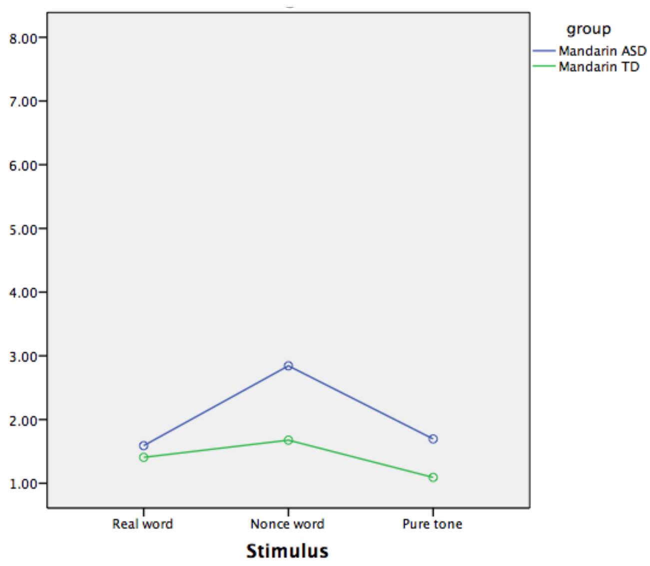

Figure 14. Mean threshold of each stimulus for Mandarin ASD and Mandarin TD

A mixed ANOVA was performed with one within-subject factor, Stimulus (Real word, Nonce word, Pure tone) and one between-subject factor, Group (Mandarin ASD, English ASD). There was no main effect of Group F(1, 32) = .075, p = .786, η^2 = .002, and there was no main effect of Stimulus F(2,64) = 1.231, p = .299 , η^2 = .037. Bonferroni post-hoc tests indicated that the threshold of three kinds of stimulus types did not significantly differ from each other. The threshold of pure tones was not significantly higher than that of real words (p = .552) nor nonce words (p = 1.000), and the threshold of nonce words was not significantly higher than that of real words (p = .468). There was a significant interaction between the within-subject factor, Stimulus and the between-subject factor, Group F(2,64) = 3.420, p = .039, η^2 = .097. Bonferroni post-tests did show any significance, but LSD post-hoc tests showed the threshold of nonce words was significantly higher than that of real words (p = .036), and marginally higher than that of pure tones (p = .083) in Mandarin ASD. In addition, the threshold of pure tones was marginally higher than that of real words (p = .089) and nonce words (p = .095) in English ASD.

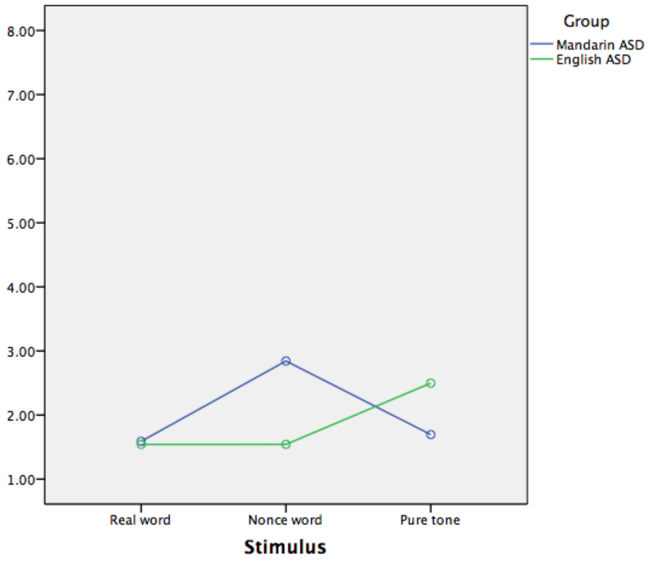

Figure 15. Mean threshold of each stimulus for Mandarin ASD and English ASD

A mixed ANOVA was performed with one within-subject factor, Stimulus (Real word, Nonce word, Pure tone) and one between-subject factor, Group (Mandarin TD, English TD). There was no main effect of Group $F(1,23) = .034$, $p = .856$, $\eta^2 = .001$. However, there was a main effect of Stimulus $F(2,46) = 4.851$, $p = .012$, $\eta^2 = .174$. Bonferroni post-hoc tests indicated that the threshold of nonce words was marginally higher than that of pure tones ($p = .069$), but was not significantly higher than that of real words ($p = .182$). The threshold of real words was not significantly higher than that of pure tones ($p = .274$). There was a significant interaction between the within-subject factor, Stimulus, and the between-subject factor, Group $F(2,46) = 3.193$, $p = .050$, $\eta^2 = .122$. Bonferroni post-hoc tests indicated that the threshold of pure tones was significantly lower than that of real words ($p = .007$) and nonce words ($p = .004$), and the threshold of real words was marginally lower than that of nonce words ($p = .088$) for Mandarin TD. In addition, Bonferroni post-hoc tests showed that the threshold of Mandarin was significantly lower than that of English in pure tones ($p = .001$).

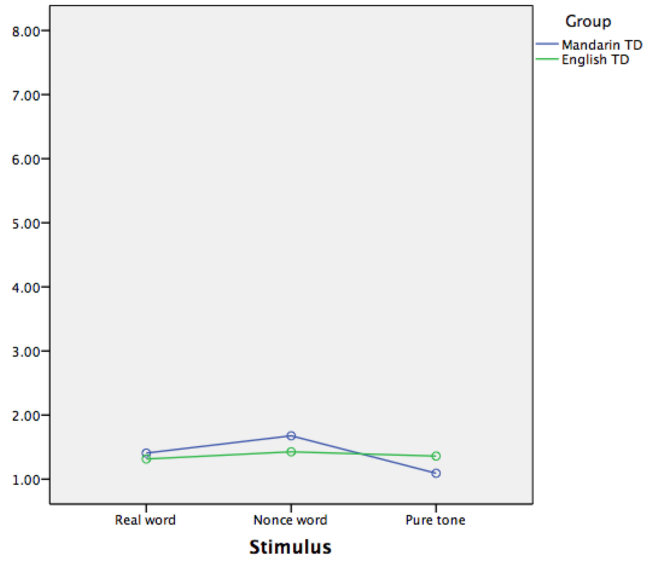

Figure 16. Mean threshold of each stimulus for Mandarin TD and English TD

3.2 Correlations between the linguistic abilities and the thresholds

For Mandarin NLP, the PPVT VMA had a significant very strong negative correlation with the threshold of nonce word (r = -.885, p = .046), whereas for Mandarin SLP and Mandarin YNLP, there was no significant correlation between the PPVT VMA and the thresholds. For overall Mandarin ASD, the PPVT VMA had a significant strong negative correlation with the threshold of nonce word (r = -.756, p = .007) and pure tone (r = -.648, p = .031). Since it is possible that the correlation tests did show a significant correlation between the PPVT VMA and the threshold of real word because of lack of power effect, we would now explore the correlation between the PPVT VMA and the average of thresholds of three stimulus types.

		Real word	Nonce word	Pure tone	PPVT
Real word	Pearson Correlation	1	-.639	-.280	.399
	Sig. (2-tailed)	-	.246	.649	.506
Nonce word	Pearson Correlation	-.639	1	-.132	-.885*
	Sig. (2-tailed)	.246	-	.833	.046
Pure tone	Pearson Correlation	-.280	-.132	1	.562
	Sig. (2-tailed)	.649	.833	-	.324
PPVT	Pearson	.399	-.885*	.562	1

	Correlation				
	Sig. (2-tailed)	.506	.046	.324	-

Table 17. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for Mandarin NLP

		Real word	Nonce word	Pure tone	PPVT
Real word	Pearson Correlation	1	.673	.612	-.339
	Sig. (2-tailed)	-	.213	.273	.577
Nonce word	Pearson Correlation	.673	1	.994**	-.700
	Sig. (2-tailed)	.213	-	.000	.121
Pure tone	Pearson Correlation	.612	.994**	1	-.683
	Sig. (2-tailed)	.273	.000	-	.135
PPVT	Pearson Correlation	-.339	-.700	-.683	1
	Sig. (2-tailed)	.577	.121	.135	-

Table 18. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for Mandarin SLP

		Real word	Nonce word	Pure tone	PPVT
Real word	Pearson Correlation	1	-.853	.761	.541
	Sig. (2-tailed)	-	.147	.239	.459
Nonce word	Pearson Correlation	-.853	1	-.338	-.860
	Sig. (2-tailed)	.147	-	.662	.140
Pure tone	Pearson Correlation	.761	-.338	1	.048
	Sig. (2-tailed)	.239	.662	-	.952
PPVT	Pearson Correlation	.541	-.860	.048	1
	Sig. (2-tailed)	.459	.140	.952	-

Table 19. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for Mandarin YNLP

		Real word	Nonce word	Pure tone	PPVT
Real word	Pearson Correlation	1	.709*	.601	-.479
	Sig. (2-tailed)	-	.022	.066	.161
Nonce word	Pearson Correlation	.709*	1	.964**	-.756**
	Sig.	.022	-	.000	.007

	(2-tailed)				
Pure tone	Pearson Correlation	.601	.964**	1	-.648*
	Sig. (2-tailed)	.066	.000	-	.031
PPVT	Pearson Correlation	-.479	-.756**	-.648	1
	Sig. (2-tailed)	.161	.007	.031	-

Table 20. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for Mandarin ASD

As we expected, the average of thresholds of three stimulus types had a significant strong negative correlation with the PPVT VMA for Mandarin ASD (r = -.633, p = .037). By contrast, for English ASD, there was no significant correlation between the average of thresholds of three stimulus types and the BPVS VMA (r = -.020, p = .949) or TROG VMA (r = .349, p = .242).

		Average	PPVT
Average	Pearson Correlation	1	-.633*
	Sig. (2-tailed)	-	.037
PPVT	Pearson Correlation	-.633*	1
	Sig. (2-tailed)	.037	-

Table 21. Correlations between the average of thresholds of three stimulus types and PPVT VMA for Mandarin ASD

		Average	BPVS	TROG
Average	Pearson Correlation	1	-.020	.349
	Sig. (2-tailed)	-	.949	.242
BPVS	Pearson Correlation	-.020	1	.429
	Sig. (2-tailed)	.949	-	.165
TROG	Pearson Correlation	.349	.429	1
	Sig. (2-tailed)	.242	.165	-

Table 22. Correlations between the average of thresholds of three stimulus types, BPVS VMA, and TROG VMA for English ASD

3.3 Correlations between the thresholds of three stimulus types

As shown in Table 20, for Mandarin ASD, the threshold of real word was significantly strongly positively correlated with the threshold of nonce word

(r = .709, p = .022). In addition, there was a significant very strong positive correlation between the threshold of nonce word and that of pure tone (r = .964, p < .001). Then the Mandarin ASD was further split into three subgroups according to their linguistic abilities as well as ages. For Mandarin NLP and Mandarin YNLP, there was no significant correlation between the thresholds of the three stimulus types. On the other hand, for Mandarin SLP, there was a significant very strong positive correlation between the threshold of nonce word and that of pure tone (r = .994, p < .001).

For Mandarin YTD, there was a significant very strong positive correlation between the threshold of real word and that of nonce word (r = .838, p = .005). Besides, there was a significant strong positive correlation between the threshold of nonce word and that of pure tone (r = .693, p = .039). This pattern was very similar to the patterns found in Mandarin ASD. As for the Mandarin OTD, there was a significant very strong positive correlation between the threshold of real word and that of nonce word (r = .713, p = .047).

Unlike Mandarin participants, there was no significant correlation between the thresholds of three stimulus types for both English ASD and TD groups.

		Real word	Nonce word	Pure tone
Real word	Pearson Correlation	1	.838**	.048
	Sig. (2-tailed)	-	.005	.895
Nonce word	Pearson Correlation	.838**	1	.693*
	Sig. (2-tailed)	.005	-	.039
Pure tone	Pearson Correlation	.048	.693*	1
	Sig. (2-tailed)	.895	.039	-

Table 23. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for Mandarin YTD

		Real word	Nonce word	Pure tone
Real word	Pearson Correlation	1	.713*	.329

	Sig. (2-tailed)	-	.047	.426
Nonce word	Pearson Correlation	.713*	1	-.002
	Sig. (2-tailed)	.047	-	.997
Pure tone	Pearson Correlation	.329	-.002	1
	Sig. (2-tailed)	.426	.997	-

Table 24. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for Mandarin OTD

		Real word	Nonce word	Pure tone
Real word	Pearson Correlation	1	-.008	-.143
	Sig. (2-tailed)	-	.976	.597
Nonce word	Pearson Correlation	-.008	1	.445
	Sig. (2-tailed)	.976	-	.097
Pure tone	Pearson Correlation	-.143	.445	1
	Sig. (2-tailed)	.597	.097	-

Table 25. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for English ASD

		Real word	Nonce word	Pure tone
Real word	Pearson Correlation	1	.153	.473
	Sig. (2-tailed)	-	.695	.199
Nonce word	Pearson Correlation	.153	1	.310
	Sig. (2-tailed)	.695	-	.281
Pure tone	Pearson Correlation	.473	.310	1
	Sig. (2-tailed)	.199	.281	-

Table 26. Correlations between the threshold of real word, the threshold of nonce word, the threshold of pure tone, and PPVT VMA for English TD

4 Discussion

Enhanced pitch perception

Q1: What is the psychoacoustic profile of pitch of ASD children and their typical peers? Can typically developing (TD) children and

children with autism spectrum disorders (ASD) discriminate the pitch contours for real words, nonce words, and pure tones?

The results of current study showed that ASD children and TD children were both sensitive to changes in pitch contours across different auditory stimulus types. While the minimum of threshold was 1 and the maximum was 16, both ASD and TD groups had around 2 for the threshold. It is worthwhile noting that there might be a ceiling effect since most of the participants had very good performance in the current study (Table 16).

Q1a: Do ASD children as the whole group show atypical pitch perception? If yes, do they demonstrate the enhanced pitch perception than their controls?

There was no significant difference in the pitch perception across stimulus types between the ASD and TD group, and this could be possibly attributed to the ceiling effect mentioned earlier.

Q1b: Do ASD children with significant language problems (ASD-SLP) perform significantly better in pitch perception than ASD children without language problems (ASD-NLP) or the typical peers?

Even when the ASD children were split into two groups according to their linguistic abilities, there was still no significant difference between the groups (Figure 13).

Q1c: Do Mandarin-speaking ASD children as well as English-speaking ASD children both show atypical pitch perception?

I found no evidence of enhanced pitch perception in my study of Tone 1-4 differences for ASD participants in English or Mandarin. There was also no evidence of enhanced perception for ASD children with language problems over ASD children with no language problems. All groups performed very well on the task, reaching very low thresholds. The ASD children did not have lower thresholds than the typically developing children in either language (Figure 13 and 14).

It is possible that the lack of differentiation between the groups is due to a ceiling effect. In other words, it is possible that either Tone 1-4 differences are too easy to perceive. In future research one could investigate Tone 2-3 differences, which are known to be harder to distinguish. Another possibility would be to make a finer-grained slicing up the Tone 1-4 continuum using 32 instead of 16 intermediate steps.

Role of stimulus type (real word, nonce word, pure tone)

Q2: Do ASD children and their controls perform differently in discrimination of pitch contour toward different kinds of stimuli (real words, nonce words, and pure tones)? Are there any interaction between the Group and the Stimulus?

Mandarin typically developing participants had significantly lower thresholds for pure tone stimuli than for nonce word stimuli (Figure 14), suggesting that it was much easier for Mandarin TD group to perceive and discriminate the non-speech stimuli than the speech stimuli. So, for these participants, we found an effect of language, as found by Heaton et al. (2008) for English participants, who also had better discrimination performance in non-speech stimuli than in speech stimuli. It was found that Mandarin participants, both typically developing and ASD participants, had a significantly higher threshold when tested on nonce words, compared to pure tone stimuli (Figure 14). In addition, we found a further disadvantage for Mandarin SLP participants compared to the other Mandarin groups, although this was mainly due to a few outlier participants (Figure 13). For Mandarin typically developing children there was also a significant difference between their threshold on pure tone stimuli and real words, and their real words' threshold was marginally lower than their nonce word threshold (Figure 16). For ASD participants, we have only found a significant difference between nonce words and pure tone stimuli.

We interpret this as nonce words presenting a more difficult task for native speakers. This effect may even be stronger for participants with ASD and a language problem. The difficulty of nonce word stimuli may come from the fact that no lexical supporting effect is present for nonce words, compared to real words. So, there is no top down effect from word recognition: participants must rely on their knowledge of the abstract tonal categories Tone 1 and Tone 4.

Furthermore, the correlation tests indicated that the performance of real word and nonce word was strongly correlated in Mandarin OTD, suggesting that Mandarin OTD actually treated the nonce word as the real word (Table 24). As discussed in the introduction section, native speakers of tone languages could store abstract schema such as the ones in (3):

(3) a. $[\ ]_{\text{word}}$ – Tone 1
    b. $[\ ]_{\text{word}}$ – Tone 4

Therefore, Mandarin OTD had the stable abstract schema based on the previous linguistic experience on real words, and then applied the abstract schema to the nonce words.

As for Mandarin YTD group, it is intriguing to see that the threshold of nonce word did not only correlate with that of real word, but also correlated with that of pure tone in a even stronger way (Table 23). This might suggest that although Mandarin YTD had established the abstract schema for the lexical tones, the abstract schema was still not stable enough for them to process the nonce words completely. As a consequence, Mandarin-speaking young children may perceive the nonce words something between the speech real words as well as non-speech pure tones.

Mandarin ASD also had very similar correlation results just as Mandarin YTD, suggesting that overall Mandarin ASD did not have sophisticated abstract schema of lexical tones as Mandarin OTD, but treated the nonce words as the way Mandarin YTD did (Table 20). Although sometimes Mandarin ASD and YTD could perceive the nonce words as real words, they actually showed a preference to treat the nonce words to be the non-speech stimuli. Moreover, the participants with ASD with significant language problems (i.e. Mandarin SLP) only had significant correlation between the threshold of nonce word and pure tone, which indicated that they mainly perceive the nonce words as the non-speech stimuli. The results of Mandarin participants altogether manifested that the linguistic abilities might play an important role in perceiving the nonce words. The better the linguistic abilities an individual possesses, the more likely that he or she would have stable abstract linguistic schema to apply to the perception of nonce words.

As for the English participants, there was no significant correlation at all between the three stimulus types (Table 25 and 26). It is not surprising since they would not have any experience in lexical tones and thus establish the abstract schema for them in the first place. Each kind of stimuli presented various levels of challenges for them to perceive the pitch contours in the task.

Role of native language

Q3: Do native languages change the auditory pitch perception? Are Mandarin-speaking participants (tone language speakers) at the group level better in discrimination of pitch contours than English-speaking individuals (non-tone language speakers) at the group level?

Mandarin speakers do not have an overall advantage, but they do seem to treat nonce words and to some extent real word stimuli differently for pure tone ones, while English participants did not do that (Table 17, 18, 19, and 20).

Negative correlation with general language abilities

Q4: Since the literature suggests that the auditory processing abilities may be the key factor for the language problems, do pitch perception abilities correlate to linguistic abilities?

While Mandarin NLP had a significant correlation between the PPVT VMA and the threshold of nonce word and Mandarin SLP had no significant correlation between the PPVT VMA and the three kinds of stimulus types, overall Mandarin ASD had a significant correlation between the linguistic abilities and the thresholds of nonce word as well as pure tones, but not with the threshold of real words. Since it could be possibly attributed to the effect of lack of power, we further explored the correlation between the PPVT VMA and the average of thresholds of three stimulus types. The results indicated that the overall performance in the tone discrimination task was strongly correlated with the linguistic abilities for Mandarin ASD. Nevertheless, for English ASD there was no significant correlation between the average of thresholds and the BPVS VMA or TROG VMA. This might be attributed that English ASD did not perceive the stimuli combined with the non-native lexical tones in a linguistic way. In addition, pitch contours in non-tone languages such as English were not as important as in tone languages like Mandarin Chinese. As a consequence, even an individual performs exceptionally well in the receptive vocabulary and grammar tasks in English, it does not guarantee the sensitive discrimination of non-native lexical tones.

Chapter 4: Categorical perception task

1 Introduction

1.1 Aim of current work

Previous linguistic scholars (Abramson, 1979; Burnham & Jones, 2002; Francis, Ciocca, & Ng, 2003; Halle, Chang, & Best, 2004) generally utilised identification, as well as discrimination tasks to explore the perception of lexical tones of speakers of a tone language as well as those of a non-tone language. They found that speakers of a tone language categorically perceived their native lexical tones, whereas speakers of a non-tone language perceived and processes lexical tones in a psychophysical way. Further studies were conducted to investigate the perception of lexical tones of advanced learners of a tone language (Chen & Kager, 2011) or bilinguals whose primary language was a tone language (Yang & Liu, 2006). The results indicated that exposure to a tone language may lead individuals to associate certain acoustic pitch contours with linguistic tonal categories and thus, identify and discriminate lexical tones with their categorical perception to some degree. However, it is worth noting that the slope of identification that functions around the category boundary was found to be not as steep in these participants as in Mandarin monolinguals. This raises the question of whether or not Mandarin children with ASD, who possess sensitive auditory perception, perceive lexical tones in a categorical way when they are sufficiently exposed to them. The way in which participants with ASD of a tone language and those of a non-tone language perceive lexical tones remains unexplained in the literature. Therefore, the aim of this experiment is to investigate how Mandarin-speaking individuals with ASD, as well as English-speaking participants with ASD, perceive lexical tones and whether or not their perception is categorical.

In spite of the fact that the previous tasks in Chapter 2 and 3 have not picked up ultrasensitive auditory perception in the current ASD populations, the results do show that some participants with ASD (especially the ones with language problems) might have weaker representations of abstract tones than their typically developing peers. As a consequence, the aim of the categorical perception task is to explore whether participants with ASD would also exhibit weaker categorical perception of tones.

1.2 Previous work

A categorical perception experiment is a classical paradigm to examine the perception of native or non-native sound categories. It is assumed that, since native speakers have established phonologically-contrasting categories of speech sounds for efficient processing, they mainly focus on the differences between category boundaries, while ignoring the irrelevant superficial variations within those boundaries. The categorical perception paradigm usually utilises an acoustical gradient continuum between two endpoint tokens representing two stable contrasting sound categories, and there are usually a pair of tasks, namely, an identification task and a discrimination task. Since the identification task demonstrates the location of the boundary category, the discrimination task can be based on this information in order to determine if there is an enhanced discrimination peak around the category boundary.

The categorical perception of segmental features of speech was illustrated by early linguistic scholars (Fry, Abramson, Eimas, & Liberman 1962; Liberman, Harris, Hofman, & Griffith 1957; Liverman, Harris, Eimas, Lisker, & Bastian 1961; Pisoni 1973; Repp 1984), among whom, Liberman et al. (1957) conducted a well-known study of the categorical perception of voiced stop consonants (/b/, /d/, and /g/) in nine English adults. Spectrograms were utilised to produce fourteen continua (Stimulus 1 to Stimulus 14) along the consonant-vowel syllables from /be/ to /de/ to /ge/. The test consisted of a labelling task and an ABX discrimination task. In the labelling task, each of the fourteen stimuli along the continuum was played once in a random order with a 6-second interval between them. The participants were tasked with labelling each stimulus as /b/, /d/, or /g/. The ABX discrimination task involved three stages: one-step, two-step, and three-step discrimination. A and B stimuli were adjacent steps (e.g. Stimulus 1 versus Stimulus 2) for the one-step discrimination condition, A and B stimuli had a two-step difference (e.g. Stimulus 1 versus Stimulus 3) for the two-step condition, A and B had a three-step difference (e.g. Stimulus 1 versus Stimulus 4) for the three-step condition. X stimuli were identical to either A or B, and the task for the participants was to determine if X was the same as A or B.

The results indicated that, while Stimuli 1 to 3 were generally labelled as /b/, there was an abrupt shift around Stimulus 4, and Stimuli 5 to 9 were primarily judged as /d/. Similarly, there was a sudden change around Stimulus 10, and Stimuli 11 to 14 were considered as /ge/. The authors believed that the participants had already established sharp and stable phoneme boundaries, which led them to perceive the consonant features in a categorical, rather than a psychoacoustic, way. Moreover, a comparison of the labelling and discrimination functions revealed a higher percentage of correct discrimination between the phoneme boundaries than within them; therefore, the researchers proposed that the labelling curves may have predicted the discrimination values. They then made scatter plots of the values obtained in the discrimination task and compared them with the predicted values from the labelling task and found that a significant relationship existed between the obtained and predicted points in the two-step and three-step discrimination data. They concluded that individuals tended to perceive the stop consonants in a categorical way; in other words, they focused on the differences between category boundaries while ignoring the irrelevant superficial variation within the category boundaries. This led to a prominent peak around the category boundaries in the discrimination task and the predictability of the two-step or three-step discrimination functions of stop consonants from the curves of a corresponding labelling task.

Despite the categorical perception of features of consonants, such as the voicing or placing of articulation, the perception of vowels appears to be more continuous. Fry, Abramson, Eimas, and Liberman (1962) explored eight English adults' perception of vowels with a similar paradigm as that utilised by Liberman et al. (1957). Spectrograms were utilised to produce thirteen continua (Stimulus 1 to Stimulus 13) evenly along the vowels from /ɪ/ to /ɛ/ to /æ/. The test consisted of an ABX discrimination task, as well as an ABX identification task. The discrimination task contained the same one-step, two-step, and three-step conditions as in Liberman et al. (1957), and while A and B were always different from each other in each step, X was either identical to A or B. The participants were tasked with determining whether X was the same as A or B. The same stimuli were utilised in a different order for the ABX identification task, and this time, the participants were tasked with labelling each stimulus as /ɪ/, /ɛ/, or /æ/.

The results indicated that the identification of the curve of the vowels was not as steep as that of the curve of the consonants. Besides, the discrimination function of the vowels did not show a marked increase in sensitivity around the region of the phoneme boundaries, unlike that of the consonants. These intriguing results demonstrated that the perception of vowels was different from that of consonants. While individuals perceived consonants categorically, they perceived vowels in a more gradient and continuous way. The researchers speculated that the sharpness of the phoneme boundaries might be correlated with the degree of articulatory discontinuity between sounds. From this perspective, it may be worth exploring the perception of other phonological units, like lexical tones, in order to shed light on the field of categorical perception from a different angle.

Halle, Chang, and Best (2004) explored the identification and discrimination of Mandarin lexical tones in Mandarin- (tone language) and French-speaking (non-tone language) adults. The fourteen Mandarin-speaking adults (aged 22-30 years) were from Taiwan, and the fourteen French-speaking adults (aged 20-31 years) were recruited in France. None of the French participants had ever been exposed to any tone language. Three Mandarin syllables /pa/, /pi/, and /kwo/ with four lexical tones were recorded for the stimuli, which were on a continuum along Tone 1 and Tone 2, Tone 2 and Tone 4, as well as Tone 3 and Tone 4. Each continuum proceeded through eight steps from one endpoint (step 1) to the other (step 8) with the result that there were seventy-two stimuli in total (three syllable x eight steps x three tonal pairs).

