Preprint Manuscript Accepted for Publication in European Journal of Neuroscience

DOI: http://onlinelibrary.wiley.com/doi/10.1111/ejn.13657/full

Distinct patterns of discrimination and orienting for temporal processing of speech and nonspeech in Chinese children with autism: An event-related potential study

Dan Huang[1,5], Luodi Yu[1,4], Xiaoyue Wang[1], Yuebo Fan[5], Suiping Wang[1,2,3*], Yang Zhang[4,6*]

1. School of Psychology, South China Normal University, Guangzhou, 510631, China
2. Center for Studies of Psychological Application, South China Normal University, 510631, China
3. Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou, 510631, China
4. Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN, 55455, USA
5. Guangzhou Rehabilitation and Research Center for Children with Autism, Guangzhou Cana School, Guangzhou, 510540, China
6. Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, 55455, USA

***Correspondence to** Suiping Wang, School of Psychology, South China Normal University, Guangzhou, 510631, China. Phone: (86) 20-85216510; Fax: (86)20-85216412, E-mail: wangsuiping@m.scnu.edu.cn
Or to Yang Zhang, Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN, 55455, USA. Phone: 1 612-624-7818; Fax: 1 612-624-7586, E-mail: zhanglab@umn.edu

Running title: Duration perception in children with autism

Keywords: Autism, temporal processing, speech perception, mismatch negativity (MMN), perceptual weighting

Abstract

Although many studies have reported domain-general impaired duration perception for speech and nonspeech sounds in children with autism, it remained unclear whether this phenomenon is universally applicable regardless of language background. In some languages such as Finnish and Japanese, vowel duration serves a phonemic role that can signify semantic distinction, and in others (e.g., Mandarin Chinese), vowel duration does not carry this phonemic function. The present event-related potential study investigated neural sensitivity to duration contrasts in speech and nonspeech contexts in Mandarin-speaking children with autism and a control group of age-matched typically developing (TD) children. A passive oddball paradigm was adopted to elicit the mismatch negativity (MMN) and involuntary orienting response (P3a) for change detection. A pure tone condition and a vowel condition were used. The MMN results showed that the autism group had diminished response amplitudes and delayed latency in the pure tone condition compared to the TD group, whereas no group difference was found in the vowel condition. The P3a results showed no significant between-group MMN difference in the pure tone condition. In the vowel condition, the autism group had smaller P3a than the TD group. Together, the distinct patterns of discrimination and orienting responses for duration contrasts in pure tones and vowels are consistent with the "allophonic perception" theory for autism, which may reflect a compromised perceptual weighting system for speech learning.

Introduction

Autism spectrum disorder (ASD) refers to a group of early-onset neuro-developmental disorders that are mainly characterized by poor social reciprocity and communication, combined with restrictive and repetitive patterns of behavior and interests (American Psychiatric Association, 2013). Some researchers found that despite impairments in language and social communication, there are high occurrences of enhanced perceptual abilities in auditory and visual domains in individuals with autism (Chen *et al.*, 2012; Pallett *et al.*, 2014). For instance, individuals with autism outperformed typically developing (TD) controls in pitch perception regardless of speechness or complexity of the testing material (Bonnel *et al.*, 2003; Heaton, 2005; Järvinen-Pasley *et al.*, 2008). However, other behavioral and neurophysiological studies have revealed that individuals with autism may exhibit deficits in sound duration perception (Szelag *et al.*, 2004; Lepistö *et al.*, 2005; Martin *et al.*, 2010; Maister & Plaisted-Grant, 2011; Falter *et al.*, 2012; Brodeur *et al.*, 2014). Together, the atypical pitch and duration perception results demonstrate that children with autism have perceptual weighting strategies that are different from typically developing children for basic spectral and temporal cues that are critical for recognizing speech sounds. Given the fundamental role of sensory perception in early development, it is of theoretical and practical importance to evaluate the underlying causes and the impact of atypical auditory skills on speech perception and verbal communication in autism (Lepistö *et al.*, 2006; Samson *et al.*, 2006; Hitoglou *et al.*, 2010; Kujala *et al.*, 2013).

An important issue here is to determine whether atypical auditory processing in autism reflects a domain-general (i.e., language non-specific) perceptual deviation or domain-specific alteration in perceptual sensitivity that is shaped by language experience. If the domain-general deficit hypothesis is correct, we would expect to observe a universal pattern of atypical basic auditory sensitivity with enhanced spectral processing and impaired temporal processing across typologically different languages such as Finnish, English, Japanese, and Chinese. If this domain-general hypothesis is incorrect, we would expect to observe different patterns of atypical spectral and temporal processing that are language-dependent in children with autism.