Test 1 for Mandarin participants contained an identification task and a discrimination task. In the identification task, the Mandarin participants were presented with a sentence “yi ge X zi” (“one character X”) in each trial, and the X was chosen from the seventy-two stimuli in a quasi-random order. The participants were asked to choose between the two Chinese characters that represented the two endpoint tones for each continuum to represent the tone of the target syllable. As for the AXB two-step discrimination task, there were six possible A-B pairs (step 1-step 3, step 2-step 4, step 3-step 5, step 4-step 6, step 5-step 7, and step 6-step 8), and four possible AXB combinations (AAB, ABB, BAA, and BBA). A and B were stimuli with the same Mandarin syllable generated from the same tonal pairs. The only difference between A and B was the step (two-step difference), and X was identical to either A or B. The Mandarin participants were presented with three stimuli, AXB, in each trial and were asked to discriminate between them and determine if X was identical to A or B by choosing the related keys.

The authors matched the Gaussian distributive function to the participants' individual identification curves in order to estimate the slopes, as well as the intercepts, of their identification function. The results indicated that the slopes did not differ significantly across the tonal continua in the identification task, whereas the intercepts differed significantly on the tonal continuum. (3.77 in Tone 1-Tone 2, 4.99 in Tone 2-Tone 4, and 5.75 in Tone 3-Tone 4).

Around 88% of the responses in the discrimination task were correct, and this did not significantly differ across the tonal continua. However, it is important to note that the correct discrimination of pair step 3-step 5 and step 4-step 6 was significantly higher than the adjacent pairs. The researchers proposed that Mandarin-speaking adults demonstrated a categorical perception of the lexical tones because the slopes were significantly steeper at the category boundary of an identification curve. They further noted that the distinction between the proportion of correct responses across category boundaries and the proportion of correct responses for within-category items were not as large as were normally found for the categorical perception of consonants. This resembled the shallower, wider patterns found with vowels.

In Test 2, the researchers conducted an AXB identification task for Mandarin- and French-speaking participants in order to compare and contrast the perception of lexical tones of native speakers of a tone language as well as of a non-tone language. The stimuli were the same as those in Test 1 (but excluding the syllable /kuo/). A, B, and X all had the same Mandarin syllables in this AXB identification test. A and B were the two endpoints in the two possible orders, and X varied from one endpoint to another along the eight steps of the tonal continuum. In each trial the participants were presented with the three stimuli A, B and X, and were asked to identify X as A or B by pressing the corresponding button. Strictly speaking, this was not a labelling task as is customary in a categorical perception test because non-native speakers cannot be expected to label the tones.

The results indicated a group effect of intercepts. While the intercepts of the French participants were around the exact centre (4.5) of the tonal continuum, those of the Mandarin participants fell left to centre for the Tone

1-Tone 2 as well as the Tone 3-Tone 4 continua. More importantly, Mandarin participants' slopes at crossover were significantly steeper than those of their French counterparts. These results manifested that French-speaking individuals perceive lexical tones in a psychophysical way, whereas Mandarin-speaking adults perceive them categorically.

The authors also conducted an AXB two-step discrimination task in Test 3 to explore French adults' discrimination of lexical tones. The stimuli and the procedure were very similar to the AXB discrimination task in Test 1 (but excluding the syllable /kuo/). The results demonstrated that, unlike their Mandarin counterparts, the French participants did not show enhanced discrimination at category boundary crossover, thereby confirming that speakers of a non-tone language do not perceive or process lexical tones in a linguistic way and categorise them differently.

In summary, as shown by the aforementioned studies of the categorical perception of stop consonants (Liberman et al., 1957, 1961), categorical perception has several defined characteristics, including a sharp slope, a corresponding discrimination peak around the category boundary, and a predictable discrimination performance from the identification results. Halle, Chang, and Best (2004) implemented several tasks to explore Mandarin and French-speaking adults' identification and discrimination of lexical tones. The intercepts and slopes gathered in the identification tasks were useful to examine the participants' perception and processing of lexical tones. If the participants treated lexical tones in a psychophysical way (like the French), the intercepts were around the exact centre of the tone continuum, and there was no significant difference between the slopes at crossover. In addition, there was no enhanced discrimination at the category boundary. On the other hand, Mandarin-speaking adults categorised Mandarin lexical tones differently; thus, the intercepts were skewed from the exact centre of the tonal continuum and the slopes at the category boundary were significantly different. Moreover, since the intercepts were the points that differentiated the lexical tones, the discrimination around the crossover was enhanced.

Chen & Kager (2011) also performed a series of tasks to explore the perception of Mandarin lexical tones of twenty native Mandarin speakers (CN), twenty native Dutch speakers with no knowledge of Mandarin (NL), and seventeen native Dutch speakers who were advanced learners of Mandarin (AL). The Mandarin syllables /ma/ with Tone 2 and Tone 3 was recorded as the stimuli and they were arranged in a continuum along Tone 2 and Tone 3, with each continuum proceeding through eight steps from one endpoint (step 1) to the other (step 8). Test 1 was the forced choice identification task. In each trial, the participants who had knowledge of Mandarin (CN and AL) heard a single stimulus from step 1 to step 8 and had to identify it and choose between the character /ma2/ "hemp" and /ma3/ "horse" by pressing the corresponding button. On the other hand, the AXB identification paradigm was utilised for NL, who had no knowledge of Mandarin. A and B were the two endpoints (Tone 2 and Tone 3) in the two possible orders, and X varied from one endpoint to the other along the eight steps of the tonal continuum. In each trial the participants were presented with three stimuli, AXB, and they were asked to identify X as A or B by pressing the related key.

The results revealed a significant group effect. CN identified steps 1-5 as Tone 2 before abruptly shifting from step 4 to step 6, and steps 6-8 were labelled as Tone 3. On the other hand, NL demonstrated a smooth identification curve along the tonal continuum and did not show a significant difference across steps. As for AL, although they did not provide consistent responses for Tone 2 or Tone 3 around the endpoints as CN did, their performance showed a steeper slope than NL and shifted around step 6 just as CN did. Therefore, the authors concluded that, while CN demonstrated evidence of a categorical perception of lexical tones and NL perceived non-native lexical tones in a psychophysical way, AL were somewhere between these two groups. Therefore, although AL may have established categories of lexical tones, they were not as distinctive as those of CN.

Test 2 was a 2-step AX discrimination task. For the "same" pairs, A stimulus could have been any step along the tonal continuum (step 1 to step 8), and X stimulus was identical to it. For the "different" pairs, X differed from A stimulus with a two-step difference, and there were six possible combinations (step 1-step 3, step 2- step 4, step 3-step 5, step 4-step 6, step 5-step 7, and step 6-step 8). There were ascending orders (e.g. step 1-step 3) as well as descending orders (e.g. step 3-step 1) to examine the effect of Tone 3 Sandhi (Tone 3 Sandhi in Mandarin Chinese occurs when two adjacent syllables both have Tone 3 and the first syllable with Tone 3 is pronounced with Tone 2). In each trial the Mandarin participants were presented with two stimuli AX, and were asked to determine if these two stimuli were the "same" or "different" by pressing the corresponding button.

The results revealed that both step and order had a significant main effect on the percentage of correct responses of each individual group. CN had a higher percentage of correct responses in the decreasing order than in the increasing order, and the authors believed this was a result of the effect of Tone 3 Sandhi in Mandarin Chinese. While NL performed poorly along the tonal continuum (below the chance level), AL and CN had similar good results, suggesting that L2 learners were also able to construct a stable representation of lexical tones and process them in a native-like way.

Test 3 was a 2-step AXB discrimination task involving six possible A-B pairs (step 1-step 3, step 2- step 4, step 3-step 5, step 4-step 6, step 5-step 7, and step 6-step 8), and four possible AXB combinations (AAB, ABB, BAA, and BBA). The only difference between A and B was the step (two-step difference), and X was either identical to A or B. In each trial, the participants were presented with three stimuli, AXB, and they were asked to discriminate the stimuli and determine if X was identical to A or B by choosing the related key.

The results indicated that CN delivered the best performance of step 4-step 6 in the BAA combination around their category boundary crossover with a clear order. In addition, the discriminative curve of BAA was parallel to that of ABB, but the curve of ABB was less accurate due to the influence of Tone 3 Sandhi; on the other hand, NL's rate of accuracy could be simplified as BBA > AAB > BAA > ABB. The researchers suggested that it would be easier for speakers with no experience of tone languages to discriminate the triplets if the first two stimuli were identical due to the low demand of memory load; besides, it could be easier for them to discriminate if B preceded A and vice versa. As for AL, the discrimination curves of AAB and BBA were below the chance level, the performance of ABB combination was significantly better than those two, and the performance of BAA was even significantly better than the previous three combinations. Nevertheless, unlike CN, AL did not exhibit a discrimination peak around the category boundary crossover. These results revealed that CN was on the way toward categorising lexical tones and partly inhibited the psychophysical processing.

This study by Chen and Kager (2011) was enlightening because, in addition to Mandarin and Dutch-speaking participants, they included Dutch speakers with an advanced level of Mandarin. In this way, they were able to explore how speakers of a non-tone language, who perceived lexical tones in a psychophysical way, began to differentiate the lexical tones and perceive them categorically. Besides, as they indicated, there may be significant differences in the performances in ascending and descending orders or a combination of AAB, ABB, BAA, and BBA. As a result, it may be worth adopting these as independent variables in future studies. Lastly, Halle, Chang and Best (2004) tested the same participants with identification, as well as discrimination tasks, in order to compare the results of these two tasks to determine if speakers of a tone language were sensitive to cross-boundary differences and ignored within-category phonetic variations.

Yang and Liu (2006) also investigated the perception of Mandarin lexical tones of eight Mandarin monolinguals (mean age: 7.2 years), eight English monolinguals (mean age: 7 years), as well as eight Mandarin-English bilinguals (mean age: 6.9 years). The Mandarin monolinguals came from China and the English monolinguals and Mandarin-English bilinguals came from the US. While the English monolinguals had not been exposed to any tone languages, the Mandarin-English bilinguals used Mandarin as the primary language at home and attended Chinese schools in the US. All of these bilinguals had been born in the US and began to learn English after pre-school or kindergarten. The Mandarin syllable /ma/ was recorded with Tone 1, Tone 2, and Tone 4. The stimuli were in a continuum along Tone 1 and Tone 2 as well as Tone 1 and Tone 4, and each continuum proceeded through eleven steps from one endpoint (step 1) to the other (step 11). The test consisted of an ABX pseudo-identification task and an identification task. For the latter, A and B were the two endpoints in the two possible orders, and X varied from one endpoint to the other along the eleven steps of the tone continuum. In each trial the participants were presented with three stimuli, ABX, and they were tasked with identifying X as A or B by pressing the related key. The three-step discrimination task utilised a three-interval, two alternative-forced-choice (2AFC) paradigm (ABA or AAB). There were eight possible tone pairs (step 1- step 4, step 2- step 5, step 3- step 6, step 4-step 7, step 5-step 8, step 6-step 9, step 7-step 10, step 8-step 11). The participants heard three stimuli in each trial and they were asked to determine which of the last two sounds (A or B) was different from the first sound (A).

The results indicated that only the Mandarin and Mandarin-English bilinguals demonstrated a sigmoid shape of identification functions. In addition, the Mandarin participants had significantly steeper slopes around the category boundary than the Mandarin-English bilinguals and the English participants in both the Tone 1-Tone 2 and Tone 1-Tone 4 continua. In addition, modest peaks (although not prominent) around the tonal boundary were found in the discriminative curves in the Mandarin and Mandarin-

English groups. These results suggested that these two groups processed Mandarin lexical tones as linguistic categories, whereas the English group perceived them on a psychoacoustic basis. However, it was intriguing to note that the slope of the identification functions around the category crossover of the Mandarin-English bilinguals was not as sharp as that of the Mandarin monolinguals. The authors speculated that the amount of exposure to the tone language might affect the sensitivity of lexical tone perception. Besides, the exposure to non-tone languages may also influence bilinguals' perception of Mandarin lexical tones (von Hapsburg & Bahng, 2009).

Hoffmann et al. (2014) further explored the within-category variance and perception of lexical tones of ten native speakers of Mandarin Chinese (mean ages: 24 years) and eight native speakers of Dutch (mean age: 25 years). The disyllabic pseudo words /a1sa3/ and /a4sa3/ were recorded. The stimuli were in a continuum along these two disyllabic pseudo words, and each continuum proceeded through nine steps from one endpoint (step 1) to the other (step 9). Test 1 consisted of an identification task and an AX discrimination task for the Mandarin participants, as well as an AXB identification task and an AX task for the Dutch participants. These tasks all had a similar paradigm as that utilised by Halle, Chang, and Best (2004) and Chen and Kager (2011). The results indicated that the slopes of identification functions of the Mandarin participants (1.88) were marginally steeper than those of the Dutch participants (0.91), and that the intercept point of the Mandarin individuals (4.49) was significantly higher than that of the Dutch participants (3.57). In addition, the discriminative peak was indicated to be around step 3-step 5 for the Mandarin group, whereas the percentage of correct responses in the Dutch group remained flat along the tonal continuum. These results again confirmed that the Mandarin speakers perceived their native lexical tones categorically, while the Dutch speakers could only process the information of pitch contours in a psychophysical way.

The researchers further explored the within-category variances with an active oddball task in Test 2. The disyllabic pseudo word /asa/ was utilised with steps 1, 2, 3, 7, 8 and 9 because they were all clearly identified as examples of two endpoints by the Mandarin participants. Numerous examples were also created by changing the pitch of the first syllable in 8 Hz from -32 to +32 Hz. There were three levels of within-category variance, the first of which only contained a single stimulus close to the centre of each category. The second level contained five stimuli close to the centre of each category, and the third level consisted of thirty-three stimuli close to the centre of each category. Each block contained 480 stimuli, 408 of which were standard stimuli from the same category (e.g. Tone 1-Tone 3), while 72 were different stimuli from the other category (e.g. Tone 4-Tone 3). Each block only contained stimuli from the same level of within-category variance. There were 36 trials per block and the participants were tasked with determining if the last two stimuli were from the same category or not by pressing the corresponding button.

The results suggested that the within-category acoustic variance greatly affected the Dutch group's discrimination of lexical tones. The Dutch participants could accurately discriminate the phonetic differences in Tone 1 from those of Tone 4 at the first level, where there was only a single stimulus close to the centre of each category. However, the acoustic variance hindered their discrimination and their performance became worse and worse at the second and third levels. On the contrary, the Mandarin participants had no problem with discriminating the lexical tones at all the three levels, thereby indicating that they could ignore the within-category variance and mainly focus on the between-category differences.

Xu, Gandour, and Francis (2006) examined the categorical perception of pitch direction of thirty Mandarin (mean age: 27.5 years) and thirty English-speaking participants (mean age: 23.2 years) with speech, as well as non-speech stimuli. None of the English individuals had been exposed to any tone language. The Mandarin syllable /yi/ with Tone 1 and Tone 2 was recorded. The speech stimuli were on a continuum along Tone 1 and Tone 2, and each continuum proceeded through seven steps from one endpoint (step 1) to the other (step 7). As for the non-speech stimuli, they were harmonic tones exhibiting the identical pitch, amplitude, and duration as the speech stimuli. The stimulus type was designed as the between-subjects variable; thus, each participant would only encounter either speech stimuli or non-speech stimuli. The test consisted of an identification task as well as a discrimination task. In the identification task, the participants would hear stimuli along the continuum from step 1 to step 7 in each trial and they had to decide if this was a "level" pitch or a "rising" pitch by pressing the related button. Around 40% of the trials in the two-step AX discrimination task were the "same" pairs, with stimulus A and stimulus X being identical. As for the "different" pairs, there were five possible combinations (step 1-step 3, step 2-step 4, step 3-step 5, step 4-step 6, step 5-step 7) with ascending or descending order. The participants were tasked with determining if the two stimuli were the "same" or "different".

The results indicated that Mandarin speakers demonstrated sharper slopes than English participants for both speech and non-speech stimuli. As for the intercept points, both Mandarin and English participants had them around the exact centre (step 4), and yet the intercepts for the speech stimuli were slightly toward the endpoint of Tone 2 (step 4.25 for the Mandarin group and step 4.18 for the English). In addition, the Chinese participants performed between-category discrimination significantly better than the English participants in both speech and non-speech stimuli, and both groups had better between-category discrimination in the non-speech stimuli than the speech stimuli. On the other hand, the English participants performed significantly better than the Mandarin participants in within-category discrimination in both speech and non-speech stimuli, and yet again, both groups had better within-category discrimination in the non-speech stimuli than the speech stimuli.

Finally, Mandarin listeners exhibited a higher discrimination peak than their English counterparts around the category boundary of both stimulus types, and the discrimination peak of the English participants was higher for non-speech stimuli than speech stimuli. These results met the defined characteristics of categorical perception and confirmed that Mandarin speakers categorically perceive their native lexical tones. While the perception of between-categories was enhanced, the discrimination of the variation within-category was reduced to some degree. Moreover, the Mandarin participants not only perceived the lexical tones of speech stimuli in a categorical way, but also extended the perception to the pitch contours of non-speech stimuli. It is also important to note that it was easier for both language groups to discriminate non-speech stimuli than speech items, which is possibly due to the different complexity.

This study led to speculation that different kinds of stimuli may lead to various results. While non-speech stimuli were less complex and easier to perceive, speech stimuli were much more difficult to process. Furthermore, it may take extra effort to perceive and process certain speech stimuli, such as nonce words. Individuals may benefit from their real life experience when identifying and discriminating real words and they may actually have examples of certain words; in contrast, it was impossible for the participants to have examples of nonce words. The successful identification and discrimination of nonce words may depend on the abstract representation of lexical tones and the application of that information to the nonce words. As a result, it would be intriguing to explore how speakers of tone languages perceive the native lexical tones in nonce words.

How Mandarin-speaking children with ASD perceive their native lexical tones and whether they have categorical perception is, as yet, unknown. Souliere et al. (2007) explored the categorical perception of the visual stimuli of sixteen individuals with high-functioning autism (mean age: 18.6 years) and sixteen match controls (mean age: 17.1 years). The test consisted of a same-different discrimination task as well as a classification task of the same materials (ten ellipses varying on a wide continuum). The height of these ten ellipses (ellipse 1 to ellipse 10) all remained 5cm, whereas the width varied from 1.4 to 4.1cm with a constant increment of 0.3cm between ellipses.

The same-different discrimination task consisted of 32 “same” pairs, which presented two identical ellipses (from ellipse 2 to ellipse 9), and there were 36 “different” pairs, which showed two adjacent stimuli (e.g. ellipse 1 vs. ellipse 2). In each trial, the participants were presented with two visual stimuli simultaneously on the computer screen, and were tasked with determining if these two stimuli were the “same” or “different” by pressing the related key. For the classification task, the participants were firstly presented with ellipse 1 and ellipse 10 as “thin” and “wide” ellipses, respectively. Then, in each trial of the main test, the participants were presented with one stimulus at a time (from ellipse 2 to ellipse 9), and were tasked with classifying the stimulus as either a “thin” or “wide” ellipse as quickly as possible.

The results demonstrated that both the clinical and typically developing groups were sensitive to the difference in width across ellipses. Both groups demonstrated similar sigmoid response curves in the classification task. However, while the matched controls performed much better in the midpoint of the continuum (ellipse 5), there was no such enhanced discrimination peak around the category boundary exhibited by the ASD group. This phenomenon was similar to the perception of Mandarin lexical tones of the Dutch-speaking individuals with an advanced level of Mandarin as shown in the study of Chen and Kager (2011). This means that, although the clinical group may have been on the way toward categorising the visual stimuli, they might still be influenced by the psychophysical processing to some degree. This leads to the question of whether participants with ASD also have a typical categorical perception of auditory information such as pitch contours.

1.3 Motivation, hypothesis and research questions

Since the way in which Mandarin-speaking children with ASD perceive their native lexical tones remains unknown, the aim of this experiment is to explore the perception of lexical tones of Mandarin participants with ASD. It is hypothesised that Mandarin participants with ASD, like their typically-developing counterparts, may exhibit a quasi-categorical perception of lexical tones if they associate acoustic pitch contours and linguistically related tonal categories. However, either because of their deficient categorisation demonstrated in the visual domain or their delayed or different language development, it is possible that they may demonstrate some diverse or delayed patterns than their controls. In addition, different kinds of stimuli will be utilised, such as real words, nonce words and non-speech pure tones in order to explore the effect of these different stimuli on the perception of lexical tones.

The aim of this project is to examine the link between the perception of lexical tones and native languages in typical and atypical development to better understand the contribution of auditory mechanisms to native languages. Since the ASD population is a particularly interesting group with which to test this link, a cross-linguistic perspective is adopted to investigate typically-developing and ASD individuals in the UK and Taiwan, in order to unravel the universal and specific aspects of language development in these two trajectories. The research questions are presented below.

Categorical perception of lexical tones

Q1: Do Mandarin-speaking children with autism spectrum disorder (ASD) categorically perceive and identify lexical tones in the same way as their typically-developing (TD) Mandarin-speaking counterparts, or do Mandarin-speaking ASD children perceive lexical tones in a psychophysical way?

Q1a: Are the intercept points around the same place for both Mandarin typically-developing and ASD participants?

Q1b: Are the slopes at the category boundary significantly different for Mandarin typically-developing and ASD participants?

Q1c: Do both Mandarin typically-developing and ASD participants have an enhanced perception of the category boundary?

Role of native language

Q2: According to the literature, speakers of a tone language have a categorical perception of their native lexical tones, whereas speakers of a non-tone language perceive and process their lexical tones in a psychophysical way. Do Mandarin- and English-speaking participants perceive lexical tones differently, as suggested in the literature?

Q2a: Are intercept points around the exact centre of the tonal continuum or at a different point for Mandarin and English participants?

Q2b: Are the slopes at the category boundary significantly different for Mandarin and English participants?

Q2c: Do both Mandarin and English participants have an enhanced perception of the category boundary?

Role of stimulus type (real word, nonce word, pure tone)

Q3: Do ASD children and their controls discriminate the pitch contours of different kinds of stimuli (real words, nonce words, and pure tones) differently? Is there any interaction between the group and type of stimulus?

Q3a: Do the different groups of participants categorically perceive nonce words equally strongly as real words? Do they have the same intercept points, the same slopes at the intercept points, and the same enhanced perception at the category boundary of nonce word and real word stimuli?

Q3b: Do the different groups of participants categorically perceive pure tones equally strongly as real words? Do they have the same intercept points, the same slopes at the intercept points, and the same enhanced perception at the category boundary of pure tone and real word stimuli?

2 Method of Naming task (Forced choice identification task)

2.1 Participants

The same Mandarin-speaking and English-speaking participants as in the previous experiment participated in the current task.

2.2 Task 1: Forced choice identification task

In this forced choice identification task, all the Mandarin participants heard one stimulus per trial and were tasked with identifying the stimulus as either Tone 1 or Tone 4. There were three sub-tests: Mandarin real words, nonce words, and pure tones. Meanwhile, the English participants were given an AXB identification task, in which they heard three stimuli per trial and were asked to determine if X was more similar to A or B. There was only one sub-test: Real English words.

2.3 Stimuli

The tone identification task for the Mandarin participants contained three kinds of stimuli: real words, nonce words, and pure tones. All the speech stimuli were on a continuum along Tone 1 and Tone 4. Three Mandarin syllables were selected for the real word stimuli: /jie/, /shu/, and /ya/ having considered the comparable imageability, frequency, syllable combination frequency, and onset frequency; therefore, three tonal linguistic contrasts were used in the task: /jie1/ “street” - /jie4/ “to borrow”, /shu1/ “book” - /shu4/ “tree”, and /ya1/ “duck” - /ya4/ “surprise”. According to the British National Corpus (BNC), a 100 million word collection of samples of written and spoken language from a wide range of sources of British English from the late twentieth century, the word frequency of /jie1/ "street" is 0.019%, the word frequency of /jie4/ "to borrow" is 0.001%, the word frequency of /shu1/ "book" is 0.024%, the word frequency of /shu4/ "tree" is 0.006%, the word frequency of /ya1/ "duck" is 0.001%, and the word frequency of /ya4/ "surprise" is 0.005%. As for the nonce word stimuli, /chei/, /tiu/ and /fi/ was utilised for a number of reasons. Firstly, these three syllables are all possible, valid, legible sound combinations in Mandarin. Secondly, these three syllables are nonce words with all the four lexical tones in Mandarin. Thirdly, their sound structures are comparable to the real words /jie/, /shu/, and /ya/. The 12 syllables (/jie1/, /jie4/, /shu1/, /shu4/, /ya1/, /ya4/, /chei1/, /chei4/, /tiu1/, /tiu4/, /fi1/ and /fi4/) were produced by a native Mandarin-speaking female, and the recordings were pre-tested for comprehension and tone identification by native Mandarin-speaking adults to ensure that the pronunciation was articulate and the tones were clear and understandable. As for the pure tone stimuli, they were created by removing segmented information from the real word stimuli words /jie/, /shu/, and /ya/ by PRAAT with the assistance of professional phoneticians. The result was a set of pure tone contours in a human voice. Then, in order to create a continuum along Tone 1 and Tone 4, the F0-range between tonal contrasts was manipulated in PRAAT and divided into 8 equidistant contours falling between Tone 1 and Tone 4. Therefore, 8 stimuli were created per word, making 24 real word stimuli for the three words (8 steps x 3 words), 24 nonce word stimuli, and 24 pure tone stimuli.

There were three sub-tests according to the kinds of stimuli: real words, nonce words, and pure tones. The stimuli were played twice in pseudo random order in each sub-test, making 48 trials for each participant. The order was designed so that there was no more than two identical words after each other and more than one identical tone step. In addition, half of the participants had T1-T4 order and the other half had T4-T1 order. This meant that half of the participants were shown Tone 1 on the left of the screen and Tone 4 on the right, while the other half were shown Tone 4 on the left and Tone 1 on the right. A total of 8 orders were created. The cartoon *Dino* was shown on the computer screen to make the experiments more interesting to young children. The Dino blinked when a sound was played. All the tasks were run using Mammoth software.

As for the AXB identification for the English participants, there was only one kind of stimulus: real English words. All the speech stimuli were in a continuum along Tone 1 and Tone 4. Three monosyllabic English words, “bowl”, “chain”, and “leaf”, were selected as the real English word stimuli because of their comparable imageability as well as their frequency based on the British National Corpus. These three words were produced by a native English-speaking female, and the comprehension and tone identification in the recordings were pre-tested by native English-speaking adults to ensure the articulation of the pronunciation. Then a set of pitch contours from the Mandarin stimuli were applied to the segmented information of these English words by PRAAT and, in order to create a continuum along Tone 1 and Tone 4, the F0-range between tonal contrasts was manipulated in PRAAT and divided into 8 equidistant contours falling between Tone 1 and Tone 4. Therefore, 8 stimuli were created along the tonal continuum (from step 0 to step 7), and there was a total of 24 real word stimuli (8 steps x 3 real English words). As for the AXB identification task, the three stimuli were the same real English words. A and B were the two endpoints in the two possible orders (Tone 1-Tone 4 or Tone 4-Tone 1), and X varied from one endpoint to the other along the eight steps of the tonal continuum. Care was taken to ensure that all the 24 real word stimuli were played once as X in a pseudo random order. In addition, there were no more than two identical words after each other, and there was no more than one identical tone step after each other. Half of the participants had Tone 1-Tone 4 order and the other half had Tone 4-Tone 1 order. All the tasks were run using Mammoth software.

2.4 Procedure

Every Mandarin participant was presented with three sub-tests: real words, nonce words, and pure tones, in a random order. Each sub-test consisted of two phases: Practice phase and Main phase. There were three trials in the practice phase, and the three acoustic examples were played once in a random order. One example was a stimulus with Tone 1 (step 0), one with Tone 4 (step 7), and one example was a continuum (step 4) between Tone 1 and Tone 4. A cartoon dinosaur shown on a computer monitor uttered an acoustic stimulus in order to make the experiment more interesting and attract the attention of young or clinical participants, who were then tasked with naming the stimulus either with Tone 1 or Tone 4 by pressing the corresponding key. The participants with Tone 1-Tone 4 order were asked to press key Q if they considered the sound to be Tone 1, and key P if they thought it was Tone 4 and vice versa for the participants with Tone 4-Tone 1 order. If they hesitated to name the stimulus Tone 1 or Tone 4, they were encouraged to make a guess based on their instinct. There were 48 stimuli for each sub-test in the main phase and no feedback was given. The next trial was presented after each click.

Every English participant was presented with an AXB identification task with real English words. The practice phase contained three trials, in each of which the participants heard three stimuli with the same real English word in various pitch contours. The task for the participants was to identify X as A or B by pressing key Q or key P, respectively. Since Tone 1-Tone 4 and Tone 4-Tone 1 orders were counterbalanced, A always represented the words with the level pitch contour (Tone 1) for half of the participants and B always indicated the words with a falling pitch contour (Tone 4), and vice versa for the other half of the participants. The main phase contained 24 trials in total and no feedback was given. The next trial was presented after each click.

This AXB identification task design helped to explore the identification of lexical tones, even of young Mandarin-speaking children, who did not fully understand lexical tones or even speakers of a non-tone language who had no concept of lexical tones. Since Mandarin KTD were still young and had not fully grasped the labels of lexical tones, they participated in the naming task as well as the AXB identification task.

3 Results of naming task

3.1 Mandarin results across stimulus types

The response data for each participant (Mandarin ASD, Mandarin YTD, and Mandarin OTD) on the computerized categorical perception naming task were imported into SPSS for analysis. The data of Mandarin KTD is not included in this section, since they were only tested on the real word stimuli. As Figure 17 and Table 27 show, the overall participants tend to distinguish the two tonal targets at both endpoints of the continuum. At step 0 and 1, overall participants could easily identify the target with only 8% and 13% of Tone 4 responses. Then the participants may face difficulty in distinguishing tone at step 2 for that their response was around the chance level (45%**).** Yet from step 3 onwards, the proportions of T4 responses rise rapidly and show consistent variability. Each one was more than 75% of T4 responses at the right-hand part of the continuum, manifesting that the participants were more likely to identify the stimuli as Tone 4 rather than Tone 1 from step 3 onwards. The rates are 75% at step 3 and 83% at step 4. Then the rate reaches a peak and is stabilized from step 5 onwards. The rates are 89% at step 5, step 6, and step 7. Just as the literature found, the categorical perception in tones is askew along the continuum. That is, the chance level (50%) between Tone 1 and Tone 4 does not lie on the middle of the continuum (step 3 or step 4), but on the step 2 in most of the cases. This manifested that between the flat and high tone contour of Tone 1 and the falling tone contour of Tone 4, as soon as the tone has the obvious falling contour, native speakers of Mandarin Chinese would incline to label it as Tone 4 rather than Tone 1 even if this falling tone contour is not as steep as the usual Tone 4. This suggests the categorical perception of tones for tone language speakers. Since it is impossible for human beings to utter Tone 4 with the exactly same pitch and slope of falling, it is crucial and essential for us to categorize the tone contours with similar falling features as Tone 4 as a group.

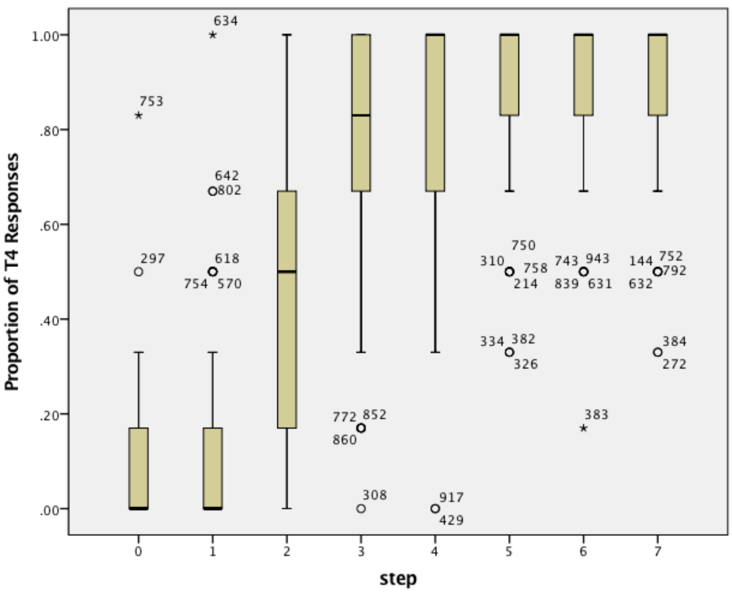

Figure 17. Mean proportion of T4 responses for overall Mandarin participants across stimulus types

		Step 0	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Mandarin ASD (n = 21)		0.10	0.13	0.42	0.72	0.81	0.89	0.89	0.90
	Mandarin NLP (n = 7)	0.08	0.15	0.40	0.73	0.80	0.94	0.85	0.90
	Mandarin SLP (n = 9)	0.12	0.09	0.30	0.60	0.73	0.82	0.85	0.86
	Mandarin YNLP (n =5)	0.09	0.18	0.68	0.90	0.96	0.97	0.98	0.97
Mandarin YTD (n = 10)		0.06	0.17	0.57	0.80	0.84	0.88	0.84	0.83
Mandarin OTD (n = 9)		0.06	0.09	0.40	0.79	0.88	0.92	0.94	0.92
Mandarin Overall (n = 40)		0.08	0.13	0.45	0.75	0.83	0.89	0.89	0.89

Table 27. Mean proportion of T4 responses across stimuli

A mixed ANOVA was performed with the within-subject factor Step (step

0, step 1, step 2, step 3, step 4, step 5, step 6, step 7) and the between-subject factor Group (Mandarin ASD, Mandarin YTD, Mandarin OTD) on the dependent variable, proportion of T4 responses, for the data of three stimulus types. There was no effect of Group $F(2, 37) = .284$, $p = .754$, $\eta^2 = .015$. Bonferroni post-hoc tests indicated that the mean proportion of T4 responses in Mandarin OTD was not significantly higher than in Mandarin YTD ($p = 1.000$) nor Mandarin ASD ($p = 1.000$), and that the mean proportion of T4 responses in Mandarin YTD was not significantly higher than in Mandarin ASD ($p = 1.000$). There was a main effect of the Step $F(7, 259) = 264.322$, $p < .001$, $\eta^2 = .877$. Bonferroni post-hoc test indicated that step 0 was significantly different from other steps (all *p*s < .005) except step 1 ($p = .464$), showing that step 0 and step 1 could be considered as being within one category. As for the confounding step 2, it was significantly different from all other steps (all *p*s < .001). Then step 3 was significantly different from other steps (all *p*s ≤ .05) and marginally different from step 4 ($p = .078$). Moreover, while step 4 was significantly different from step 0, 1, and 2 (all *p*s < .001), it was not significantly different from step 3, 5, 6, and 7 ($p = .078$, $p = .257$, $p = .493$, $p = 1.000$, respectively). In addition, there was a significant interaction between Step and Group $F(14, 259) = 1.932$, $p = .024$, $\eta^2 = .095$. However, Bonferroni post-hoc tests did not indicate any significant difference between the three groups along the steps.

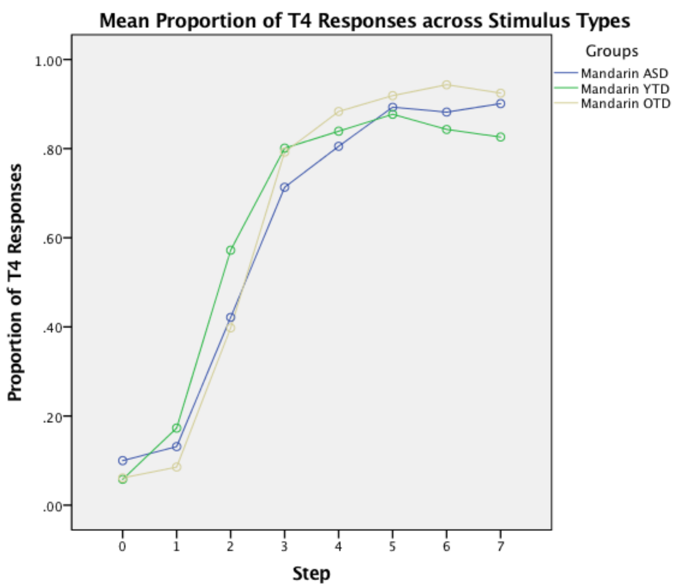

Figure 18. Mean proportion of T4 responses across stimulus types for ASD, Mandarin YTD, and Mandarin OTD

Since the data of ASD group are heterogeneous as usual, we once again split it into Mandarin NLP, Mandarin SLP, and Mandarin YNLP, and thus have five groups to explore. A mixed ANOVA was conducted with the within-subject factor Step (step 0, step 1, step 2, step 3, step 4, step 5, step 6, step 7) and the between-subject factor Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the dependent variable, proportion of T4 responses, for the data of three stimulus types. There was a major effect of Group $F(4, 35) = 5.896$, $p = .001$, $\eta^2 = .403$. Bonferroni post-hoc tests indicated that the percentage of T4 responses in Mandarin YNLP was significantly higher than Mandarin SLP ($p < .001$), and was marginally higher than that of in Mandarin NLP ($p = .070$). There was also a main effect of Step $F(7, 245) = 298.715$, $p < .001$, $\eta^2 = .895$. Bonferroni post-hoc tests unsurprisingly demonstrated the similar results as the previous analysis on the three groups. That is, step 0 and 1 could be considered being within one category, whereas step 3, 4, 5, 6, and 7 were basically categorized together. In addition, there was a significant interaction between the within-subject factor, Step, and the between-subject factor, Group $F(28, 245) = 2.011$, $p = .003$, $\eta^2 = .187$. Bonferroni post-hoc tests indicated that at step 2, the percentage of T4 responses in Mandarin SLP was significantly lower than that of in Mandarin YNLP ($p = .007$) and Mandarin YTD ($p = .024$). At step 3, the percentage of T4 responses in Mandarin SLP was significantly lower than that of in Mandarin YNLP ($p = .007$) and Mandarin YTD ($p = .043$), and marginally lower than that of in Mandarin OTD ($p = .074$). At step 4, the percentage of T4 responses in Mandarin SLP was significantly lower than that of in Mandarin YNLP ($p = .034$).

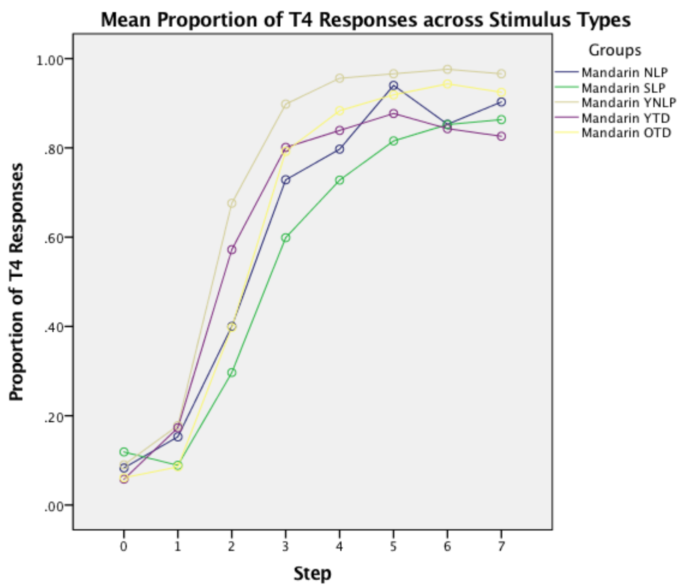

Figure 19. Mean proportion of T4 responses across stimulus types for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, and Mandarin OTD

3.2 Mandarin results for the three stimulus types

3.2.1 Interaction between Stimulus and Step

After examining the proportion of T4 responses across stimulus types, now we would like to explore the results for each stimulus type. A two-way ANOVA was conducted with the within-subject factors Step (step 0 to step 7) and Stimulus (real word, nonce word, pure tone) on the data of all Mandarin groups. There was a main effect of Stimulus $F(2, 688) = 19.945$, $p < .001$. Bonferroni post-hoc tests revealed that the three kinds of stimulus type (real word, nonce word, and pure tone) are statistically significantly different from each other (all *p*s < .05). Table 28 and Figure 20 may give us a clearer idea for their differences. Generally speaking, our participants tend to have higher proportion of T4 responses for the pure tones, then the real words, and lower proportion of T4 responses for the nonce words. It is worthwhile to note that, however, the proportion of T4 responses for the nonce words is higher than the pure tones and the real words at step 0, and the proportion of T4 responses for the nonce words is higher than the real words at step 1. Overall, it is easier to perceive and discriminate the pure tones in the categorical perception naming test, since the participants could focus on the auditory pitch without the distraction of linguistic information.

As for the real words and nonce words, although they both provide auditory pitch as well as linguistic information for discriminating the tones, it turns out that the real words are more helpful in the way that the participants have already learnt and categorized the real words in their lexicon.

There was a main effect of Step $F(7, 688) = 294.874$, $p < .001$. Bonferroni post-hoc tests revealed that that step 0 and step 1 were significantly different from other steps (all *p*s < .001) except for each other ($p = 1.000$). As for the step 2, it is significantly different from all other steps (all *p*s < .001). Further, while step 3 is significantly different from step 0, 1, 2, 5 and 6 (all *p*s < .021), it is not significantly different from step 4 and 7 ($p = 1.000$, $p = .052$, respectively). Step 4 was significantly different from step 0, 1, and 2 (*p*s < .001), and yet it was not significantly different from step 3, 5, 6, and 7 (*p*s = 1.00). As a consequence, step 0 and 1 could be considered being within one category, whereas step 3, 4, 5, 6, and 7 could be categorized together. There was no significant interaction found between Stimulus and Step $F(14, 688) = 1.477$, $p = .114$.

	Step 0	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Real word	0.05	0.09	0.48	0.81	0.88	0.86	0.92	0.87
Nonce word	0.10	0.11	0.38	0.71	0.77	0.85	0.85	0.84
Pure tone	0.08	0.20	0.62	0.87	0.90	0.96	0.93	0.94

Table 28. Mean proportion of T4 responses across groups for each stimulus types

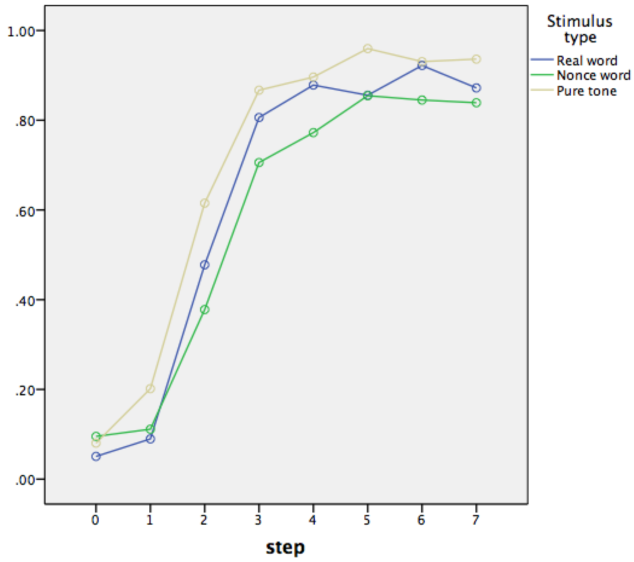

Figure 20. Mean proportion of T4 responses for each stimulus types across groups

3.2.2 Interaction between Stimulus and Group

A mixed ANOVA was conducted with a within-subject factor Stimulus (real word, nonce word, pure tone) and a between-subject factor Group (Mandarin ASD, Mandarin YTD, Mandarin OTD). There was no main effect of Group $F(2, 35) = .242$, $p = .786$, $\eta^2 = .014$. There was a main effect of Stimulus $F(2, 70) = 14.196$, $p < .001$, $\eta^2 = .289$. Bonferroni post-hoc tests indicated that the percentage of T4 responses in pure tones was significantly higher than that of in real words ($p = .013$), and the proportion of T4 responses in real words was significantly higher than that of in nonce words ($p = .024$). There was a marginal interaction between the within-subject factor, Stimulus, and the between-subject factor, Group $F(4, 70) = 2.249$, $p = .072$, $\eta^2 = .114$. Bonferroni post-hoc tests indicated that for Mandarin YTD, the percentage of T4 responses in pure tones was significantly higher than in real words ($p = .008$) and in nonce words ($p < .001$). In addition, the percentage of T4 responses in pure tones in Mandarin YTD was marginally higher than that of in Mandarin ASD ($p = .052$).

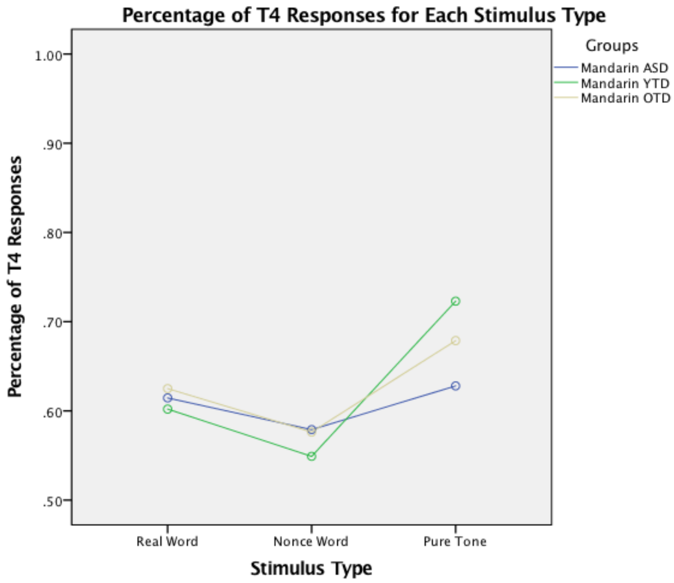

Figure 21. Percentage of T4 responses for each stimulus type for Mandarin ASD, Mandarin YTD, and Mandarin OTD

In addition, the ASD group was split again to run a mixed ANOVA with a within-subject factor Stimulus (real word, nonce word, pure tone) and a between-subject factor Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD). There was a major effect of Group $F(4, 33) = 5.580$, $p = .002$, $\eta^2 = .403$. Bonferroni post-hoc tests indicated that the percentage of T4 responses in Mandarin SLP was significantly lower than that of in Mandarin YNLP ($p = .001$). There was a main effect of Stimulus $F(2, 66) = 8.174$, $p = .001$, $\eta^2 = .199$. Bonferroni post-hoc tests indicated that the percentage of T4 responses in nonce words was significantly lower than that of in pure tones ($p = .004$), and marginally lower than that of in real words ($p = .053$). There was no interaction between the within-subject factor, Stimulus, and the between-subject factor, Group $F(8, 66) = 1.220$, $p = .302$, $\eta^2 =.129$.

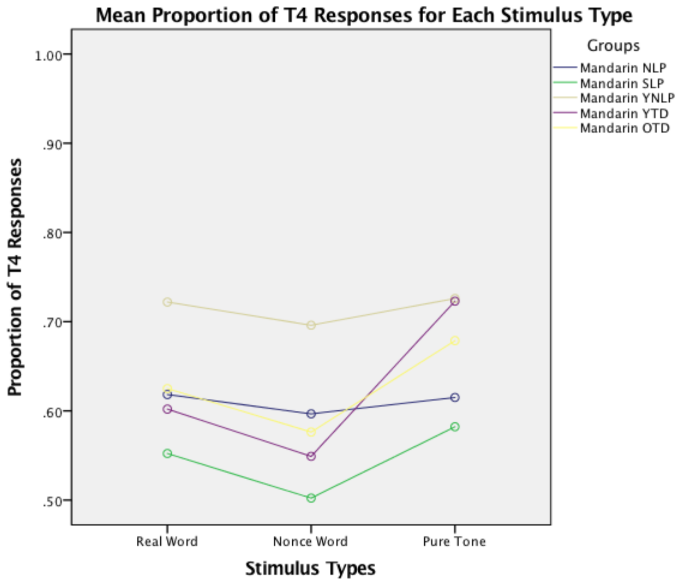

Figure 22. Percentage of T4 responses for each stimulus type for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, and Mandarin OTD

3.3 Mandarin and English results for real word

After examine the data of Mandarin participants for the three stimulus types, now we would include the data of Mandarin KTD as well as English participants and only focus on the real word conditions.

		Step 0	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Mandarin ASD (n = 21)		0.06	0.11	0.47	0.75	0.86	0.87	0.88	0.89
	Mandarin NLP (n = 7)	0.10	0.12	0.48	0.76	0.86	0.90	0.83	0.86
	Mandarin SLP (n = 9)	0.07	0.06	0.28	0.67	0.78	0.81	0.87	0.87
	Mandarin YNLP (n = 5)	0.00	0.20	0.80	0.90	1.00	0.93	0.97	0.97
Mandarin YTD (n = 10)		0.05	0.08	0.47	.082	0.85	0.85	0.87	0.82
Mandarin OTD (n = 9)		0.04	0.02	0.46	0.80	0.91	0.89	0.96	0.93
Mandarin KTD (n = 18)		0.15	0.17	0.37	0.67	0.82	0.76	0.78	0.80
Mandarin		0.08	0.11	0.44	0.74	0.85	0.84	0.86	0.85

Overall (n = 58)									
English ASD (n = 15)		0.11	0.24	0.48	0.73	0.76	0.82	0.91	0.91
	English NLP (n = 8)	0.00	0.21	0.46	0.71	0.84	0.83	0.88	0.88
	English SLP (n = 7)	0.24	0.28	0.50	0.76	0.67	0.81	0.95	0.95
English YTD (n= 11)		0.24	0.15	0.55	0.64	0.76	0.76	0.73	0.82
English OTD (n = 5)		0.13	0.13	0.73	0.93	1.00	1.00	1.00	0.93
English Overall (n = 31)		0.16	0.19	0.54	0.73	0.80	0.83	0.86	0.88
Overall (n = 89)		0.11	0.14	0.47	0.74	0.83	0.83	0.86	0.86

Table 29. Mean proportion of T4 responses for real words for Mandarin as well as English participants

All the data for real words was subjected to a mixed ANOVA with one within-subject factor, Step (step 0 to step 7), and one between-subject factor, Group (Mandarin participants, English participants) with the dependent variable, the proportion of T4 responses. Group ($F(1, 87) = 1.591$, $p = .211$, $\eta^2 = .018$) had no major effect, but Step ($F(7, 609) = 191.896$, $p < .001$, $\eta^2 = .688$) did have a major effect. Bonferroni post-hoc tests unsurprisingly demonstrated the similar results as the previous analysis on the three groups. That is, step 0 and 1 could be considered being within one category, whereas step 3, 4, 5, 6, and 7 were basically categorized together. There was no statistically significant interaction between the within-subject factor Step and the one between-subject factor Group ($F(7, 609) = 1.495$, $p = .166$, $\eta^2 = .017$).

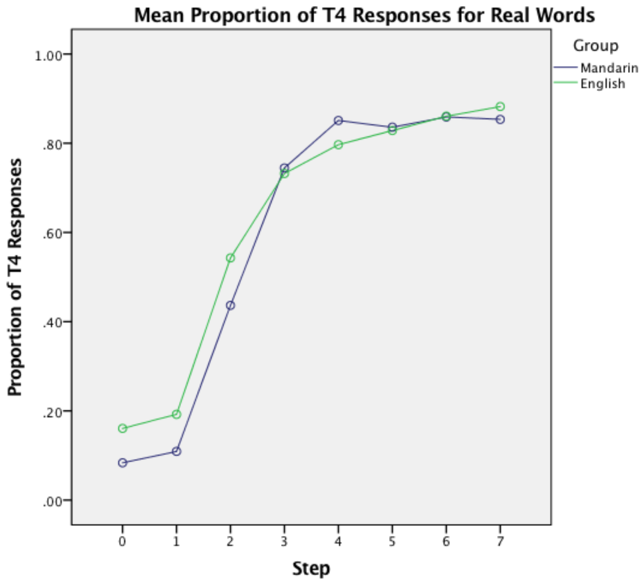

Figure 23. Mean proportion of T4 responses for real word for Mandarin and English participants

All the data for real words was subjected to a mixed ANOVA with one within-subject factor, Step (step 0 to step 7), and one between-subject factor, Group (Mandarin ASD, English ASD) with the dependent variable, the proportion of T4 responses. Group ($F(1, 34) = 0.87$, $p = .770$, $\eta^2 = .003$) had no major effect, but Step ($F(7, 238) = 100.617$, $p < .001$, $\eta^2 = .747$) did have a major effect. Bonferroni post-hoc tests unsurprisingly demonstrated the similar results as the previous analysis on the three groups. That is, step 0 and 1 could be considered being within one category, whereas step 3, 4, 5, 6, and 7 were basically categorized together. There was no statistically significant interaction between the within-subject factor Step and the one between-subject factor Group ($F(7, 238) = 1.098$, $p = .365$, $\eta^2 = .031$).