As speech sounds are cued by different acoustic features and different languages employ a different set of sounds and phonological contrasts, speech perception deficits in autism may take different forms depending on the sound category and language. Speech acquisition in normal listeners involves a largely implicit neural commitment process to map out the phonological system of their native language by perceptually "tuning out" irrelevant acoustic information (Kuhl, 2004). Unlike typically developing children, individuals with autism who have a different perceptual weighting system of spectral and temporal cues are expected to have deficits in representing the phonological categories if something went wrong in the early development of native language neural commitment (NLNC). Recent studies examining pitch perception in the context of tonal language have provided insightful information regarding this issue. In tonal languages, such as Chinese and Thai, pitch is used to differentiate word meaning at the syllabic level (Yip, 2002). Unlike the previously reported domain-general pitch superiority in autism (Jarvinen-Pasley & Heaton, 2007; Haesen *et al.*, 2011), Chinese-speaking children with autism showed enhanced mismatch negativities (MMNs) to pitch contrasts in nonspeech stimuli but diminished responses to lexical tone changes (Yu *et al.*, 2015). The MMN reflects pre-attentive automatic detection of acoustic stimulus change, which is correlated with perceptual discrimination ability (Näätänen *et al.*, 2011). A follow-up study has further demonstrated a categorical perception deficit for lexical tones in that hypersensitivity to pitch variation in Chinese children with autism may have interrupted their proper acquisition of language-specific tonal categories, which requires the inhibition of neural sensitivity to within-category differences and heightened sensitivity to across-category differences (Wang *et al.*, 2017). Additionally, Jiang et al. (2015) examined whether tonal language experience would alleviate deficits in speech intonation perception in Chinese-speaking children with autism using behavioral discrimination and identification tasks. Despite their superior performance in melodic contour condition, children with autism showed impaired intonation perception compared with the TD controls (Jiang *et al.*, 2015). The co-existence of "pitch superiority" for non-speech stimuli and language-specific pitch processing deficit in Mandarin-speaking children with autism appears to indicate NLNC abnormality for the higher-order phonological processing of lexical tones (Wang et al., 2017; Yu et al., 2015). These results suggest domain-specific pitch perception deficit in autism in relation to the tonal language background.

In contrast to the domain-specific pitch processing deficits for speech stimuli in Chinese children with autism, a number of studies have reported domain-general duration perception deficits for both speech and nonspeech stimuli in individuals with autism whose native languages employ vowel length contrast phonemically to mark semantic distinction (Kasai et al., 2005; Lepistö et al., 2005, 2006). For example, /tuli/ means "fire" in Finnish whereas /tuuli/ means "wind", and detecting/producing the phonemic length difference in the first vowel here is critical for the proper understanding and use of these two Finnish words. A study of Finnish children with autism found diminished MMNs to the phonemic Finnish vowel duration contrast as well as duration change in nonspeech sounds in comparison with age-matched TD controls (Lepistö *et al.*, 2005). In a follow-up study by the same research team, Finnish children with Asperger Syndrome (AS) did poorer in detecting vowel duration changes as indexed by lower hit rate and longer reaction time along with diminished MMN amplitude (Lepistö *et al.*, 2006). In Japanese-speaking adults with autism, Kasai *et al.* (2005) observed a similar trend of reduced mismatch field (MMF) response to sound duration change. Along with other evidence showing deficits in detecting auditory temporal features in individuals with autism (Alcantara *et al.*, 2004; Groen *et al.*, 2009), it is tantalizing to consider lower-level (i.e., language non-specific) temporal processing deficit as an underlying cause, which might reflect altered brain networks for auditory temporal processing in autism (Foss-Feig *et al.*, 2012).

As the vowel duration perception deficit in autism has only been found in speakers of Finnish and Japanese, it remains a question whether it reflects an impairment in domain-general (language non-specific) temporal processing or altered perception that can be shaped by the language-specific phonological system. To address this question, it is necessary to examine languages such as Mandarin Chinese in which vowel duration change does not signify a phonemic change. If the domain-general temporal processing deficit hypothesis is correct, we would expect to see a similar duration perception deficit for speech and nonspeech stimuli in autism across the different languages such as Finnish, Japanese, and Chinese. Another possibility is that Chinese children with autism may exhibit a different pattern of duration perception in the speech context from their Finnish and Japanese peers. There is mounting evidence that children with autism may demonstrate not only a deficit in discriminating across-phonemic-category differences but also enhanced sensitivity to detect acoustic differences within a phonemic category (Wang *et al.*, 2017; You *et al.*, 2017). In contrast, typically developing children show reduced within-category discrimination and enhanced sensitivity to across-category differences as reflected in the categorical perception tests (Wang *et al.*, 2017). Following the "allophonic perception" theory for autism (You *et al.*, 2017), which is consistent with the notion of abnormal NLNC in autism, the Chinese subjects with autism in our study may not necessarily demonstrate a deficit in perceiving vowel duration difference in comparison with the TD controls as the vowel duration contrast in Chinese reflects within-category allophonic variation.