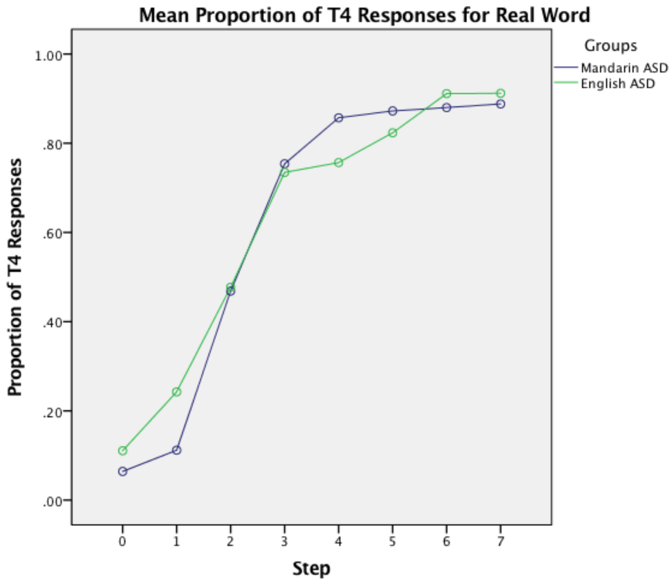

Figure 24. Mean proportion of T4 responses for real word for Mandarin ASD and English ASD

All the data for real words was subjected to a mixed ANOVA with one within-subject factor, Step (step 0 to step 7), and one between-subject factor, Group (Mandarin TD, English TD) with the dependent variable, the proportion of T4 responses. Group ($F(1, 51) = 1.588$, $p = .213$, $\eta^2 = .030$) had no major effect, but Step ($F(7, 357) = 93.844$, $p < .001$, $\eta^2 = .648$) did have a major effect. Bonferroni post-hoc tests unsurprisingly demonstrated the similar results as the previous analysis on the three groups. That is, step 0 and 1 could be considered being within one category, whereas step 3, 4, 5, 6, and 7 were basically categorized together. There was no statistically significant interaction between the within-subject factor Step and the one between-subject factor Group ($F(7, 357) = 1.335$, $p = .233$, $\eta^2 = .026$).

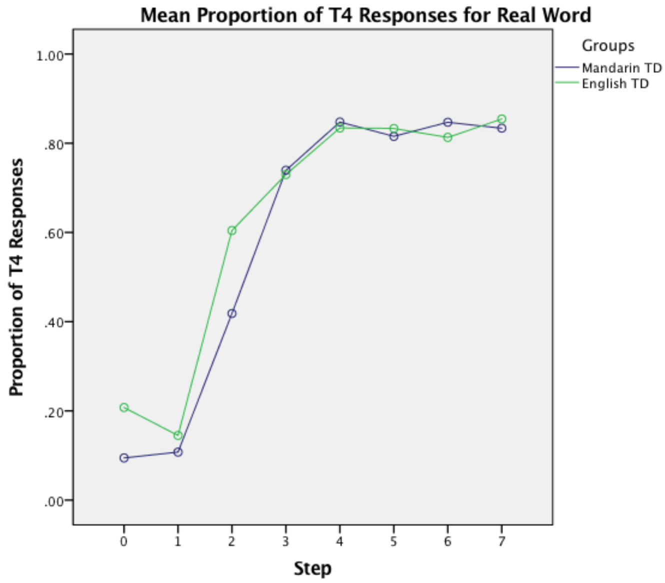

Figure 25. Mean proportion of T4 responses for real word for Mandarin TD and English TD

3.4 Intercept points

The intercept points were calculated based on the linear functions of individual of identification curves. It was the point at the chance level, where 50% of the responses identifying the stimuli as Tone 1, and 50% of the responses were naming the stimuli as Tone 4. For example, if a participant had 50% of T4 responses at step 2, then the intercept point for this participant would be 2. If a participant had 40% of T4 responses at step 2 and 60% of T4 responses at step 3, then the intercept point for this individual would be 2.5.

A mixed ANOVA was performed with one within-subject factor, Stimulus (Real word, Nonce word, Pure tone), and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the dependent variable, Intercept. (The data of Mandarin KTD is not included because they were only tested on the real word stimuli.) There was a marginal effect of Group $F(4, 32) = 2.423$, $p = .068$, $\eta^2 = .232$, but Bonferroni post-hoc tests did not reveal any significant difference between the five groups. There was a main effect of Stimulus $F(2, 64) = 8.905$, $p < .001$, $\eta^2 = .218$. Bonferroni post-hoc tests indicated that the intercept of nonce words was significantly higher than that of real words (p = .015) and pure tones (p = .002). There was no interaction between the within-subject factor, Stimulus, and the between-subject factor, Group $F(8, 64) = .861, p = .554, \eta^2 = .097$.

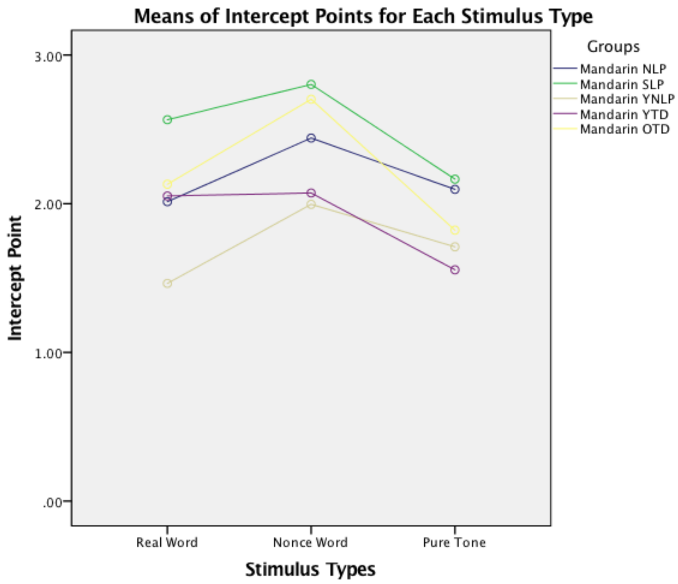

Figure 26. Means of intercept points for each stimulus type for four Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD)

Now we would like to focus on the analysis of real words, so that the data of Mandarin KTD as well as English participants could also be included in the following seven tests on the intercept points.

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD, Mandarin KTD) on the dependent variable, Intercept, in real word conditions. The results revealed a main effect of Group $F(5, 52) = 3.585$, p = .007. Bonferroni post-hoc tests revealed the intercept of Mandarin YNLP was significantly lower than that of Mandarin SLP (p = .005) and Mandarin KTD (p = .032)

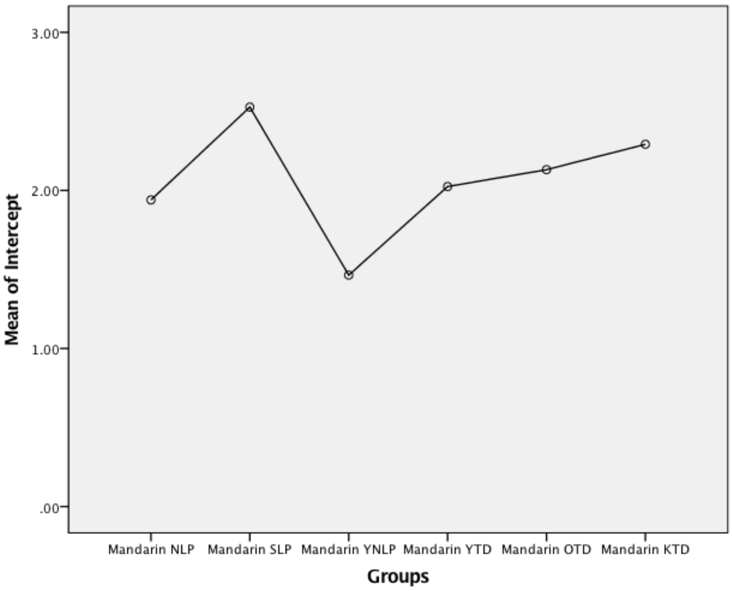

Figure 27. Means of intercept points for real word for five Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD, Mandarin KTD)

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (Mandarin ASD, Mandarin TD) on the dependent variable, Intercept, in real word conditions. The results did not reveal a main effect of Group $F(1, 56) = .744$, $p = .392$.

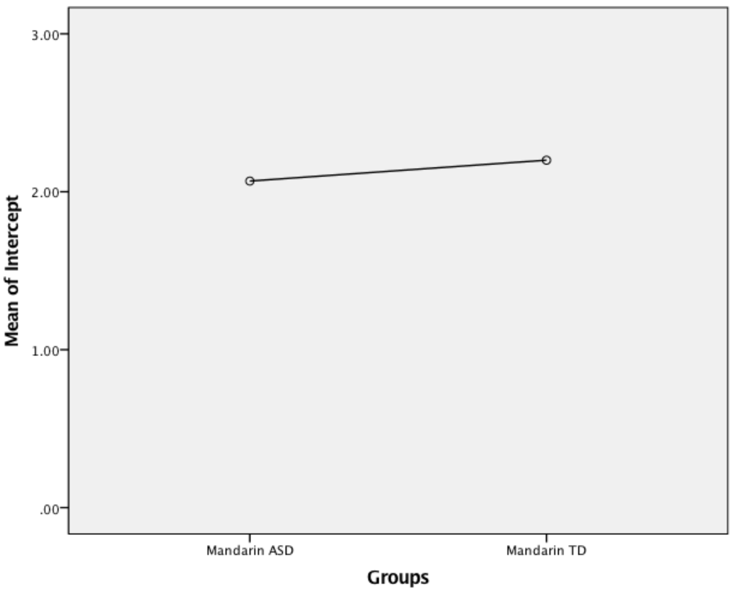

Figure 28. Means of intercept points for real word for Mandarin ASD and Mandarin TD

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (English NLP, English SLP, English YTD, English

OTD) on the dependent variable, Intercept, in real word conditions. There was no main effect of Group $F(3, 28) = 1.110$, $p = .362$.

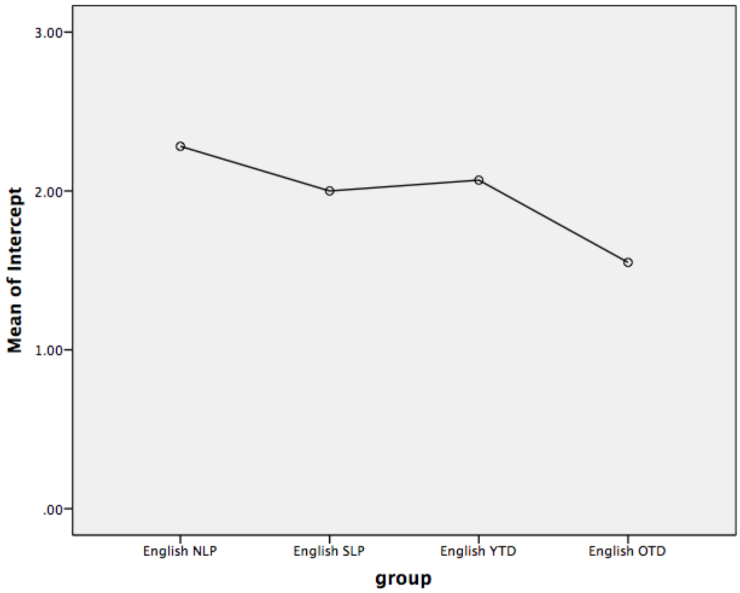

Figure 29. Means of intercept points for real word for four English groups (English NLP, English SLP, English YTD, English OTD)

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (English ASD, English TD) on the dependent variable, Intercept, in real word conditions. There was no main effect of Group $F(1, 30) = .859$, $p = .362$.

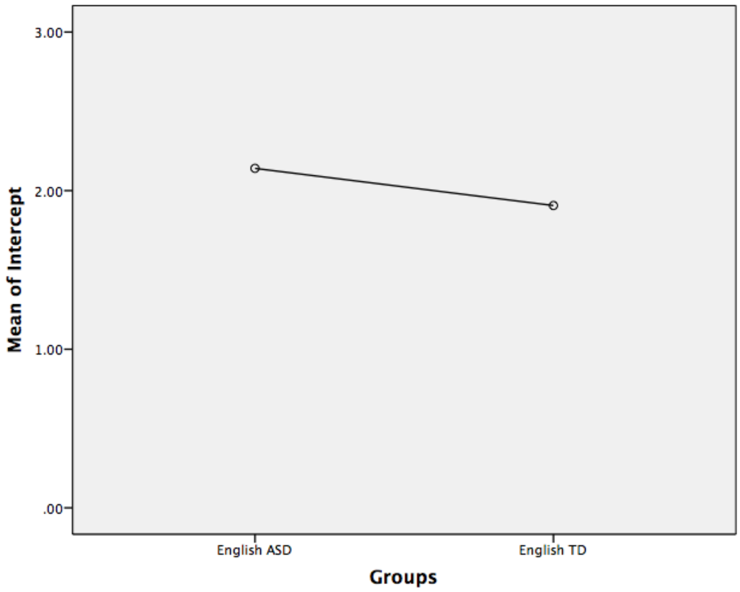

Figure 30. Means of intercept points for real word for English ASD and English TD

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (Mandarin ASD, Mandarin TD, English ASD, English TD) on the dependent variable, Intercept, in real word conditions. There was no main effect of Group F(3, 86) = .881, *p* = .454.

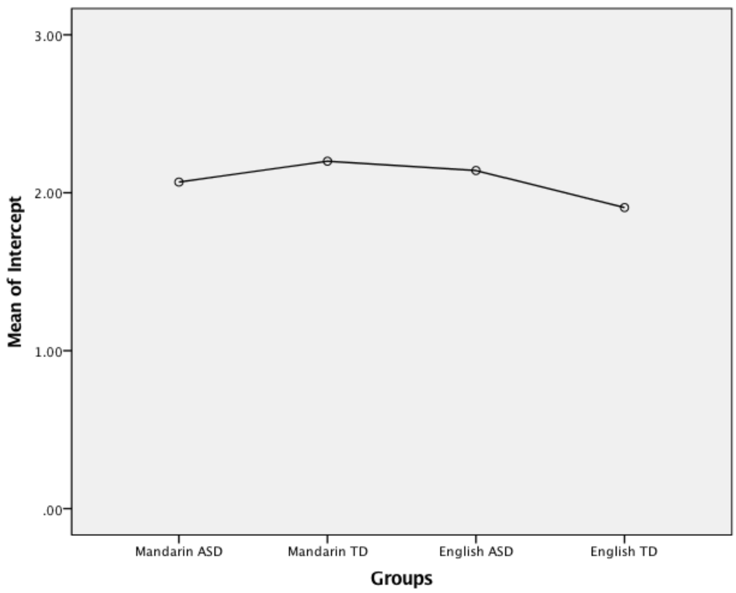

Figure 31. Means of intercept points for real word for Mandarin ASD, Mandarin TD, English ASD, and English TD

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (Mandarin ASD, English ASD) on the dependent variable, Intercept, in real word conditions. There was no main effect of Group $F(1, 35) = .097$, $p = .757$.

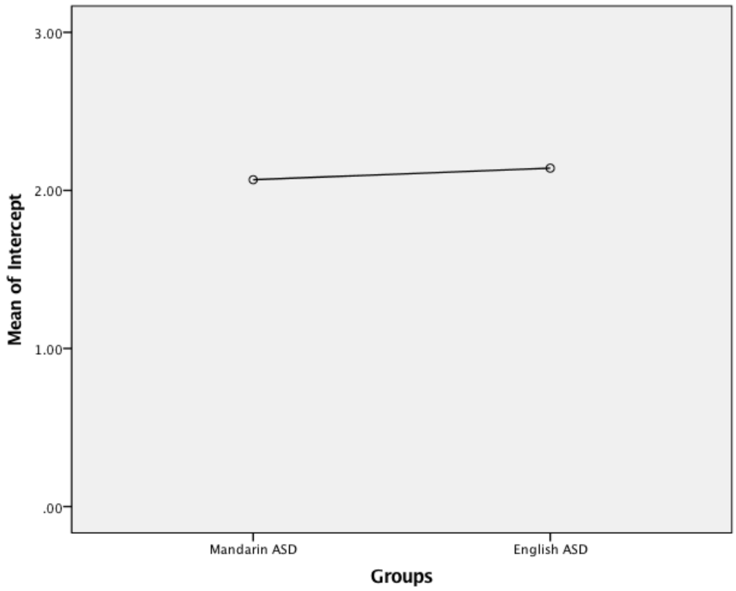

Figure 32. Means of intercept points for real word for Mandarin ASD and English ASD

A one-way ANOVA was performed to explore the effect of the between-subject factor, Group (Mandarin TD, English TD) on the dependent variable, Intercept, in real word conditions. There was a marginal effect of Group $F(1, 51) = 3.154$, $p = .082$.

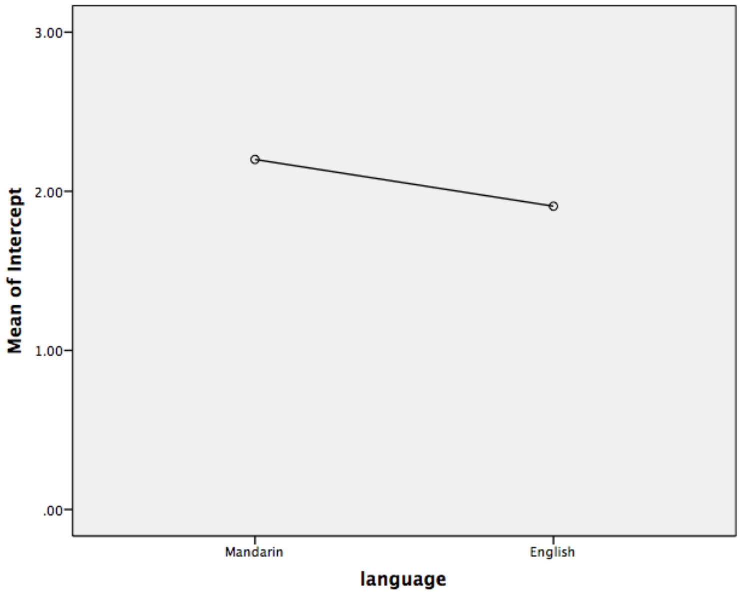

Figure 33. Means of intercept points for real word for Mandarin TD and English TD

3.5 Slopes

The slopes were also calculated based on the linear functions of individual of identification curves. We mainly focused on the slopes which crossed the chance level (50%). If the intercept point of a particular participant was 2.5, meaning the chance level was between step 2 and step 3, then the slope would be the actual number of T4 responses at step 3 minus that of at step 2. If the intercept point of an individual was 2, meaning that the chance level was just on step 2, then the slope would be the average of the actual number of T4 responses at step 3 minus that of at step 2, and the actual number of T4 responses at step 2 minus that of at step 1.

A mixed ANOVA was performed with one within-subject factor, Stimulus (Real word, Nonce word, Pure tone), and one between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the dependent variable, Slope. (Once again the data of Mandarin KTD is not included because they were only tested on the real word stimuli.) There was no main effect of Group $F(4, 31) = .593$, $p = .671$, $\eta^2 = .071$, and there was no main effect of Stimulus $F(2, 62) = 1.135$, $p = .328$, $\eta^2 = .035$. There was no interaction between the within-subject factor, Stimulus, and the between-subject factor, Group F(8, 62) = .866, p = .550, η^2 = .100.

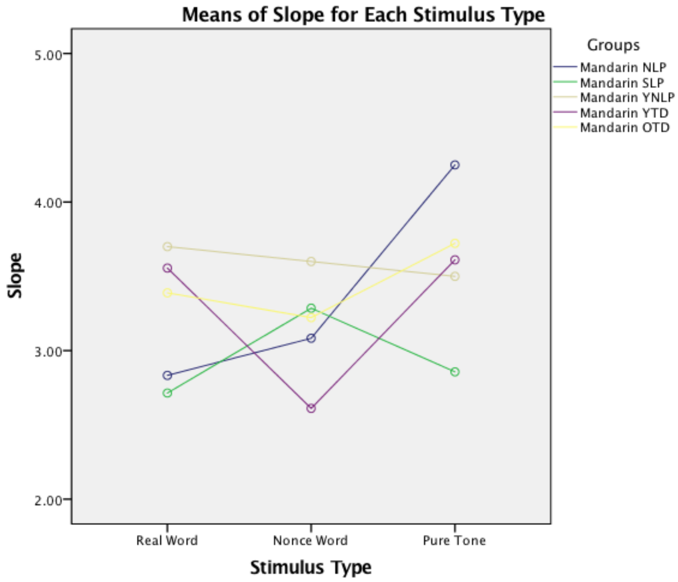

Figure 34. Means of slope for each stimulus type for four Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD)

Now we would like to focus on the analysis of real words, so that the data of Mandarin KTD as well as English participants could also be included in the following four tests on the slopes.

A one-way ANOVA was performed with a between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD, Mandarin KTD) on the dependent variable, Slope, for the data of real word. There was no main effect of Group F(5, 52) = .810, p = .548.

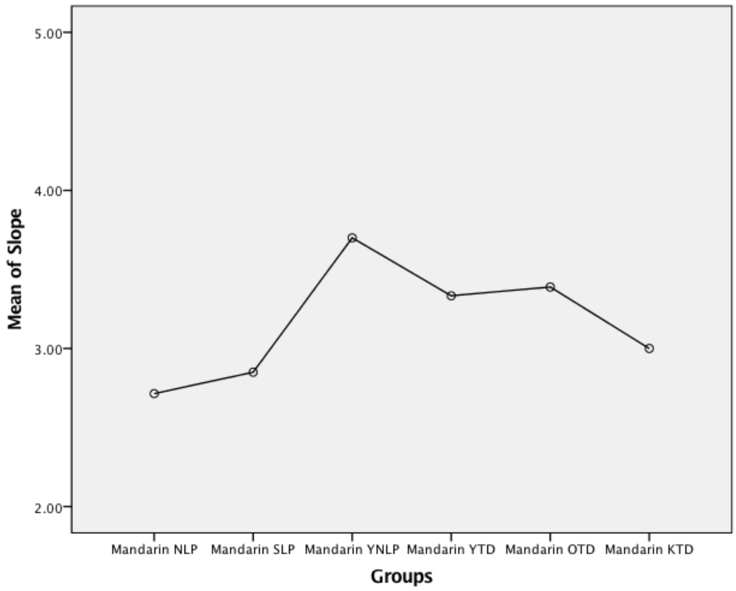

Figure 35. Means of slope for real word for five Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD, Mandarin KTD)

A one-way ANOVA was performed with a between-subject factor, Group (English NLP, English SLP, English YTD, English OTD) on the dependent variable, Slope, for the data of real word. There was no main effect of Group F(3, 28) = .857, *p* = .475.

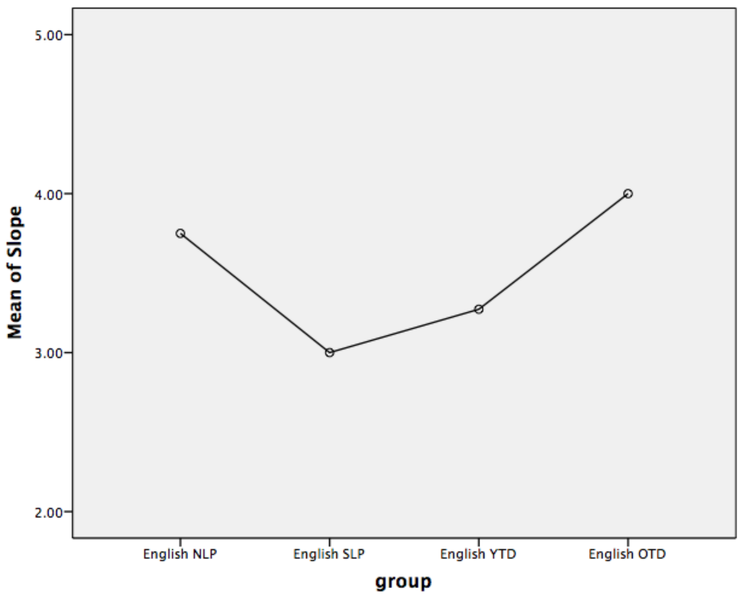

Figure 36. Means of slope for real word for four English groups (English NLP, English SLP, English YTD, English OTD)

A one-way ANOVA was performed with a between-subject factor, Group

(Mandarin, English participants) on the dependent variable, Slope, for the data of real word. There was no main effect of Group $F(1, 88) = 1.637$, $p = .204$.

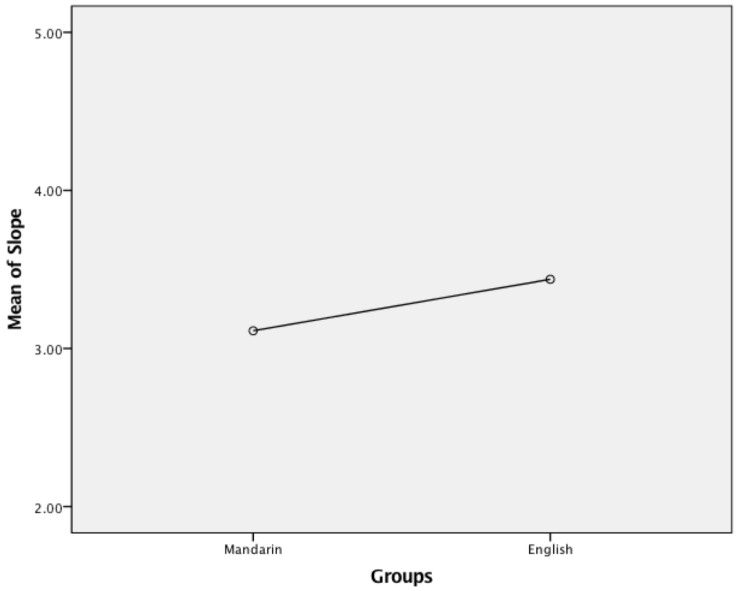

Figure 37. Means of slope for real word for Mandarin and English groups

A one-way ANOVA was performed with a between-subject factor, Group (Mandarin ASD, Mandarin TD, English ASD, English TD) on the dependent variable, Slope, for the data of real word. There was no main effect of Group $F(3, 86) = .775$, $p = .511$.

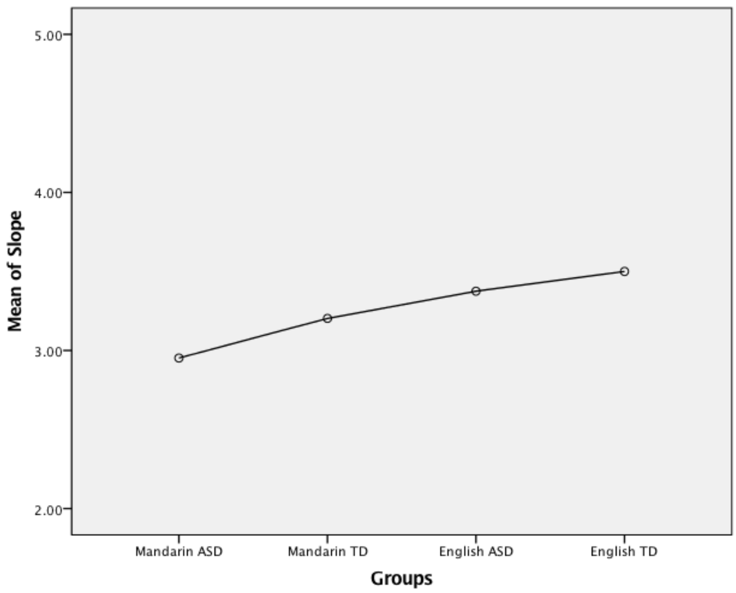

Figure 38. Means of slope for real word for Mandarin ASD, Mandarin TD, English ASD, and English TD

4 Discussion of naming task results

Categorical perception of lexical tones

Q1: Do Mandarin-speaking children with autism spectrum disorder (ASD) categorically perceive and identify lexical tones in the same way as their typically-developing (TD) Mandarin-speaking counterparts, or do Mandarin-speaking ASD children perceive lexical tones in a psychophysical way?

The forced-choice identification task for Mandarin participants showed that Mandarin ASD did not perform significantly differently in the identification of Mandarin lexical tones from Mandarin YTD and OTD (Figure 18). These three groups all had low proportion of T4 responses at step 0 and step 1, then had a sharp slope and reached around 50% of T4 responses at step 2, and finally had high percentage of T4 responses at step 3, step 4, step 5, step 6, and step 7. This suggested that the Mandarin ASD and TD both perceived the lexical tones in a categorical way instead of a psychoacoustic way (Figure 17 and Table 27).

Nevertheless, since the data of ASD group were heterogeneous, we once again split it into Mandarin NLP and Mandarin SLP for further investigations (Figure 24). It turned out that the identification of lexical tones in Mandarin NLP was significantly different from that of in Mandarin SLP and Mandarin YTD. Just as indicated in Table 27 and Figure 19, Mandarin NLP had higher proportions of T4 responses than other groups for every step except step 0, suggesting that they were more inclined to label an item with pitch contours along the T1-T4 continuum as Tone 4 instead of Tone 1. Moreover, the proportions of T4 responses for step 5, 6, and 7 were almost 100%, indicating that Mandarin NLP categorized these steps altogether as Tone 4.

On the other hand, the percentage of T4 responses in Mandarin SLP was significantly higher than Mandarin YTD, and marginally higher than Mandarin NLP and OTD at step 0 (i.e. Tone 1). Except for this, the proportions of T4 responses in Mandarin SLP were consistently lower than which of Mandarin NLP. These results revealed that Mandarin SLP had more difficulty in identifying Tone 1 itself than other three groups. In addition, while Mandarin NLP tended to label an item with pitch contours along the T1-T4 continuum as Tone 4 instead of Tone 1, Mandarin SLP participants, just as Mandarin YTD participants, were less inclined to identify the items as Tone 4 along the tone continuum.

It is important to point out that Mandarin NLP and SLP might perceive the lexical tones differently, or they might just have certain preferences or take various strategies in identifying Tone 1 and Tone 4. While a less sharp falling in pitch contour was also considered to be Tone 4 by Mandarin NLP, it seemed that Mandarin SLP participants were stricter and more conservative for the classification of Tone 4.

Q1a: Are the intercept points around the same place for both Mandarin typically-developing and ASD participants?

The intercept points across stimulus types for Mandarin ASD (Mandarin NLP, Mandarin SLP, and Mandarin YNLP) and TD (Mandarin YTD and Mandarin OTD) were both around step 2 and only showed a marginal difference between groups (Figure 26). However, the intercept points for real word (Mandarin KTD was also included) were significantly lower in Mandarin YNLP than in Mandarin SLP as well as Mandarin KTD (Figure 27).

Q1b: Are the slopes at the category boundary significantly different for Mandarin typically-developing and ASD participants?

The slopes across stimulus types for Mandarin ASD and TD (Mandarin YTD and Mandarin OTD) were both around 3 and did not show any statistical difference (50). In addition, although the slope for real word in Mandarin SLP was slightly lower than other Mandarin groups including KTD, there was still no statistical significance.

Role of native language

Q2: According to the literature, speakers of a tone language have a categorical perception of their native lexical tones, whereas speakers of a non-tone language perceive and process their lexical tones in a psychophysical way. Do Mandarin- and English-speaking participants perceive lexical tones differently, as suggested in the literature?

The AXB identification task for English participants showed that English ASD did not perform significantly differently in the identification of non-native lexical tones from English TD. There was a significant correlation between the Mandarin and the English participants. Just as shown in Figure 37, although the slopes around the category boundary in Mandarin group was not much steeper than in English participants, the percentage of T4 responses at step 0 and step 1 was significantly lower in Mandarin than in English participants. Therefore, while the slopes around the category boundary in Mandarin group was not significantly sharper than in English participants, the slopes along the tone continuum were actually significantly steeper in Mandarin than in English participants. This suggested that the Mandarin listeners tended to identify the native lexical tones in a more categorical way, and the English individuals perceived the non-native lexical tones in a more continuous way. The similar pattern was also found in the significant interaction between Mandarin TD and English TD. Although the results of Mandarin ASD and English ASD were not that clear and neat as the TD groups and did not have significant interaction, it could still be observed that the identification curve along the tone continuum in English ASD was less sharp than which of in Mandarin ASD (Figure 24 and 25).

Q2a: Are intercept points around the exact centre of the tonal continuum or at a different point for Mandarin and English participants?

English ASD and English TD both had an intercept point around step 2 and did not differ significantly (Figure 30). While the intercept points in Mandarin overall participants and English overall participants were not significantly different, the intercept point in Mandarin TD was marginally higher than in English TD.

Q2b: Are the slopes at the category boundary significantly different for Mandarin and English participants?

English ASD and English TD both had a slope around 3.5 and did not differ significantly (Figure 38). In addition, although the slope for real word was slightly higher in English participants than in Mandarin participants (including KTD), this was not statistical significant. Nevertheless, as discussed in Q2, while the slopes of category boundary did not differ in Mandarin and English groups, the slopes along the continuum were actually significantly sharper in Mandarin than in English individuals. It would be more precise to explore the slopes along the tone continuum rather than only in the category boundary.

Role of stimulus type (real word, nonce word, pure tone)

Q3: Do ASD children and their controls identify the pitch contours of different kinds of stimuli (real words, nonce words, and pure tones) differently? Is there any interaction between the group and type of stimulus?

Mandarin ASD and Mandarin TD both behaved differently for different kinds of stimuli. Generally speaking, Mandarin participants had the highest percentage of T4 responses for pure tone, then for real word, and the lowest for nonce word. However, there was a significant interaction between the group and type of stimulus, which suggested that certain groups might behave differently in identifying these three stimulus types. The post-hoc tests revealed Mandarin ASD and OTD both had significantly higher percentage of T4 responses for pure tone than for nonce word. On the other hand, Mandarin YTD had significantly higher percentage of T4 responses for pure tone than for nonce word as well as real word. As shown in Figure 21 and 22, this could be attributed to the particularly high percentage of T4 responses for pure tone in Mandarin YTD.

Q3a: Do the different groups of participants categorically perceive nonce words and pure tones equally strongly as real words? Do they have the same intercept points, the same slopes at the intercept points, and the same enhanced perception at the category boundary of nonce word, pure tone, and real word stimuli?

Just as discussed above, the overall Mandarin data showed that the percentage of T4 responses was significantly higher in pure tone than in real word than in nonce word along the tone continuum (Figure 20). Overall, it is easier to perceive and discriminate the pure tones in the categorical perception naming test, since the participants could focus on the auditory pitch without the distraction of linguistic information. As for the real words and nonce words, although they both provide auditory pitch as well as linguistic information for discriminating the tones, it turned out that the real words are more helpful in the way that the participants have already learnt and categorized the real words in their lexicon.

On the other hand, these three stimulus types all had similar main effect of Step. That is, step 0 and 1 could be considered being within one category, whereas step 3, 4, 5, 6, and 7 could be categorized together. Therefore, Mandarin ASD and TD generally perceived these three kinds of stimuli in a categorical way.

Mandarin ASD and TD groups both had higher intercept points of nonce word than that of real word and pure tone. However, the slopes did not differ between these three stimulus types.

5 Method of Task 2: Two-step discrimination task

5.1 Participants

The same Mandarin- and English-speaking participants as in the previous experiment participated in the current task.

5.2 Two-step discrimination task

This was an AXB discrimination task. The participants heard three stimuli per trial and had to determine whether the second sound (X) was the same as the first (A) or the third sound (B). Just as Experiment 3, there were three subtests for the Mandarin participants: real words, nonce words, and pure tone, whereas the English participants only had one real word subtest.

5.3 Stimuli

The speech materials were the same stimuli as those in Experiment 3. There were also three kinds of stimuli: real words, nonce words, and pure tones, and they were distinguished and presented in different subtests. Three stimuli from the same continuum were played in a row for each trial. The first stimulus (A) and the third (B) for each continuum were two steps apart, so that there were six possible A-B pairs (step 0-step2, step 1-step 3, step 2-step 4, step 3-step 5, step 4-step 6, and step 5-step 7). In each trial, A and B corresponded to one two-step pair of a continuum, while the second stimulus (X) was identical to either A or B. As a consequence, the AXB trials had four possible combinations (AAB, ABB, BAA, and BBA). Nevertheless, in order to explore and compare the impact of the ascending, as well as the descending order, half of the participants only encountered AAB and ABB, while the other half were presented with BAA and BBA. In addition, the trials were organised in a pseudo-random order, making 36 trials (6 tone pairs x 3 syllables x 2 combination) for each participant in each subtest. The orders were created so that there were not more than two identical words after each other and more than one identical tone pair. Sixteen different lists were created in total. Three cartoon dinosaurs in different colours were shown on computer monitor to make the experiments more interesting to young children. The first sound (A) was produced by the red Dino on the left of the screen, the second (X) was made by the yellow Dino in the middle of the screen, and the third (B) by the blue Dino on the right of the screen. A Dino jumped when a sound was played. All the tasks were run on Mammoth software.

5.4 Procedure

Every participant was presented with three subtests: real words, nonce words, and pure tones, and the order of the three subtests were randomised. In this task, the participants heard three stimuli per trial along one continuum and were required to make a forced choice as to whether the second stimulus (X) sounded the same as the first (A) or the third (B). If the participants believed that the second sound (X) made by the yellow dinosaur in the middle of the screen was the same as the first stimulus (A) produced by the red dinosaur on the left of the screen, they should press the key Q. If they considered that the second stimulus (X) was the same as the third sound made by the blue dinosaur on the right of the screen, they should press the key P. No feedback was given in the main phase and the next trial was presented after each click.

6 Results of discrimination task

6.1 Mandarin results across stimulus types

The results across participants are displayed in Table 30. The percentages of correct responses were all above the chance level (50%) in all steps. As manifested in Figure 39 and 40, the lines were askew and the performance on the left was relatively higher than on the right. The accuracy rate was the highest at pair 0-2 and pair 1-3 (both around 90%), then became lower and lower along the continuum, and finally reached the bottom at pair 5-7 (73%). Mandarin NLP has higher percentages of correct responses than other four groups at pair 0-2 (93%), pair 1-3 (94%), and pair 5-7 (79%), while Mandarin OTD has higher percentages of correct responses than other four groups at pair 3-5 (88%) and pair 4-6 (84%). On the other hand, Mandarin SLP has lower percentages of correct responses than other four groups at pair 0-2 (84%) and pair 1-3 (82%), while Mandarin YTD has lower percentages of correct responses than other four groups at pair 2-4 (65%), pair 3-5 (64%), pair 4-6 (56%) and pair 5-7 (57%).

		0-2	1-3	2-4	3-5	4-6	5-7
Mandarin ASD (*n* = 21)		0.88	0.87	0.83	0.76	0.73	0.71
	Mandarin NLP (*n* = 7)	0.93	0.94	0.83	0.79	0.81	0.79
	Mandarin SLP (*n* = 9)	0.84	0.82	0.88	0.77	0.74	0.68
	Mandarin YNLP (*n* = 5)	0.91	0.86	0.65	0.64	0.56	0.57
Mandarin YTD (*n* = 10)		0.89	0.92	0.88	0.79	0.76	0.78
Mandarin OTD (*n* = 9)		0.91	0.92	0.85	0.88	0.84	0.73
Overall (*n* = 40)		0.89	0.90	0.84	0.80	0.77	0.73

Table 30. Percentage of correct responses of two-step discrimination across stimulus type for each Mandarin group

In order to further explore if various groups would behave differently in this task, whether there is a significant effect of the steps, and whether there is a correlation between the groups and the steps, a mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor Group (Mandarin ASD, Mandarin TD). There was no effect of Group $F(1, 34) = 1.443$, $p =$

.238, η^2 = .041, but here was a main effect of Step F (5, 170) = 12.994, $p <$ .001, η^2 = .276. Bonferroni post-hoc tests indicated that the performance at step 0-2 was significantly higher than the performance at step 3-5 (p = .008), step 4-6 (p = .003), and step 5-7 (p = .001). In addition, the performance at step 1-3 was significantly higher than the performance at step 3-5 (p < .001), step 4-6 (p < .001), and step 5-7 (p < .001). Further, the performance at step 2-4 was significantly higher than the performance at step 5-7 (p = .015). There was no interaction between the within-subject factor, Step, and the between-subject factor, Group F (5, 170) = .340, p = .888, η^2 = .010.

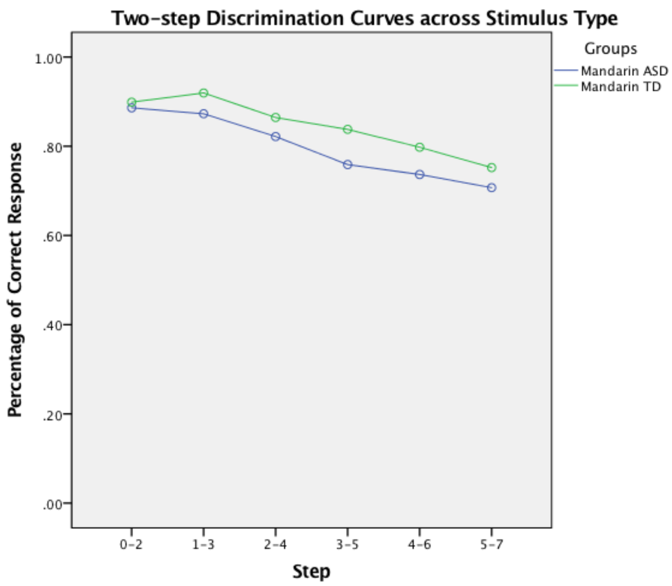

Figure 39. Two-step discrimination curves across stimulus type for Mandarin ASD and Mandarin TD

A mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD). There was no effect of Group F (4, 31) = 1.244, p = .313, η^2 = .138, but here was a main effect of Step F (5, 155) = 14.846, $p <$ .001, η^2 = .324. Bonferroni post-hoc tests indicated that the performance at step 0-2 was significantly higher than the performance at step 3-5 (p = .001), step 4-6 (p < .001), and step 5-7 (p = .001). In addition, the performance at step 1-3 was significantly higher than the performance at step 3-5 (p < .001), step 4-6 (p < .001), and step 5-7 (p < .001). Further, the performance at step 2-4 was marginally higher than the performance at step 5-7 (p = .053). There was no interaction between the within-subject factor, Step, and the between-subject factor, Group F (20, 155) = 1.247, p = .224, η^2 = .139.

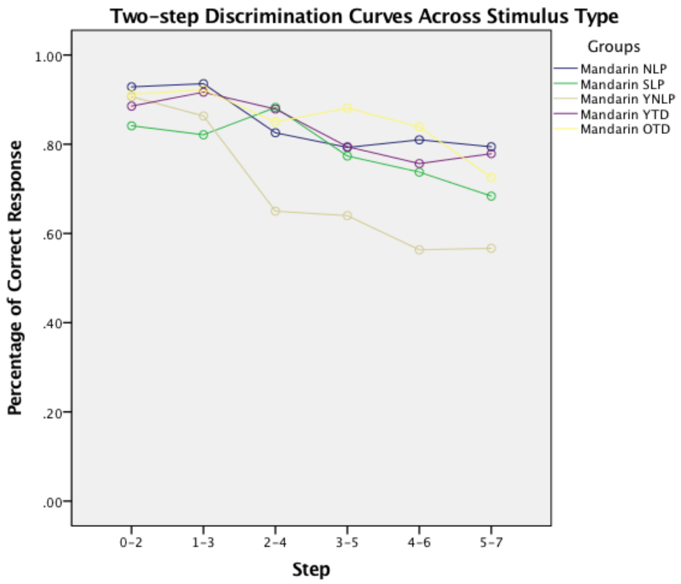

Figure 40. Two-step discrimination curves across stimulus type for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, and Mandarin OTD

6.2 Breakdown of Mandarin results for the three stimulus types

6.2.1 Real word

A mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor, Group (Mandarin ASD, Mandarin YTD, Mandarin OTD) on the real word stimuli. There was no effect of Group (F(2, 27) = .935, p = .405), but there was a main effect of Step F(5, 135) = 12.668, p < .001. Bonferroni post-hoc tests further indicated that the performance at step 0-2 was significantly higher than the performance at step 3-5 (p = .019), step 4-6 (p = .002), and step 5-7 (p = .001). In addition, the performance at step 1-3 was significantly higher than the performance at step 3-5 (p = .005), step 4-6 (p < .001), and step 5-7 (p < .001). Further, the performance at step 2-4 was significantly higher than the performance at step 3-5 (p = .049), step 4-6 (p = .027), and step 5-7 (p = .015). There was no interaction between the two factors Step and Group (F(10, 135) = 1.553, p = .127).

Since the data of ASD group are heterogeneous as usual, we once again split it into Mandarin NLP, Mandarin SLP, and Mandarin YNLP, and thus have five groups to explore. A mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the data of real word. There was no effect of Group $F(4, 25) = 1.251$, $p = .315$, $\eta^2 = .167$, but there was a main effect of Step $F(5, 125) = 14.120$, $p < .001$, $\eta^2 = .361$. Bonferroni post-hoc tests further indicated that the performance at step 0-2 was significantly higher than the performance at step 2-4 ($p = .005$), step 3-5 ($p = .001$), step 4-6 ($p < .001$), and step 5-7 ($p = .001$). In addition, the performance at step 1-3 was significantly higher than the performance at step 3-5 ($p = .003$), step 4-6 ($p < .001$), and step 5-7 ($p = .001$). Further, the performance at step 2-4 was significantly higher than the performance step 4-6 ($p = .013$). There was a significant interaction between the two factors Step and Group $F(20, 125) = 1.904$, $p = .017$, $\eta^2 = .233$. Bonferroni post-hoc test indicated that at step 4-6 the performance in Mandarin YNLP was significantly lower than the performance in Mandarin NLP ($p = .024$), and was marginally lower than the performance in Mandarin YTD ($p = .053$)

		0-2	1-3	2-4	3-5	4-6	5-7
Mandarin ASD (n = 21)		0.85	0.82	0.74	0.68	0.70	0.65
	Mandarin NLP (n = 7)	0.90	0.93	0.71	0.69	0.81	0.72
	Mandarin SLP (n = 9)	0.79	0.77	0.81	0.71	0.71	0.65
	Mandarin YNLP (n = 5)	0.92	0.67	0.59	0.50	0.25	0.42
Mandarin YTD (n = 10)		0.95	0.90	0.86	0.72	0.76	0.67
Mandarin OTD (n = 9)		0.92	0.97	1.00	0.89	0.69	0.59

Overall (*n* = 40)	0.89	0.87	0.82	0.73	0.71	0.64

Table 31. Percentage of correct responses of two-step discrimination for real word for each group

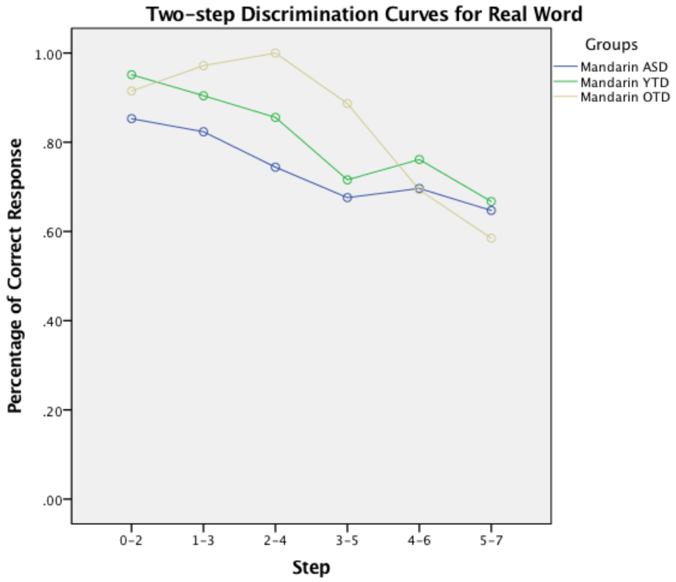

Figure 41. Two-step discrimination curves for real word for Mandarin ASD, Mandarin YTD, and Mandarin OTD

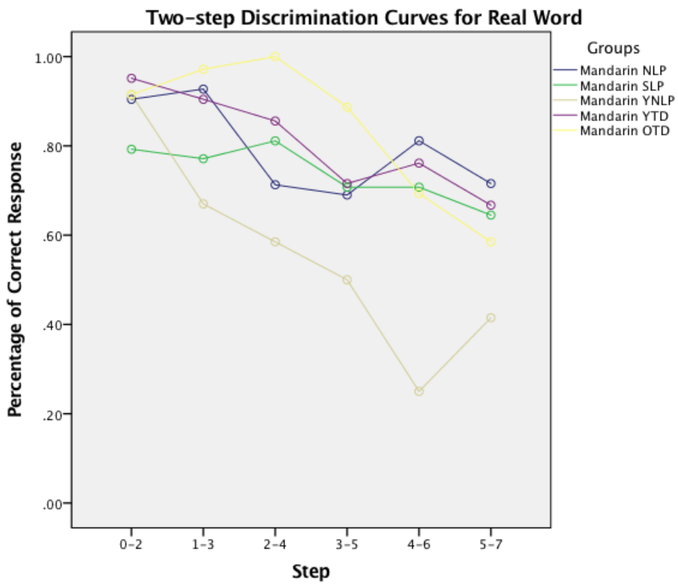

Figure 42. Two-step discrimination curves for real word for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, and Mandarin OTD

6.2.2 Nonce word

A mixed ANOVA was performed with a within-subject factor, Step (step 0-

2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor, Group (Mandarin ASD, Mandarin YTD, Mandarin OTD) on the data of nonce word. There was no main effect of Group F(2, 26) = 0.094, p = .910, but there was a main effect of Step F(5, 130) = 4.294, p = .001). Bonferroni post-hoc tests further indicated that the performance at step 1-3 was marginally higher than the performance at step 5-7 (p = .067). There was no interaction between the two factors Step and Group F(10, 130) = 1.423, p = .177.

Since the data of ASD group are heterogeneous as usual, we once again split it into Mandarin NLP, Mandarin SLP, Mandarin YNLP, and thus have five groups to explore. A mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the data of nonce word. There was no main effect Group F(4, 24) = .922, p = .467, η^2 = .133, but there was a main effect of Step F(5, 120) = 7.740, p < .001, η^2 = .244). Bonferroni post-hoc tests further indicated that the performance at step 0-2 was significantly higher than the performance at step 4-6 (p = .029) and step 5-7 (p = .015). In addition, the performance at step 1-3 was significantly higher than the performance at step 4-6 (p = .031) and step 5-7 (p = .003). The performance at step 2-4 was significantly higher than the performance step 5-7 (p = .012). Further, the performance at step 3-5 was marginally higher than the performance at step 5-7 (p = .083). There was no interaction between the two factors Step and Group F(20, 120) = 1.436, p = .119, η^2 = .193.

		0-2	1-3	2-4	3-5	4-6	5-7
Mandarin ASD (n = 21)		0.91	0.89	0.93	0.83	0.73	0.70
	Mandarin NLP (n = 7)	0.94	0.92	0.97	0.89	0.78	0.81
	Mandarin SLP (n = 9)	0.89	0.83	1.00	0.89	0.72	0.70

	Mandarin YNLP (*n* = 5)	0.89	0.94	0.72	0.61	0.66	0.50
Mandarin YTD (*n* = 10)		0.86	0.90	0.88	0.79	0.74	0.83
Mandarin OTD (*n* = 9)		0.90	0.90	0.81	0.93	0.86	0.74
Overall (*n* = 40)		0.90	0.90	0.89	0.84	0.76	0.74

Table 32. Percentage of correct responses of two-step discrimination for nonce word for each group

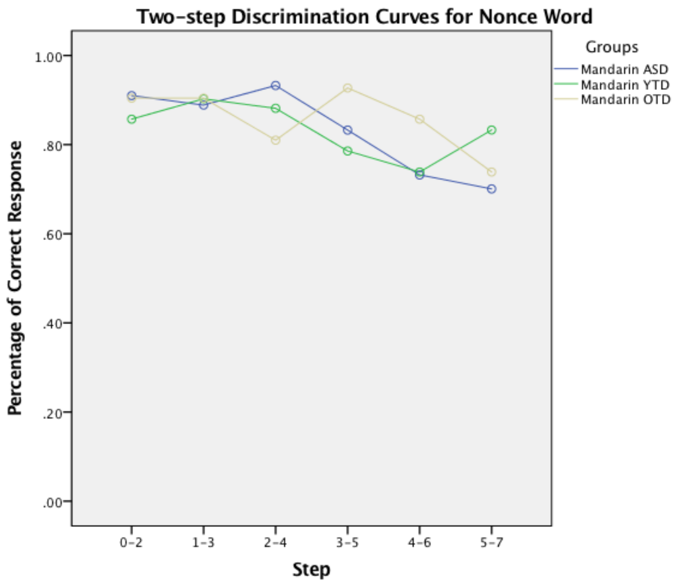

Figure 43. Two-step discrimination curves for nonce word for Mandarin ASD, Mandarin YTD, and Mandarin OTD

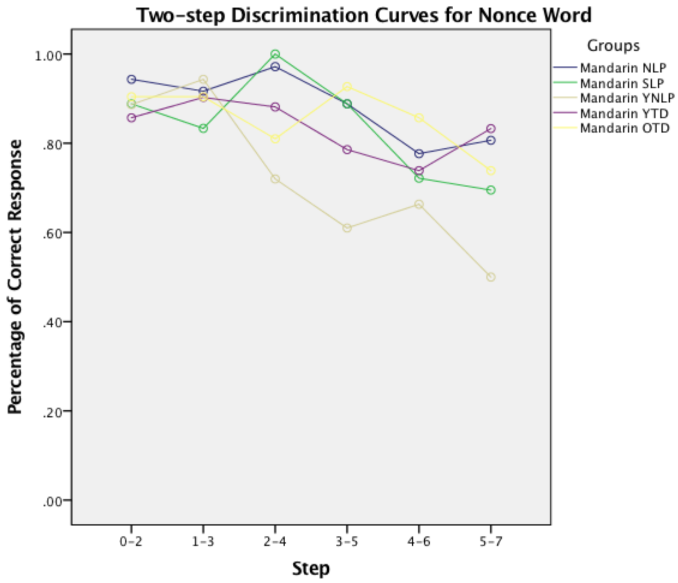

Figure 44. Two-step discrimination curves for nonce word for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, and Mandarin OTD

6.2.3 Pure tone

A mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor, Group (Mandarin ASD, Mandarin YTD, Mandarin OTD) on the data of pure tone. There was no effect of Group F(2, 25) = .688, p = .512, but there was a main effect of Step F(5, 125) = 2.872, p = .017. Bonferroni post-hoc tests further indicated that the performance at step 1-3 was marginally higher than the performance at step 5-7 (p = .066). There was no interaction between the two factors Step and Group F(10, 125) = .746, p = .680.

Since the data of ASD group are heterogeneous as usual, we once again split it into Mandarin NLP, Mandarin SLP, and Mandarin YNLP, and thus have five groups to explore. A mixed ANOVA was performed with a within-subject factor, Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7), and a between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the data of pure tone. There was no effect of Group F(4, 23) = .467, p = .759, η^2 = .075, but there was a main effect of Step F(5, 115) = 3.870, p = .003, η^2 = .144**.** Bonferroni post-hoc tests indicated that the performance at step 1-3 was marginally higher than the performance at step 5-7 (p = .059). There was no interaction between the two factors Step and Group $F(20, 115) = .648$, $p = .868$, $\eta^2 = .101$.

		0-2	1-3	2-4	3-5	4-6	5-7
Mandarin ASD (n = 21)		0.89	0.92	0.81	0.77	0.77	0.77
	Mandarin NLP (n = 7)	0.92	0.94	0.80	0.75	0.78	0.81
	Mandarin SLP (n = 9)	0.86	0.90	0.88	0.79	0.81	0.76
	Mandarin YNLP (n = 5)	0.94	0.89	0.67	0.78	0.67	0.72
Mandarin YTD (n =10)		0.88	0.96	0.83	0.96	0.92	0.83
Mandarin OTD (n = 9)		0.94	0.92	0.79	0.83	0.87	0.71
Overall (n = 40)		0.90	0.92	0.81	0.82	0.82	0.76

Table 33. Percentage of correct responses of two-step discrimination for pure tone for each group

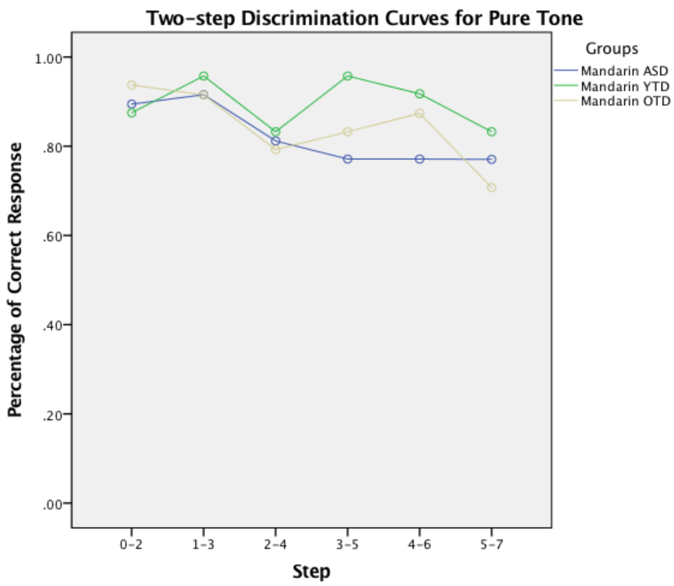

Figure 45. Two-step discrimination curves for pure tone for Mandarin ASD, Mandarin YTD, and Mandarin OTD

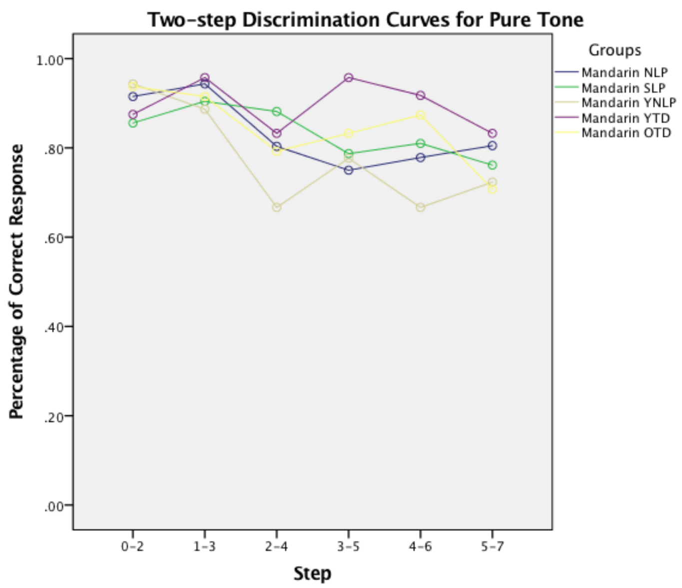

Figure 46. Two-step discrimination curves for pure tone for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, and Mandarin OTD

6.3 Breakdown of Mandarin results for the different participant groups

After examining the percentage of correct responses in two-step discrimination across stimulus types, now we would like to explore the results for each participant group. Two-way repeated measures ANOVA was performed with two within-subject factors Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and Stimulus (real word, nonce word, pure tone) on the data of Mandarin participants. There was a main effect of Stimulus F(2, 38) = 8.283, p = .001, η^2 = .304. Bonferroni post-hoc tests indicated that the performance in real word was significantly lower than the performance in nonce word (p = .007) and pure tone (p = .024). There was a main effect of Step F(5, 95) = 13.424, p < .001, η^2 = .414. Bonferroni post-hoc tests further indicated that the performance at step 0-2 was significantly higher than the performance at step 4-6 (p = .002), and step 5-7 (p = .002), and marginally higher than the performance at step 3-5 (p = .060), In addition, the performance at step 1-3 was significantly higher than the performance at step 3-5 (p = .001), step 4-6 (p = .001), and step 5-7 (p < .001). Further, the performance at step 2-4 was significantly higher than the performance at step 5-7 (p = .033). There was a marginal interaction between the two within-subject factors Step and Stimulus F(10, 190) = 1.673, p = .089, η^2 = .081. Bonferroni post-hoc tests indicated that the performance in real word was marginally lower than the performance in nonce word at step 0-2 (p = .095). In addition, at step 3-5, the performance in real word was significantly lower than the performance in nonce word (p = .010), and was marginally lower than the performance in pure tone (p = .087). Further, the performance in real word was marginally lower than the performance in pure tone at step 4-6 (p = .092) and step 5-7 (p = .067).

	0-2	1-3	2-4	3-5	4-6	5-7	Total
Real word	0.87	0.86	0.80	0.68	0.68	0.58	0.75
Nonce word	0.92	0.90	0.92	0.86	0.77	0.71	0.85
Pure tone	0.91	0.93	0.80	0.83	0.80	0.76	0.84
Total	0.90	0.90	0.84	0.79	0.75	0.68	0.81

Table 34. Percentage of correct responses of two-step discrimination for each stimulus type across group

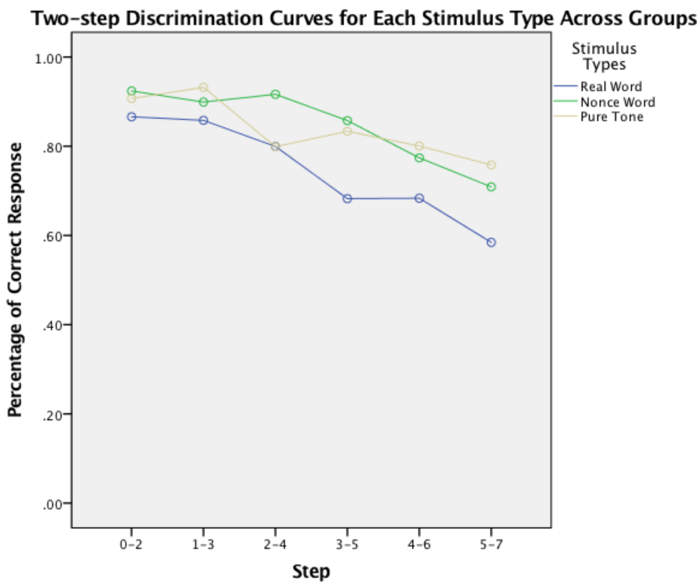

Figure 47. Two-step discrimination curves for each stimulus type across groups

6.3.1 Mandarin ASD

Two-way repeated measures ANOVA was performed with two within-subject factors Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and Stimulus (real word, nonce word, pure tone) on the data of Mandarin ASD. There was a main effect of Stimulus $F(2, 24) = 8.553$, $p = .002$, $\eta^2 = .416$, Bonferroni post-hoc tests indicated that the performance in real word was significantly lower than the performance in nonce word (p = .011) and in pure tone (p = .042). There was a main effect of Step F(5, 60) = 7.491, p < .001, η^2 = .384. Bonferroni post-hoc tests manifested the performance at step 0-2 was marginally higher than the performance at step 4-6 (p = .051). Besides, the performance at step 1-3 was significantly higher than the performance at step 3-5 (p = .006), step 4-6 (p = .008), and step 5-7 (p = .012). There was a significant interaction between the two factors Step and Stimulus F(10, 120) = 1.934, p = .047, η^2 = .139. Bonferroni post-hoc tests indicated that the performance in real word was significantly lower than the performance in nonce word at step 2-4 (p = .017). Besides, at step 3-5, the performance in real word was significantly lower than the performance in nonce word (p = .023), and was marginally lower than the performance in pure tone (p = .063). Further, the performance in real word was marginally lower than the performance in pure tone at step 5-7 (p = .053).

	0-2	1-3	2-4	3-5	4-6	5-7	Total
Real word	0.85	0.83	0.73	0.63	0.68	0.62	0.72
Nonce word	0.91	0.90	0.97	0.86	0.77	0.74	0.86
Pure tone	0.91	0.93	0.82	0.82	0.77	0.81	0.84
Total	0.89	0.89	0.84	0.77	0.74	0.72	0.81

Table 35. Percentage of correct responses of two-step discrimination for each stimulus type for ASD

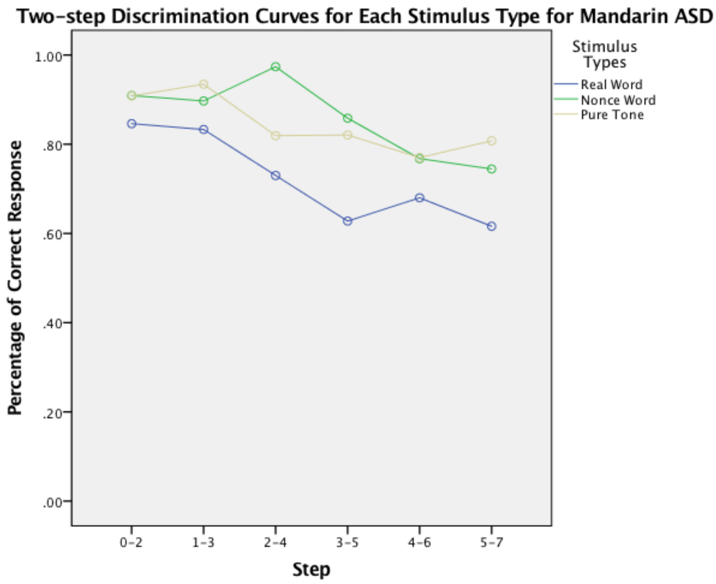

Figure 48. Two-step discrimination curves for each stimulus type for Mandarin ASD

6.3.2 Mandarin TD

Two-way repeated measures ANOVA was performed with two within-subject factors Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and Stimulus (real word, nonce word, pure tone) on the data of Mandarin TD. There was no effect of Stimulus F (2, 12) = .678, p = .526, η^2 = .102, but there was a main effect of Step F(5, 30) = 7.580, p < .001, η^2 = .558. Bonferroni post-hoc tests manifested that the performance at step 0-2 was marginally higher than the performance ate step 5-7 (p = .060). There was no interaction between the two factors Step and Stimulus F(10, 60) = 1.056, p = .410, η^2 = .150.

	0-2	1-3	2-4	3-5	4-6	5-7	Total
Real word	0.90	0.90	0.93	0.78	0.69	0.53	0.78
Nonce word	0.95	0.90	0.81	0.86	0.79	0.64	0.83
Pure tone	0.90	0.93	0.76	0.86	0.86	0.67	0.83
Total	0.92	0.91	0.83	0.83	0.78	0.61	0.81

Table 36. Percentage of correct responses of two-step discrimination for each stimulus type for Mandarin TD

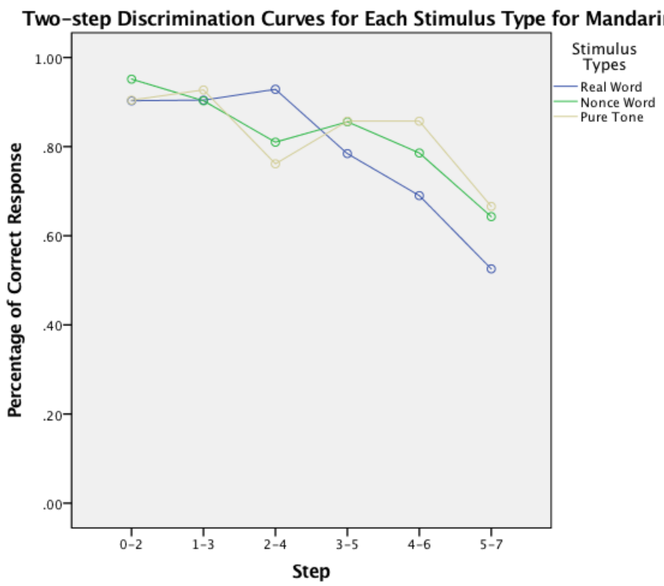

Figure 49. Two-step discrimination curves for each stimulus type for Mandarin TD

6.4 Order

After examining the percentage of correct responses of two-step discrimination for each stimulus type, now we would like to explore the results for increasing and decreasing orders, respectively. A mixed ANOVA was performed with a within-subject factors Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and a between-subject factor Order (increasing order, decreasing order) on the data of Mandarin subjects. There was no main effect of Order $F(1, 34) = .000$, $p = 1.000$, $\eta^2 = .000$, but there was a main effect of Step $F(5, 170) = 17.470$, $p < .001$, $\eta^2 = .339$. Bonferroni post-hoc tests manifested that the performance at step 0-2 was significantly higher than the performance at step 3-5, step 4-6 and step 5-7 ($p = .003$, $p < .001$, and $p < .001$, respectively). Besides, the performance at step 1-3 was significantly higher than the performance at step 3-5, step 4-6 and step 5-7 (all ps $< .001$). Further, the performance at step 2-4 was significantly higher than the performance at step 5-7 ($p = .001$), and the performance at step 3-5 was significantly higher than the performance at step 5-7 ($p = .041$). There was a significant interaction between the two factors Step and Stimulus $F(5, 170) = 6.909$, $p < .001$, $\eta^2 = .169$. Bonferroni post-hoc tests indicated that the performance in increasing order was significantly lower than the performance in decreasing order at step 0-2 ($p = .033$). Nevertheless, the performance in increasing order was significantly higher than the performance in decreasing order at step 5-7 ($p = .018$).

	0-2	1-3	2-4	3-5	4-6	5-7	Total
Increasing	0.85	0.88	0.82	0.79	0.79	0.80	0.82
Decreasing	0.95	0.92	0.87	0.81	0.74	0.64	0.82
Total	0.89	0.90	0.84	0.80	0.77	0.73	0.82

Table 37. Percentage of correct responses of two-step discrimination for increasing and decreasing orders across stimulus types across groups

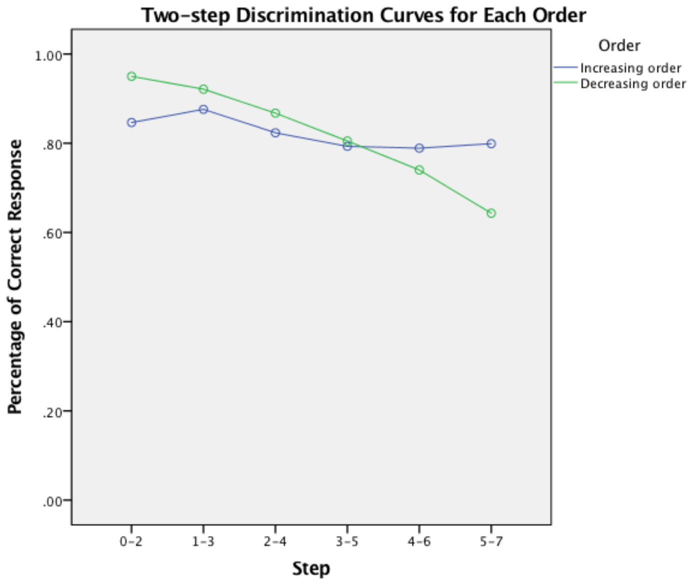

Figure 50. Two-step discrimination curves for increasing and decreasing orders across stimulus types across groups

6.5 AXB

After examining the percentage of correct responses of two-step discrimination for increasing and decreasing orders, now we would like to explore the results for AAB as well as ABB orders, respectively. Two-way repeated measures ANOVA was performed with two within-subject factors Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and AXB (AAB, ABB) on the data of Mandarin participants. There was a main effect of AXB $F(1, 35) = 16.690$, $p < .001$, $\eta^2 = .327$. The percentage of correct responses was significantly higher in AAB order than that of in ABB order. There was also a main effect of Step $F(5, 175) = 14.132$, $p < .001$, $\eta^2 = .288$. Bonferroni post-hoc tests manifested that the performance at step 0-2 was significantly higher than the performance at step 3-5, step 4-6, and step 5-7 ($p = .001$, $p = .001$, and $p < .001$, respectively). Besides, the performance at step 1-3 was significantly higher than the performance at step 3-5, step 4-6, and step 5-7 (*ps* < .001). Further, the performance at step 2-4 was significantly higher than the performance at step 5-7 ($p = .004$). There was no interaction between the two factors Step and AXB $F(5, 175) = 1.110$, $p = .357$, $\eta^2 = .031$.

	0-2	1-3	2-4	3-5	4-6	5-7	Total

AAB	0.93	0.91	0.90	0.86	0.82	0.76	0.86
ABB	0.87	0.88	0.80	0.73	0.71	0.70	0.78
Total	0.90	0.90	0.85	0.80	0.77	0.73	0.82

Table 38. Percentage of correct responses of two-step discrimination for AAB and ABB across groups

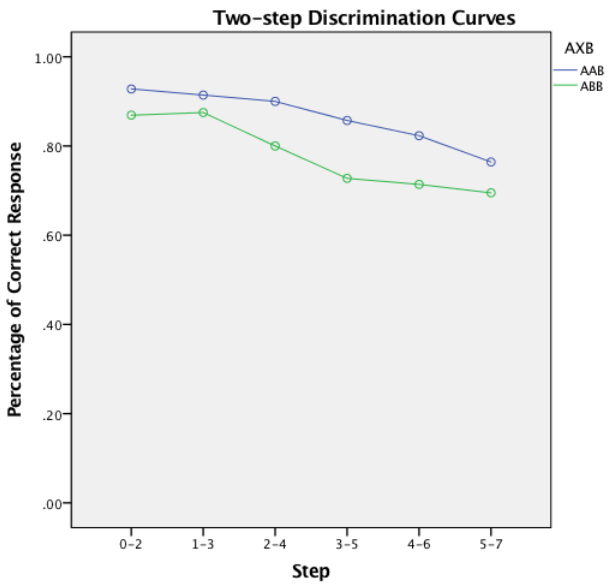

Figure 51. Two-step discrimination curves for AAB and ABB across groups

6.6 Mandarin and English results for real word stimuli

The results for real word stimuli are displayed in Table 39. Most of the percentages of correct responses were above the chance level (50%) except that Mandarin YNLP only had 25% and 42% of correct responses at step 4-6 and step 5-7, respectively. Overall speaking, the percentage of correct responses was higher on the left than on the right continuum. The accuracy rate was the highest at pair 0-2 and pair 1-3 (both around 80%), then became lower and lower along the continuum, and finally reached the bottom at pair 5-7 (65%). While Mandarin participants had higher percentages of correct responses than English participants on the left continuum (step 0-2, step 1-3, and step 2-4), they had lower percentage of correct responses than English participants on the right continuum (step 3-5, step 4-6, and step 5-7). Mandarin TD participants had higher percentages of correct responses than Mandarin ASD participant all along the continuum except at step 5-7, and English TD participants had higher percentages of correct responses than English ASD participant all along the continuum except at step 4-6. When the five Mandarin sub-groups were compared together, Mandarin YTD had the highest percentage of correct responses at step 0-2 (95%), and Mandarin OTD had the highest percentages of correct responses at step 1-3 (97%), step 2-4 (100%), and step 3-5 (89%). On the other hand, Mandarin NLP had the highest percentages of correct responses on the right continuum at step 4-6 (81%) and step 5-7 (72%). In addition, while Mandarin SLP had the lowest percentage of correct responses at step 0-2 (79%), Mandarin YNLP had the lowest percentages of correct responses for the rest along the continuum (67% at step 1-3, 59% at step 2-4, 50% at step 3-5, 25% at step 4-6, and 42% at step 5-7). As for the four English groups, English OTD had the highest percentages along the continuum (93% at step 0-2, 83% at step 1-3, 80% at step 2-4, 80% at step 3-5, 83% at step 4-6, and 77% at step 5-7).

		0-2	1-3	2-4	3-5	4-6	5-7
Mandarin ASD (*n* = 21)		0.85	0.82	0.74	0.68	0.70	0.65
	Mandarin NLP (*n* = 7)	0.90	0.93	0.71	0.69	0.81	0.72
	Mandarin SLP (*n* = 9)	0.79	0.77	0.81	0.71	0.71	0.65
	Mandarin YNLP (*n* = 5)	0.92	0.67	0.59	0.50	0.25	0.42
Mandarin TD (*n* = 19)		0.93	0.94	0.92	0.79	0.73	0.63
	Mandarin YTD (*n* = 10)	0.95	0.90	0.86	0.72	0.76	0.67
	Mandarin OTD (*n* = 9)	0.92	0.97	1.00	0.89	0.69	0.59
Mandarin Overall (*n* = 40)		0.89	0.87	0.82	0.73	0.71	0.64
English ASD (*n* = 15)		0.69	0.73	0.69	0.73	0.74	0.62
	English	0.63	0.76	0.65	0.75	0.74	0.60

	NLP (n = 8)						
	English SLP (n = 7)	0.76	0.70	0.74	0.71	0.74	0.64
English TD (n = 16)		0.80	0.82	0.80	0.74	0.71	0.71
	English YTD (n = 11)	0.73	0.81	0.80	0.71	0.65	0.68
	English OTD (n = 5)	0.93	0.83	0.80	0.80	0.83	0.77
English Overall (n = 31)		0.74	0.77	0.74	0.74	0.73	0.67
Overall (n = 71)		0.82	0.82	0.78	0.73	0.72	0.65

Table 39. Percentage of correct responses of two-step discrimination for real words for each group

6.6.1 Comparison of Mandarin groups

A mixed ANOVA was performed with a within-subject factor Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and a between-subject factor Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the dependent variable, percentage of correct responses, in real word conditions. There was no main effect of Group $F(4, 25) = 1.251$, $p = .315$, $\eta^2 = .167$, but there was a main effect of Step $F(5, 125) = 14.120$, $p < .001$, $\eta^2 = .361$. Bonferroni post-hoc tests indicated that the performance at step 0-2 was significantly higher than step 3-5 ($p = .001$), step 4-6 ($p < .001$) and step 5-7 ($p = .001$). In addition, the performance at step 1-3 was significantly higher than step 3-5 ($p = .003$), step 4-6 ($p < .001$) and step 5-7 ($p = .001$). Further, the performance at step 2-4 was significantly higher than step 4-6 ($p = .013$). There was a significant interaction between the within-subject factor, Step, and the between-subject factor, Group, $F(20, 125) = 1.904$, $p = .017$, $\eta^2 = .233$. Bonferroni post-hoc tests indicated that at step 4-6 the performance in Mandarin YNLP was significantly lower than Mandarin NLP ($p = .025$), and was marginally lower than Mandarin YTD ($p = .053$).

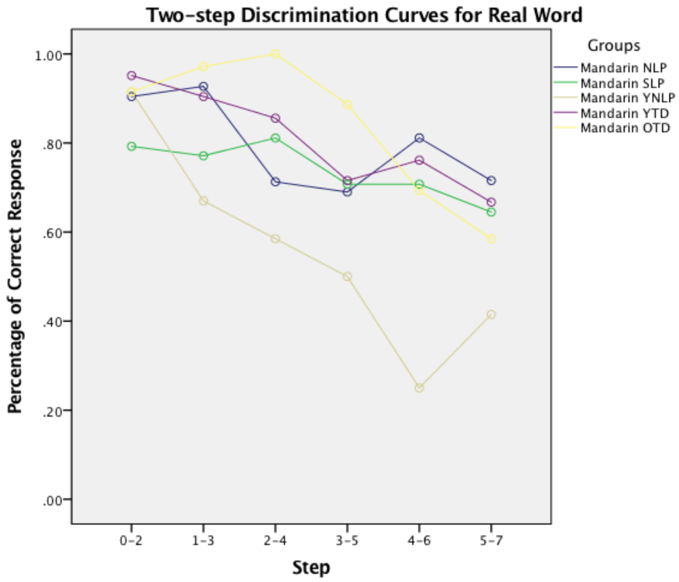

Figure 52. Percentage of correct responses for real word for four Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD)

6.6.2 Comparison of English groups

A mixed ANOVA was performed with a within-subject factor Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and a between-subject factor Group (English NLP, English SLP, English YTD, English OTD) on the dependent variable, percentage of correct responses, in real word conditions. There was no effect of Group F(3, 26) = 1.474, p = .245, η^2 = .145, and there was no effect of Step F(5, 130) = 1.002, p = .419, η^2 =.037. There was no interaction between the two factors Step and Group F(15,130) = .664, p = .816, η^2 = .071.

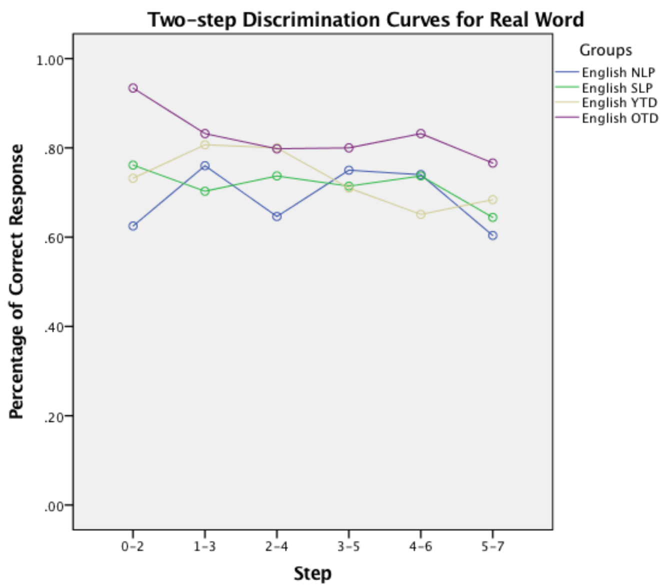

Figure 53. Percentage of correct responses for real word for four English groups (English NLP, English SLP, English YTD, English OTD)

6.6.3 Comparison of Mandarin and English groups

A mixed ANOVA was performed with a within-subject factor Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and a between-subject factor Group (Mandarin, English) on the dependent variable, percentage of correct responses, in real word conditions. There was no effect of Group F (1, 58) = 1.345, p = .251, η^2 = .023, but here was a main effect of Step F(5, 290) = 8.848, p < .001, η^2 = .132. Bonferroni post-hoc tests indicated that the performance at step 0-2 was significantly higher than step 4-6 and step 5-7 (p = .017 and p < .001, respectively). In addition, the performance at step 1-3 was significantly higher than step 3-5, step 4-6 and step 5-7 (p = .013, p = .008 and p < .001, respectively). Further, the performance at step 2-4 was significantly higher than step 5-7 (p = .021). There was a significant interaction between the two factors Step and Group F(5, 290) = 2.623, p = .024, η^2 = .043. Bonferroni post-hoc tests indicated that the performance in Mandarin participants was significantly higher than English participants at step 0-2 (p = .008), and marginally higher than English participants at step 1-3 (p = .063).

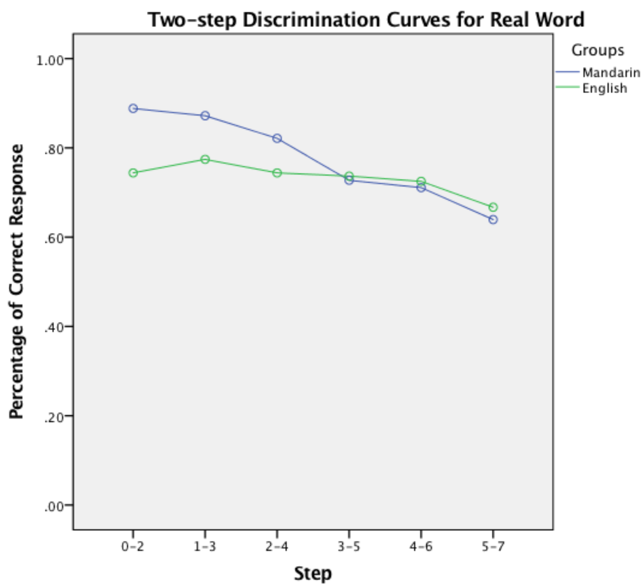

Figure 54. Percentage of correct responses for real word for Mandarin and English groups

6.6.4 Comparison of Mandarin ASD, Mandarin TD, English ASD, and English TD

A mixed ANOVA was performed with a within-subject factor Step (step 0-2, step 1-3, step 2-4, step 3-5, step 4-6, step 5-7) and a between-subject factor Group (Mandarin ASD, Mandarin TD, English ASD, English TD) on the dependent variable, percentage of correct responses, in real word conditions. There was no effect of Group (3, 56) = 1.711, p = .175, η^2 = .084, but there was a main effect of Step (5, 280) = 9.223, p < .001, η^2 = .141. Bonferroni post-hoc tests showed that the performance at step 0-2 was significantly higher than step 4-6 and step 5-7 (p = .014 and p < .001, respectively). In addition, the performance at step 1-3 was significantly higher than step 3-5, step 4-6, and step 5-7 (p = .016, p = .016, and p < .001, respectively). Further, the performance at step 2-4 was significantly higher than step 5-7 (p = .012). There was a marginal interaction between the two factors Step and Group F(15, 280) = 1.569, p = .082, η^2 = .078. Bonferroni post-hoc test indicated that the performance in Mandarin TD was significantly higher than English ASD at step 0-2 and step 2-4 (p = .013 and p = .036, respectively), and was marginally higher than English ASD at step 1-3 (p = .054).

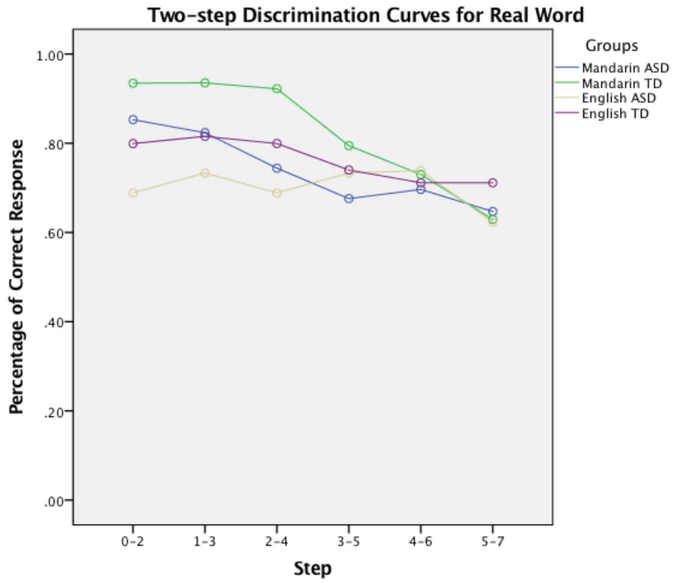

Figure 55. Percentage of correct responses for real word for Mandarin ASD, Mandarin TD, English ASD, and English TD

6.7 Category boundary

The chance level (50%) in the identification curves was considered to be the category boundary. Just as suggested by the previous studies, listeners of tone languages would show enhanced discrimination performance around the category boundary. In order to examine this effect, we would like to calculate the percentage of correct responses in between-category boundary as well as within-category boundary in the two-step discrimination task for each individual. For the between-category boundary, if the intercept point of an individual was 2, then the percentage of correct responses at step 1-step 3 would be taken as the performance in between-category boundary for this participant. If the intercept of an individual was 2.5, meaning that the intercept point was between step 2 and step 3, then the average of percentage of correct responses at step 1-step 3 and percentage of correct responses at step 2-step 4 would be the performance in between-category boundary for this participant. Since each individual might have different intercept points, the levels of the between-category boundary also varied along the step. On the other hand, for the performance in within-category boundary, it was always the average of the percentage of correct responses at step 3-step 5, that of at step 4-step 6, and that of at step 5-step 7 for every participant.

6.7.1 Comparison of different participant groups

A mixed ANOVA was performed with a within-subject factor, Boundary (between-boundary, within-boundary) and a between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD, English NLP, English SLP, English YTD, English OTD) on the data of real word. There was no main effect of Group $F(8, 49) = 1.211$, $p = .312$, $\eta^2 = .165$, but there was a significant main effect of Boundary $F(1, 49) = 41.424$, $p < .001$, $\eta^2 = .458$. In addition, there was a significant interaction between the within-subject factor, Boundary, and the between-subject factor, Group $F(8, 49) = 2.392$, $p = .029$, $\eta^2 = .281$. Bonferroni post-hoc tests indicated that percentage of correct responses in between-category boundary was significantly higher than that of in within-category boundary for Mandarin NLP, Mandarin YNLP, Mandarin YTD and Mandarin OTD ($p = .027$, $p < .001$, $p = .003$ and $p < .001$, respectively).

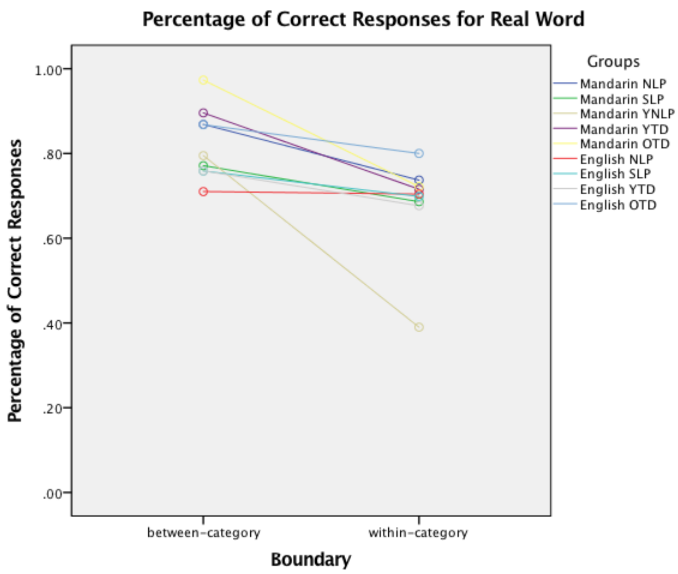

Figure 56. Percentage of correct responses for real word for between-category boundary and within-category boundary for Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin, YTD, Mandarin OTD, English NLP, English SLP, English YTD, and English OTD

6.7.2 Comparison of Mandarin groups

A mixed ANOVA was performed with a within-subject factor, Boundary (between-boundary, within-boundary) and a between-subject factor, Group (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD) on the data of real word. There was no main effect of Group $F(4, 25) = 1.144$, $p = .359$, $\eta^2 = .155$, but there was a significant main effect of Boundary $F(1, 25) = 37.787$, $p < .001$, $\eta^2 = .602$. There was no interaction between the within-subject factor, Boundary, and the between-subject factor, Group $F(4, 25) = 2.011$, $p = .124$, $\eta^2 = .243$. However, Bonferroni post-hoc tests indicated that percentage of correct responses in between-category boundary was significantly higher than that of in within-category boundary for Mandarin NLP, Mandarin YNLP, Mandarin YTD and Mandarin OTD ($p = .045$, $p = .002$, $p = .008$ and $p = .001$, respectively).

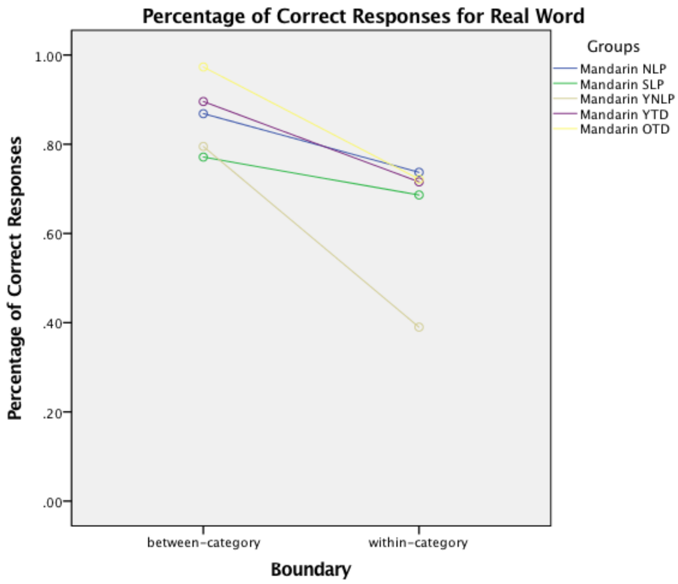

Figure 57. Percentage of correct responses for real word for between-category boundary and within-category boundary for five Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YNLP, Mandarin YTD, Mandarin OTD)

6.7.3 Comparison of English groups

A mixed ANOVA was performed with a within-subject factor, Boundary (between-boundary, within-boundary) and a between-subject factor, Group (English NLP, English SLP, English YTD, English OTD). The results indicated that there was a marginal main effect of Boundary F(1, 24) =

4.105, p = .054, but there was no main effect of Group F(3, 24) = 1.288, p = .301. There was no interaction between the within-subject factor, Boundary, and the between-subject factor, Group F(3, 24) = .432, p = .732.

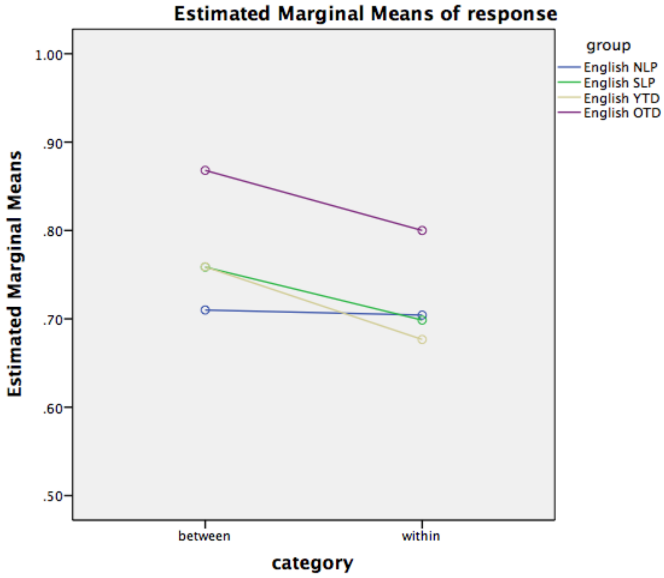

Figure 58. Percentage of correct responses for real word for between-category boundary and within-category boundary for four English groups (English NLP, English SLP, English YTD, English OTD)

7 Discussion of discrimination task results

Categorical perception of lexical tones

Q1: Do Mandarin-speaking children with autism spectrum disorder (ASD) categorically perceive and discriminate lexical tones in the same way as their typically-developing (TD) Mandarin-speaking counterparts?

In spite of the fact that the percentage of correct responses in Mandarin ASD was not significantly lower than in Mandarin TD in the two-step discrimination task across stimulus type, Table 30 and Figure 39 revealed the percentage of correct responses in Mandarin ASD was consistently lower than in Mandarin TD along the tone continuum except at endpoints of the tone continuum. When Mandarin ASD group was further split into three subgroups according to their linguistic abilities and ages, it turned out that

Mandarin YNLP, the younger group with ASD who had no language problems, had an overall lower result than other four Mandarin groups (Mandarin NLP, Mandarin SLP, Mandarin YTD, and Mandarin OTD) especially at the right end of the continuum (Figure 40). Although this did not reach the threshold of significance due to the power problems, it is an intriguing result to show that there is a potential delay in discrimination of lexical tones in ASD participants even without language problem. This should be subjected to further investigations in the future.

The Mandarin groups altogether showed the enhanced perception around the category boundary (Figure 39). The percentage of correct responses in within-category boundary (step 3-step 5, step 4-step 6, and step 5-step 7) was significantly lower than that of around category boundary (step 0-step 2 and step 1-step 3). Just as Figure 39 indicated, the peak was not as sharp as we might find for the categorical perception of consonant. Instead, its wide and shallow curve resembled more of the patterns found with vowels just as in Halle et al. (2004). Further investigations indicated that while Mandarin NLP, Mandarin YNLP, Mandarin YTD as well as Mandarin OTD indeed showed the significantly enhanced perception of category boundary, such significant improvement was not demonstrated in Mandarin SLP. In addition, the percentage of correct responses in between-category boundary was significantly lower in Mandarin SLP than which of in Mandarin OTD. These results altogether show that unlike other four Mandarin groups including Mandarin YNLP, Mandarin SLP had less prominent improvement around the category boundary and thus less pronounced categorical perception of lexical tones.

Role of native language

Q2: Do Mandarin- and English-speaking participants discriminate lexical tones differently?

Unlike the Mandarin listeners, the English participants did not perform significantly differently in between-category boundary and within-category boundary. Importantly, there was a significant interaction between Group

(Mandarin participants, English participants) and Step (step 0-step 2, step1-step 3, step 2-step 4, step3-5, step 4-step 6, step 5-step 7) (Figure 54). Just as indicated in Figure 56, while Mandarin participants had higher percentage of correct responses than English participants in between-category boundary, they had lower percentage of correct responses than the English listeners in within-category boundary. As a consequence, there was a sharp slope in the discrimination curve in Mandarin group, whereas the discrimination curve was comparatively flat in English participants. Just as the literature suggested, English participants perceive the non-native lexical tones in a psychoacoustic way and did not show the enhanced perception of category boundary as the listeners of tone languages (Figure 53 and Figure 58). It is worth noticing that Mandarin SLP perceive the native lexical tones just as the English participants and did not show the significant advantage for the categorical perception as native speakers.

Role of stimulus type (real word, nonce word, pure tone)

Q3: Do ASD children and the control groups discriminate the pitch contours of different kinds of stimuli (real words, nonce words, and pure tones) differently? Is there any interaction between the group and type of stimulus?

Mandarin ASD and TD both had marginally lower percentage of correct responses for real word than for nonce word or pure tone, especially in within-category boundary (Figure 48 and Figure 49). It was intriguing that compared to nonce word and pure tone, real word actually presented more difficulty in discriminating the lexical tones. This could be possibly attributed to the fact that the linguistic information provided by real word interfered with the discrimination. Speakers of tone languages tended to focus on the discrimination of lexical tones around the category boundary and ignore the subtle variations within the same category in order to process the tonal information efficiently. Therefore, when they were presented with real word stimuli in the discrimination task, they were still inclined to ignore the variations within the same category.

Role of orders

Q4: Do ASD children and their controls discriminate the pitch contours in ascending and descending orders differently? Do ASD children and their controls discriminate the pitch contours in AAB and ABB orders differently?

Mandarin ASD and TD groups did not discriminate the pitch contours in ascending and descending orders differently (Figure 50). However, they did discriminate the pitch contours differently in AAB and ABB orders (Figure 51). The percentage of correct responses was higher in AAB order than that of in ABB order. Similar results were also found in Chen and Kager (2011), and this effect was possibly due to the memory load. The identical pairs were the former two stimuli in AAB order, whereas the identical pairs were the last two stimuli in ABB order, the first condition might actually pose less difficulty for the participants' to discriminate the pitch contours of lexical tones.

Chapter 5: Conclusion

The current thesis has explored the tone processing and the acquisition of tone in Mandarin- and English-speaking typically developing children and children with Autism Spectrum Disorder with a series of experiments. The tone comprehension task for Mandarin participants in Chapter 2 showed that the Mandarin TD children and Mandarin ASD children were both able to perceive the four Mandarin tones at word level and differentiate lexical items with moderate accuracy based on the pitch contours. Mandarin NLP and SLP did not perceive the lexical tones significantly differently from Mandarin OTD. They displayed a similar rate of acquisition of tones in their comprehension without a significant delay in development (Table 3). That is, Mandarin NLP as well as Mandarin SLP had similar percentages of correct responses as Mandarin OTD. Also, Mandarin NLP and Mandarin SLP both had similar percentages of tonal errors, semantic errors, and phonetic errors as Mandarin OTD. In addition, the Mandarin ASD and TD children were both able to identify Tone 1 and Tone 4 easily and correctly, and yet had difficulty in perceiving Tone 2 and Tone 3. However, while Mandarin TD found the condition where the target was Tone 2 and the distractor was Tone 3 particularly difficult, Mandarin ASD had the opposite directional asymmetry.

In addition to the numerous tonal errors, the Mandarin ASD and TD children both made just a few semantic and phonetic errors. The correlation test indicated that the age of the Mandarin TD participants was correlated with the numbers of correct responses, tonal errors, semantic errors, but not with phonetic errors. On the other hand, the age of the Mandarin ASD was correlated with the numbers of correct responses, semantic errors, phonetic errors, but not with tonal errors. Further, the linguistic abilities of the Mandarin ASD participants were significantly correlated with the number of correct responses, tonal errors, semantic errors, as well as phonetic errors.

We can conclude from the results of the tone comprehension task that Mandarin ASD speakers do not have major problems with tone comprehension. They show the same order and rate of acquisition as their typically developing peers. At the same time, we identified some group-specific error patterns in an overreliance on Tone 2 over Tone 3, which is the opposite of the pattern found in typically developing children's comprehension. In future research, one could explore this further.

The psychoacoustic tone discrimination task for Mandarin and English participants in Chapter 3 indicated that Mandarin ASD, Mandarin TD, English ASD, and English TD were all sensitive to changes in pitch contours and did not show any significant difference. There was no evidence of enhanced pitch perception in the study of Tone 1-4 differences for ASD participants in English or Mandarin. There was also no evidence of enhanced perception for ASD children with language problems over ASD children with no language problems. All groups performed very well on the task, reaching very low thresholds. Therefore, it is possible that the lack of differentiation between the groups is due to a ceiling effect. This kind of ceiling effect might be avoided by including younger age group in the future research, especially given the literature review signals acquisition as earlier.

It was found that both Mandarin ASD and TD participants had a significantly higher threshold when tested on nonce words, compared to pure tone stimuli (Figure 14). We interpret this as nonce words presenting a more difficult task for native speakers. This effect may even be stronger for participants with ASD and a language problem. The difficulty of nonce word stimuli may come from the fact that no lexical supporting effect is present for nonce words, compared to real words. So, there is no top down effect from word recognition: participants must rely on their knowledge of the abstract tonal categories Tone 1 and Tone 4. If the representation of abstract tones is weaker, tone identification becomes harder. The results of Chapter 3 thus point to the direction that Mandarin SLP participants have weaker representations of abstract tones than their typically developing peers or Mandarin NLP.

In addition, the correlation tests indicated that the performance of real word and nonce word was strongly correlated in Mandarin OTD, while the threshold of nonce word did not only correlate with that of real word, but also correlated with that of pure tone in a even stronger way for Mandarin YTD and Mandarin ASD. Moreover, the participants with ASD with significant language problems (i.e. Mandarin SLP) only had significant correlation between the threshold of nonce word and pure tone.

These findings can be explained as follows. Mandarin OTD speakers have strong abstract representations of tones, thereby being able to treat real word and nonce word stimuli alike. They rely on general acoustic abilities to identify pure tone stimuli. Mandarin YTD and Mandarin ASD participants have weaker abstract representation of tones. As a result they treat nonce word stimuli partially like real word stimuli and partially like pure tone stimuli. They partly rely on their general acoustic abilities to identify nonce word stimuli. Mandarin SLP speakers treat nonce words only like pure tones stimuli, relying on their general acoustic abilities even more. This suggests that their grammatical representations of abstract tones are not stable enough.

The emerging picture is thus one where ASD participants fall in Scenario 2 from the Introduction with weaker representation of abstract tones than Mandarin OTD. However, it seems that Mandarin NLP pattern with Mandarin YTD. Just as discussed in Chapter 1, it is important to examine whether children with ASD display a delayed or a deviant developmental pattern of tone perception. Since Mandarin NLP pattern with Mandarin YTD, we see the occurrence of a developmental delay. In the case of Mandarin SLP, however, the behaviour suggests impairment over and above a developmental delay in the grammatical representation of abstract tones.

As for the English participants, there was no significant correlation at all between the three stimulus types. The results altogether revealed that the linguistic abilities in Mandarin might play an important role in perceiving the Mandarin lexical tones, reaching significance for nonce words. Also, the overall performance in the tone discrimination task was strongly correlated with the linguistic abilities (PPVT) for Mandarin ASD, whereas there was no significant correlation between the average of thresholds and the BPVS VMA or TROG VMA for English ASD.

The categorical perception tasks for Mandarin and English individuals in Chapter 4 consisted an identification task as well as a two-step discrimination task. The identification task for Mandarin participants showed that Mandarin ASD did not perform significantly differently in the identification of Mandarin lexical tones from Mandarin TD. They both perceived the lexical tones in a categorical way instead of a psychoacoustic way. While Mandarin YNLP had higher proportions of T4 responses than other groups for every step except step 0, Mandarin SLP participants were less inclined to identify the items as Tone 4 along the tone continuum.

The intercept points across the stimulus types for Mandarin ASD and TD (Mandarin YTD and Mandarin OTD) were both around step 2 and did not show any statistical difference. However, the intercept points for real word were marginally lower in Mandarin NLP than other groups. English ASD and English TD both had an intercept point around step 2 and did not differ significantly. While the intercept points in Mandarin overall participants and English overall participants were not significantly different, the intercept point in Mandarin TD was marginally higher than in English TD.

The slopes across stimulus types for Mandarin ASD and TD (Mandarin YTD and Mandarin OTD) were both around 3 and did not show any statistical difference. English ASD and English TD both had a slope around 3.5 and did not differ significantly. Although there was no significant difference in slopes around category boundary between the Mandarin and

English participants, there was a significant interaction between these two groups along the tone continuum in the identification curve. This showed that Mandarin participants tended to have a lower proportion of Tone 4 responses at low step-sizes than English participants, while they had a higher proportion of Tone 4 responses than English participants at high step-sizes. This makes for a steeper identification curve along the crucial part involving the category change. So, overall, although English participants certainly did not perceive steps along the Tone 1-4 in a purely psychophysical way, Mandarin speakers' identification curve had a more categorical step-function shape.

Mandarin ASD and TD generally had the highest percentage of T4 responses for pure tone, then for real word, and the lowest for nonce word. In addition, they both perceived these three kinds of stimuli in a categorical way. While these two groups both had higher intercept points of nonce word than that of real word and pure tone, the slopes did not differ between these three stimulus types.

The two-step discrimination task indicated that Mandarin SLP had significantly lower percentage of correct responses than Mandarin NLP, Mandarin YTD, and Mandarin OTD. Mandarin groups altogether had significantly better performance in between-category boundary than in within-category boundary, and thus demonstrated the enhanced perception around the category boundary. However, while Mandarin TD indeed showed the significantly enhanced perception of category boundary, which is a hallmark of categorical perception, such improvement did not demonstrated significantly in Mandarin NLP or SLP.[5]

[5] On the other hand, the English participants did not show such enhanced perception of category boundary. The percentage of correct response in between-category boundary was not significantly higher than that of in within-category boundary for English listeners due to the comparatively flat discrimination curve along the tone continuum. Just as the literature suggested, while Mandarin subjects perceived the native lexical tones in a categorical way, English participants perceive the non-native lexical tones in a psychoacoustic way.

These results fall in line with earlier results from Chapters 2 and 3 indicating subtle but persistent differences between the grammatical representation of tones for Mandarin ASD participants compared to their typically developing peers. Here we found that although their performance in the naming task was not distinguishable from their typically developing peers, nevertheless, the two-step identification task revealed a less strongly categorical perception of the Tone 1-4 continuum for the Mandarin ASD groups. The performance of the ASD SLP groups was also overall worse. To the extent that categorical perception tasks tap into the nature of the grammatical representation of tones, the results of Chapter 4 suggest significant issues for the ASD population for this. If these results are on the right track, then they could possibly substantiate a grammatical impairment of this population, which was hitherto uncovered, due to the lack of research on tones in this population. The results point to the general direction that potentially people living with ASD might have prosodic impairments relating their pitch perception, and their ability to categorise pitch contours in a grammatical fashion, in addition to their pragmatic difficulties.

As for the future directions, the current study only utilized Tone 1 - Tone 4 pair, the easiest tone pair to distinguish in Mandarin, as the experimental stimuli. As a consequence, the follow-up study could use Tone 2 - Tone 3 pair, the most difficult pair to discriminate in Mandarin, as the experimental stimuli to avoid the ceiling effect and further explore the tone processing and the acquisition of tone in typically developing children and children with ASD. Furthermore, since Tone 1 (High) - Tone 4 (Falling) pair and Tone 2 (Rising) - Tone 3 (Low) pair are both the combination of a level tone and a contour tone, it would be interesting to explore Tone 2 - Tone 4 pair, the combination of two contour tones, to see how participants would perceive and discriminate the contour tone pair.

In addition to the low-level pitch perception, investigations on higher level linguistic functions of prosody could also be included in the future studies to chart the full territory along the level of abstractness. If it becomes clear that children with ASD can properly perceive tones, then it is important to compare and contrast the processing of lexical tone and focal accent with eye tracking facilities and electroencephalography (EEG) to see the real-time responses. If it turns out that the participants can in fact process tones without difficulty, then the next task of the research will be examining the role of focus in the phrasal/sentential domain to further explore the communication function in ASD. The texperiment could employ the design of Szendroi et al. (2010) developed for testing the abilities of typically developing children in production and comprehension of focal accent as well as focal meaning.

Further, it would be intriguing to explore the potential delay by a larger younger ASD group with and without language impairment. The current study only managed to recruit younger ASD group without language problem (Mandarin YNLP) from six to ten years old. Therefore, in the future it would be enlightening to compare four ASD groups according to their ages and language abilities: older NLP, older SLP, younger NLP, as well as younger SLP. In this wise, it would provide a clear and thorough picture for the interaction between the ages and language abilities in children with ASD.

All in all, further studies are greatly needed to explore more participants in different languages, clinical backgrounds, language abilities, as well as ages in a larger scale to chart the full territory along the level of abstractness. More importantly, the effort could be made on finding individually matched TD for each participant with ASD for a more precise and clear result in the future studies.

References

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech, 21*, 319–325.

Alcantara, J. I., 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. doi:10.1016/j.neuropsychologia.2012.01.034

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

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Arthur S. Abramson. (1977). Noncategorical perception of tone categories in Thai. *The Journal of the Acoustical Society of America, 61*, S66.

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *J Autism Dev Disord, 38*(7), 1230-1240. doi:10.1007/s10803-007-0504-z

Bailey, P., & Snowling, M. (2002). Auditory processing and the development of language and literacy. *British Medical Bulletin, 63*(1), 135-146.

Baltaxe, C. (1984). Use of contrastive stress in normal, aphasic and autistic children. *Journal of Speech and Hearing Research, 27*, 97-105.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The Autism-Spectrum Quotient (AQ)--adolescent version. *J Autism Dev Disord, 36*(3), 343-350. doi:10.1007/s10803-006-0073-6

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very

High Functioning Adults with Autism or Asperger Syndrome. *Jornal of Child Psychology and Psychiatry, 38*(7), 813-822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition, 21*(1), 37-46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord, 31*(1), 5-17.

Bhatara, A., Babikian, T., Laugeson, E., Tachdjian, R., & Sininger, Y. S. (2013). Impaired timing and frequency discrimination in high-functioning autism spectrum disorders. *J Autism Dev Disord, 43*(10), 2312-2328. doi:10.1007/s10803-013-1778-y

Bishop, D. V. M. (1983). *Test for Reception of Grammar* (TROG). Manchester, UK: Chapel Press.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin tone 2/tone 3distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*, 37-49.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International, 5*, 341– 345.

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

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *J Child Psychol Psychiatry, 41*(7), 847-857.

Brock, J., Bzishvili, S., Reid, M., Hautus, M., & Johnson, B. W. (2013). Brief report: atypical neuromagnetic responses to illusory auditory pitch in children with autism spectrum disorders. *J Autism Dev Disord, 43*(11), 2726-2731. doi:10.1007/s10803-013-1805-z

Burnham, D., & Jones, C. (2002). *Categorical perception of lexical tones by tonal and non-tone language speakers*. Paper presented at the 9th Australian International Conference on Speech Science & Technology, Melbourne.

Ceponiene, R., Lepisto, T., Shestakova, A., Vanhala, R., Alku, P., Naatanen, R., & Yaguchi, K. (2003). Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc Natl Acad Sci U S A, 100*(9), 5567-5572. doi:10.1073/pnas.0835631100

Chao, Y. R. (1973). The Cantian idiolect: An analysis of the Chinese spoken by a twenty-eight-month-old child. In C.A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*, 13-33. New York: Holt. (Original work published 1951).

Charman, T. (2003). Why Is Joint Attention a Pivotal Skill in Autism? *Philosophical Transactions: Biological Sciences, 358*(1430), 315-324.

Chen, A. (2013). *Universal biases in the perception of Mandarin tones: From infancy to adulthood*. Doctoral thesis, University Utrecht.

Chen, A, and Kager, R. (2011). The perception of lexical tones and tone sandhi in L2: Success or failure?" In W.S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences ICPhS XVII 2011*, 444-447. Hong Kong: City University of Hong Kong.

Chen, A, and Kager, R. (2013). Perceptual constraints on the early discrimination of lexical tones. To appear in BUCLD 37: *Proceedings of the 37nd annual conference on language development*. Somersville: Cascadilla Press.

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

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2009). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with Asperger Syndrome. *Research in Autism Spectrum Disorders, 3*(2), 502-516. doi:10.1016/j.rasd.2008.10.004

Chevallier, C., Wilson, D., Happe, F., & Noveck, I. (2010). Scalar inferences in Autism Spectrum Disorders. *J Autism Dev Disord, 40*(9), 1104-1117. doi:10.1007/s10803-010-0960-8

Cilibrasi, L., Stojanovik, V., & Riddell, P. (2015). Word position and stress effects in consonant cluster perception and production. *Dyslexia, 21*(1), 50-59. doi:10.1002/dys.1488

Cleland, J., Gibbon, F., 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.

Clumeck, H. V. (1977*). Studies in the acquisition of Mandarin phonology*. Unpublished doctoral dissertation, University of California, Berkeley.

Collet, L., Roge, B., Descouens, D., Moron, P., Duverdy, F., & Urgell, H. (1993). Objective auditory dysfunction in infantile autism. [Letter]. *Lancet, 342*(8876), 923-924.

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

Cutler, A., Oahan, D., & van Donselaar, W. Prosody in the comprehension of spoken language: a literature review. (1997). *Language and Speech, 40*(2), 141-201.

DeMyer, M., Bartons, S., DeMyer, W., Norton, J., Allen, J., & Stelle, R. (1973). Prognosis in autism: A follow-up study. *Journal of Autism and Childhood Schizophrenia, 3*, 199-246.

DePape, A. M., Hall, G. B., Tillmann, B., & Trainor, L. J. (2012). Auditory processing in high-functioning adolescents with Autism Spectrum Disorder. *PLoS One, 7*(9), e44084. doi:10.1371/journal.pone.0044084

Dodd, Z. H. B. (2015). Putonghua (modern standard Chinese)-speaking children with speech disorder. *Clinical Linguistics & Phonetics, 14*(3), 165-191. doi:10.1080/026992000298814

Dolan, E. (2014). *Discrimination and categorical perception of lexical tones in autism spectrum disorder*. (Master), University College London.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. (PPVT-R). Circle Pines, MN: American Guidance Service.

Dunn, L. M., Whetton, C., & Pintilie, D. (1997). *British Picture Vocabulary Scale*. Windsor, UK: NFER-Nelson.

Fay, W., Schuler, A. (1980). *Emerging language in Autistic Children*. London: Edward Arnold.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America, 89*, 412-424.

Findlay, J. M. (1978). Estimates on probability functions: A more virulent PEST. *Perception and Psychophysics, 23*, 181–185.

Fine, J., Bartolucci, G., Ginsberg, G., Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 32*, 771-782.

Fodor, J., & Sag, D. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy, 5*(3), 355-398.

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*(Pt 12), 2703-2709. doi:10.1093/brain/awg274

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268-294. doi:10.1016/j.wocn.2007.06.005

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Attention, Perception, & Psychophysics, 65*(7), 1029-1044.

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

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). Identification and discrimination of synthetic vowels. *Language & Speech, 5*, 171-189.

Frye, R. E., & Beauchamp, M. S. (2009). Receptive language organization in high-functioning autism. *J Child Neurol, 24*(2), 231-236. doi:10.1177/0883073808329572

Fu, Q.-J., & Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing, 5*, 45-57.

Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey*, 71-76. New York: Academic Press.

Gandour, J. (1984). Tone perception in far eastern languages. *Journal of Phonetics, 11*(2), 149-175.

Gandour, J., & Dardarananda, R. (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language, 18*, 98–114.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., & Luritot, J.. (2001). Differential neural circuitry underlies the perception of Chinese tone and intonation: An fMRI study. *Neuroimage, 13*(6), 533.

Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1–33.

Gandour, J., Petty, S., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language, 35*, 201–240.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, D. (2000). A cross-linguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12*, 207–222.

Gandour, J., Wong, D., & Hutchins, D. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport, 9*, 2115–2119.

Gao, J., Shi, R., & Li, A. (2010). *Categorization of lexical tones in Mandarin-learning infants*. Paper presented at the Speech Prosody 2010, Chicago.

Gao, J., Shi, R., & Li, A. (2011). Lexical tone perception in non-tone-learning infants. *12th International Congress for the Study of Child Language*, Montreal.

Garding, E., Kratochvil, P., & Svantesson, J. (1986). Tone 4 and tone 3 discrimination in modern standard Chinese. *Language and Speech, 29*, 281-293.

Gauthier, B., & Shi, R. (2011). A connectionist study on the role of pitch in infant-directed speech. *J Acoust Soc Am, 130*(6), El380-386. doi:10.1121/1.3653546

Gauthier, B., Shi, R., & Xu, Y. (2006). Learning to categorize lexical tones in sentences with variable narrow focus: A neural network approach. *The Journal of the Acoustical Society of America, 119*, 3420.

Gauthier, B., Shi, R., & Xu, Y. (2007). Simulating the acquisition of lexical tones from continuous dynamic input. *J Acoust Soc Am, 121*(5 Pt1), El190-195.

Ghaziuddin, M., & Gerstein, L. (1996). Pedantic speaking style differentiates Asperger syndrome from high-functioning autism. *Journal of Autism and Developmental Disorders, 26*, 585-595.

Ghaziuddin, M., Thomas, P., Napier, E., Kearney, G., Tsai, L., Welch, K., & Fraser, W. (2000). Brief report: Brief syntactic analysis in Asperger syndrome: a preliminary study. *J Autism Dev Disord, 30*(1), 67-70.

Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., & Pearce, P. S. (2001). Children with a history of expressive vocabulary delay: Outcomes at 5 years of age. *American Journal of Speech-Language Pathology*, *10*(4), 358.

Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Atten Percept Psychophys, 75*(8), 1799-1810. doi:10.3758/s13414-013-0518-x

Gomot, M., Bernard, F. A., Davis, M. H., Belmonte, M. K., Ashwin, C., Bullmore, E. T., & Baron-Cohen, S. (2006). Change detection in children with autism: an auditory event-related fMRI study. *Neuroimage, 29*(2), 475-484. doi:10.1016/j.neuroimage.2005.07.027

Haesen, Birgitt, Boets, Bart, & Wagemans, Johan. (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.

Hallé, P. (1994). Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese. *Language and Speech, 37*, 103–124.

Hallé, P., Best, C., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics, 27*, 281–306.

Hallé, P., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421. doi:10.1016/s0095-4470(03)00016-0

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tone language speakers. *Journal of Phonetics, 40*, 269-279.

Happe, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences, 3*, 216– 222.

Happe, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J Autism Dev Disord, 36*(1), 5-25. doi:10.1007/s10803-005-0039-0

Heaton, P., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cogn Neuropsychol, 25*(6), 771-782. doi:10.1080/02643290802336277

Hoffmann, C. W. G., Sadakata, M., Chen, A., Desain, P., & McQueen, J. M. (2014). *Within-Category Variance and Lexica Tone Discrimination in Native and Non-Native speakers*. Paper presented at the 4th International Symposium on Tonal Aspects of Languages.

Huang, T. (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (OSU working papers in linguistics, No. 55), 23-42.

Huang, T. (2004). Language-*specificity in auditory perception of Chinese tones*. Unpublished doctoral dissertation. The Ohio State University, Ohio.

Huang, T. (2007). *Perception of Mandarin tones by Chinese- and English-speaking listeners*. Paper presented at the International Congress of Phonetic Science.

Huang, T. & Johnson, K. (2010). Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. *Phonetica, 67*(4), 243–67.

Hume, E., & Johnson, K. (2003). *The impact of partial phonological contrast on speech perception*. Paper presented at the 15th International Congress of Phonetic Sciences.

Jansson-Verkasalo, E., Ceponiene, R., Kielinen, M., Suominen, K., Jantti, V., Linna, S. L., . . . Naatanen, R. (2003). Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neurosci Lett, 338*(3), 197-200.

Jansson-Verkasalo, E., Kujala, T., Jussila, K., Mattila, M. L., Moilanen, I., Naatanen, R., . . . Korpilahti, P. (2005). Similarities in the phenotype of the auditory neural substrate in children with Asperger syndrome and their parents. *Eur J Neurosci, 22*(4), 986-990. doi:10.1111/j.1460-9568.2005.04216.x

Jarvinen-Pasley, A., Pasley, J., & Heaton, P. (2008). Is the linguistic content of speech less salient than its perceptual features in autism? *J Autism Dev Disord, 38*(2), 239-248. doi:10.1007/s10803-007-0386-0

Jarvinen-Pasley, A., Wallace, G. L., Ramus, F., Happe, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Dev Sci, 11*(1), 109-121. doi:10.1111/j.1467-7687.2007.00644.x

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia, 48*(9), 2630-2639.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Mem Cognit, 40*(7), 1109-1121. doi:10.3758/s13421-012-0208-2

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of Melodic Contour and Intonation in Autism Spectrum Disorder: Evidence From Mandarin Speakers. *J Autism Dev Disord, 45*(7), 2067-2075. doi:10.1007/s10803-015-2370-4

Jones, C. R., Happe, F., Baird, G., Simonoff, E., Marsden, A. J., Tregay, J., . . . Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia, 47*(13), 2850-2858. doi:10.1016/j.neuropsychologia.2009.06.015

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neurosci, 9*, 53. doi:10.1186/1471-2202-9-53

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: an event-related potential study. *Brain Res, 1148*, 113-122. doi:10.1016/j.brainres.2007.02.019

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

Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S., . . . Schroeder, J. L. (2006). The pattern of sensory processing abnormalities in autism. *Autism, 10*(5), 480-494. doi:10.1177/1362361306066564

Khalfa, S., Bruneau, N., Roge, B., Georgieff, N., Veuillet, E., Adrien, J. L., . . . Collet, L. (2001). Peripheral auditory asymmetry in infantile autism. [Research Support, Non-U.S. Gov't]. *Eur J Neurosci, 13*(3), 628-632.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *J Autism Dev Disord, 21*(1), 29-42.

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. *Dev Sci, 8*(1), F1-f12. doi:10.1111/j.1467-7687.2004.00384.x

Kujala, T., Lepisto, T., & Naatanen, R. (2013). The neural basis of aberrant speech and audition in autism spectrum disorders. *Neurosci Biobehav Rev, 37*(4), 697-704. doi:10.1016/j.neubiorev.2013.01.006

Lee, Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. *J Psycholinguist Res, 25*(5), 527-542.

Lepisto, T., Kuitunen, A., Sussman, E., Saalasti, S., Jansson-Verkasalo, E., Nieminen-von Wendt, T., & Kujala, T. (2009). Auditory stream segregation in children with Asperger syndrome. *Biol Psychol, 82*(3), 301-307. doi:10.1016/j.biopsycho.2009.09.004

Lepisto, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Naatanen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Res, 1066*(1-2), 147-157. doi:10.1016/j.brainres.2005.10.052

Lepisto, T., Nieminen-von Wendt, T., von Wendt, L., Naatanen, R., & Kujala, T. (2007). Auditory cortical change detection in adults with Asperger syndrome. *Neurosci Lett, 414*(2), 136-140. doi:10.1016/j.neulet.2006.12.009

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

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., & Dzemidzic, M. (2010). Hemispheric asymmetries in phonological processing of tones versus segmental units. *Neuroreport, 21*(10), 690-694. doi:10.1097/WNR.0b013e32833b0a10

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*(02), 185-199

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.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61*(5), 379-388.

Lin B.-G., Chi P.-H. (2007). *The Assessment of School Language Disorders*. Taipei, National Taiwan Normal UP.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Dev Psychol, 43*(4), 912-917. doi:10.1037/0012-1649.43.4.912

Liu, L. (2010). *Is perceptual reorganization affected by statistical learning? Dutch infants' sensitivity to lexical tone*. Unpublished Master thesis, Utrecht University.

Liu, L., & Kager, R. (2011a). Is perceptual reorganization affected by statistical learning? Dutch infants' sensitivity to lexical tones. In N. Danis, K. Mesh & H. Sung (eds.), *BUCLD 35: Proceedings of the 35nd annual conference on language development*, 404-413. Somersville: Cascadilla Press.

Liu, L. & Kager, R. (2011b). How do statistical learning and perceptual reorganization alter Dutch infant's perception to lexical tones?" *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS), 1270-1273, Hong Kong.

Liu, L., & Kager, R. (2012). Non-native tone perception from infant to adult: How consistent and flexible is it? *Proceedings of Speech Prosody 2012*, Shanghai, China.

Liu, L, and Kager, R. (2013). How does bilingualism alter infants' tone perception during perceptual reorganization? To appear in *BUCLD*

37: Proceedings of the 37nd annual conference on language development. Somersville: Cascadilla Press.

Lu, Y. (2011). Perception and representation of Mandarin fricatives. *LSA Annual Meeting Entended Abstracts, 2*, 22:1-5.

Makarova, A. O. (2010). *Acquisition of three vowel contrasts by Russian speakers of American English*. Doctoral thesis, Harvard University.

Massaro, D. W., Cohen, M. M., Tseng, C.-Y. (1985). The evaluation and integration of pitch height and pitch contour. *Journal of Chinese Linguistics, 13*, 267-290.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367-1381. doi:10.1016/j.cognition.2007.07.002

McCann, J., & Peppe, S. (2003). Prosody in autism spectrum disorders: a critical review. [Review]. *Int J Lang Commun Disord, 38*(4), 325-350.

Mirak, J., & Rescorla, L. (1998). Phonetic skills and vocabulary size in late talkers: Concurrent and predictive relationships. *Applied Psycholinguistics*, *19*, 1–18.

Morton, J. M., Haith, M. M., & Gibson, J. J. (1976). The Relation between Audition and Vision in the Human Newborn. *Monographs of the Society for Research in Child Development, 41*(4), 1-72.

Mottron, L., & Burack, J. (2001). Enhanced perceptual functioning in the development of persons with autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131– 148). Mahwah, NJ: Lawrence Erlbaum Associates.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *J Autism Dev Disord, 36*(1), 27-43. doi:10.1007/s10803-005-0040-7

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neurosci Biobehav Rev, 36*(2), 836-854. doi:10.1016/j.neubiorev.2011.11.008

Ouimet, T., Foster, N. E., Tryfon, A., & Hyde, K. L. (2012). Auditory-musical processing in autism spectrum disorders: a review of behavioral and brain imaging studies. *Ann N Y Acad Sci, 1252*, 325-331. doi:10.1111/j.1749-6632.2012.06453.x

Paul, R., Shriberg, L., McSweeneym J., Cicchetti, D., Klin, A., &Volkmar, F. (2005). Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35*, 205-220.

Pisoni, D. (1971). *On the nature of categorical perception of speech sounds*. Unpublished doctoral dissertation, Michigan University at Ann Arbor.

Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, & Psychophysics, 13*(2), 253-260.

Pisoni, D., & Lazarus, J. (1975). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55*, 328–333.

Pisoni, D., & Tash, J. (1974). Reaction times to comparisons within and across phonetic boundaries. *Perception and Psychophysics, 15*, 285–290.

Peppé, S., Cleland, J., Gibbon, F., O'Hare, A., & Castilla, P. M. (2011). Expressive prosody in children with autism spectrum conditions. *Journal of Neurolinguistics, 24*(1), 41-53. doi:10.1016/j.jneuroling.2010.07.005

Peppe, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. *Clin Linguist Phon, 17*(4-5), 345-354.

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.

Perovic, A., Modyanova, N., & Wexler, K. (2013). Comparison of Grammar in Neurodevelopmental Disorders: The Case of Binding in Williams Syndrome and Autism With and Without Language Impairment. *Lang Acquis, 20*(2), 133-154. doi:10.1080/10489223.2013.766742

Peters, A. M. (1983). *The Units of Language Acquisition*. New York: Cambridge UP.

Pexman, P. M., Rostad, K. R., McMorris, C. A., Climie, E. A., Stowkowy, J., Glenwright, M. R. (2011). Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of Autism and Developmental Diorders, 41*(8); 1097-1112.

Qin, Z., & Mok, P. (2012). *The perception of speech and non-speech tones by tone and non-tone language listeners*. Paper presented at the Speech Prosody 2012, Shanghai.

Rapin, I. (1991). Autistic children: diagnosis and clinical features. [Research Support, U.S. Gov't, P.H.S. Review]. *Pediatrics, 87*(5 Pt 2), 751-760.

Repp, B. (1984a). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and Speech, 27*(3), 245-254.

Repp, B. (1984b). Categorical perception: Issues, methods, findings. *Speech and language: Advances in Basic Research and Practice, 10*, 243–335.

Repp, B., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18*, 481–495.

Rundblad, G., & Annaz, D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism, 14(1)*, 29-46.

Sanders, R. (2008). *Tonetic sound change in Taiwan Mandarin: the case of Tone 2 and Tone 3 citation contours*. Paper presented at the 20th North American Conference on Chinese Linguistics (NACCL-20).

Schneider, K. (2012). *The German boundary tones: Categorical perception, perceptual magnets, and the perceptual reference space*. Doctoral thesis, University Stuttgart.

Shen, X. (1990). On Mandarin tone 4. *Australian Journal of Linguistics, 10*, 41-59.

Shen, X., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*, 145-156.

Shi, R., Gao, J., & Li, A. (2011). *Perception of lexical tones in infants*. Poster presentation in 36th Annual Boston University Conference on Child Language Development, Boston. , 36.

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

Siegal, M., & Blades, M. (2003). Language and auditory processing in autism. *Trends Cogn Sci, 7*(9), 378-380.

Simmons, J., & Baltaxe, C. (1975). Language patterns in adolescent autistics. *Journal of Autism and Childhood Schizophrenia 5*, 333–351.

Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., & Place, M. (2004). The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry, 43*(5), 548-558. doi:10.1097/00004583-200405000-00008

Soulieres, I., Mottron, L., Saumier, D., & Larochelle, S. (2007). Atypical categorical perception in autism: autonomy of discrimination? *J Autism Dev Disord, 37*(3), 481-490. doi:10.1007/s10803-006-0172-4

Stanutz, S., Wapnick, J., & Burack, J. A. (2014). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism, 18*(2), 137-147. doi:10.1177/1362361312462905

Sutcliffe, P., & Bishop, D. (2005). Psychophysical design influences frequency discrimination performance in young children. *J Exp Child Psychol, 91*(3), 249-270. doi:10.1016/j.jecp.2005.03.004

Tager-Flusberg H. (1999). A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry, 11*, 325–334.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, *41*, 782–787.

Tsao, F. M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology, 50*(2), 111-124.

Tseng, H.-C. (2014). *Categorical perception of lexical tones in Mandarin-speaking children at age of 6*. Unpublished Master thesis, University College London.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition, 18*(6), 559-567.

van der Lely, H. K. (2005). Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends Cogn Sci, 9*(2), 53-59. doi:10.1016/j.tics.2004.12.002

Vihman, M. M. (1996). *Phonological Development*. Oxford: Blackwell.

Wang, Y, Jongman, A., Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone production before and after perceptual training. *Journal of the Acoustical Society of America, 113*, 1033-1043.

Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A. D., & Obrig, H. (2007). The processing of prosody: Evidence of interhemispheric specialization at the age of four. *Neuroimage, 34*(1), 416-425. doi:10.1016/j.neuroimage.2006.09.009

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition Manual*. Bloomington, MN: Pearson.

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. *Cogn Neuropsychol, 23*(5), 748-761. doi:10.1080/02643290500438607

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

Wong, P. (2011). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*, 141-151.

Wong, P. (2012a). Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and in the United States: interjudge reliability and perceptual results. *J Speech Lang Hear Res, 55*(5), 1423-1437. doi:10.1044/1092-4388(2012/11-0273)

Wong, P. (2012b). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141-151.

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *J Acoust Soc Am, 133*(1), 434-443. doi:10.1121/1.4768883

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *J Speech Lang Hear Res, 48*(5), 1065-1079. doi:10.1044/1092-4388(2005/074)

Wong, P., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*, 420-422.

Wong, Patrick C. M., & Perrachione, Tyler K. (2007). Learning Pitch Patterns in Lexical Identification by Native English-Speaking Adults. *Applied Psycholinguistics, 28*(4), 565-585.

World Health Organization. (1993). *Mental disorders: A glossary and guide to their classification in accordance with the 10th revision of the international classification of disease-research diagnostic criteria* (ICD-10). Geneva: WHO.

Wu, X. & Lin, H., 2008. Perception of Mandarin Tones by Mandarin and English Listeners. *Journal of Chinese Language and Computing, 18*(4), pp.175–187.

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. doi:10.1016/j.neuroscience.2010.06.077

Xu, L., Li, Y., Hao, J., Chen, X., Xue, S. A., & Han, D. (2004). Tone production in Mandarin-speaking children with cochlear implants: a preliminary study. *Acta Otolaryngol, 124*(4), 363-367.

Xu, Y. (1994). Production and perception of coarticulated tones. Journal of the *Acoustical Society of America, 95*(4), 2240-2253.

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

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*, 179-203.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics, 5*, 757-797.

Xu, Y., Chen, S.-w., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review, 29*(1). doi:10.1515/tlr-2012-0006

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America, 120*(2), 1063-1074.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33*, 319–337.

Xu, Y., & Whalen, D. H. (1990). Information for Mandarin tones in the amplitude contour and in brief segments. *The Journal of the Acoustical Society of America, 87*, S46.

Yang, J., & Liu, C. (2012). Categorical perception of lexical tone in 6 to 8-year-old monolingual and bilingual children. *International Journal of Asian Language Processing, 22*(2), 49-62.

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., & Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *J Child Psychol Psychiatry, 53*(8), 874-882. doi:10.1111/j.1469-7610.2012.02528.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. doi:10.1080/01690965.2010.520493

Zhou, N., & Xu, L. (2008). Development and evaluation of methods for assessing tone production skills in Mandarin-speaking children with cochlear implants. *J Acoust Soc Am, 123*(3), 1653-1664. doi:10.1121/1.2832623

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon, England, Multilingual Matters.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*, 3-24.

Zhu, H., & Dodd, B. (2006). *Phonological development and disorders in children : A multilingual perspective* (Child language and child development ; 8). Clevedon, [England] ; Buffalo: Multilingual Matters.