As in the case of pitch perception in relation to language background (Yu et al., 2015), we were particularly motivated in the present study to address whether there were distinct neural sensitivity patterns for duration perception in simple nonspeech stimuli and complex speech stimuli in Chinese-speaking children with autism in comparison with age-matched TD controls. A passive listening oddball paradigm was adopted to study neural sensitivity and orienting responses, which has been widely used in developmental research on auditory and linguistic processing with individuals with autism (Gomot *et al.*, 2002; Ferri *et al.*, 2003; Gage *et al.*, 2003; Jansson-Verkasalo *et al.*, 2003; Lepistö *et al.*, 2005; Kujala *et al.*, 2007; Lepistö *et al.*, 2008; Yu *et al.*, 2015). As the experimental protocol does not require focused attention or any overt responses in testing young children with relatively limited communication ability, this neurophysiological approach can serve as an objective tool to measure auditory discrimination and speech processing (Näätänen *et al.*, 2011; Kujala *et al.*, 2013). Abnormalities in the MMN component may reflect pre-attentive neural sensitivity problems in speech discrimination. Previous research has shown that children with autism had diminished P3a responses to speech stimuli but not to nonspeech signals, indicating deficits in the control of attentional resources in the context of novelty detection for socially significant information (Lepistö *et al.*, 2005; Kujala *et al.*, 2013).

Our experimental design included two stimulus conditions. In the nonspeech condition, the stimuli were a train of pure tones interleaved with deviant longer tones. This condition aimed to replicate previous findings of duration discrimination deficit with acoustically simple sounds. In the vowel condition, nonsense syllables with different vowel durations were used. If duration perception deficit in autism is universal regardless of language background, we would observe diminished neural sensitivity as assessed by the MMN response in both nonspeech and speechconditions in the autism group. Alternatively, as vowel duration difference reflects within-category variation in Mandarin Chinese, the allophonic mode of perception in autism (You et al., 2017) would predict that MMN deficits for duration discrimination might be observed for the tone condition but not for the vowel condition due to the enhanced within-category duration discrimination in the Chinese children with autism in comparison with categorical mode of speech perception with reduced within-category discrimination in TD controls.

Method

Participants

The experiments were conducted with approval from the institutional review board of South China Normal University in compliance with the 1964 Helsinki declaration and its later amendments of ethical standards. All children in the study had normal hearing assessed by standard audiometric screening with pure tones before study admission. All participants were monolingual native speakers of Mandarin Chinese. Informed consent was obtained from each individual's parent or guardian following the approved protocol.

Children in the autism group were recruited from the Guangzhou Cana School (Guangzhou Rehabilitation Research Center for Children with Autism). The diagnosis was established according to the DSM-IV criteria for autistic disorder (American Psychiatric Association, 2000) by certified pediatricians and child psychiatrists. All the children with autism were verbal with limited communication ability, and had delayed onset of speech as measured by their use of two-word utterances. We confirmed the diagnoses using the Chinese version of the Gilliam Autism Rating Scale–Second Edition (GARS-2; Gilliam 2006), as the Autism Diagnostic Observation Schedule (ADOS, Lord et al., 2000) and the Autism Diagnostic Interview –Revised ADI-R (Lord *et al.*, 1994) have not been officially validated and widely adopted in mainland China (Huang *et al.*, 2013; Sun *et al.*, 2013). Detailed usage of GARS-2 and the Autism Index scores are included in the supplementary material. The children in the TD group were recruited from a local public elementary school. These TD children had no known psychiatric or neurological disorders.

Originally, twenty-three children with autism and 20 TD controls participated in the study. After EEG artifact inspection and rejection, only those children whose accepted trials were above 60% of the total presented trials entered final analyses. Specifically, the reported data included 22 children with autism (mean age 9.6 yrs, SD =1.88, 20 boys) and 20 controls (mean age 9.4, SD =1.71, 16 boys) for the pure tone condition, and 18 children with autism (mean age 9.8, SD =2.03, 16 boys) and 17 controls (mean age 9.4, SD=1.85, 15 boys) for the vowel condition. The two groups were closely matched by age (F (1, 40) = 0.12, p = 0.731 for the tone condition; F (1, 33) =0.27, p = 0.607 for the vowel condition) and sex. Non-verbal IQ scores were measured using the Raven's Standard Progressive Matrices Test (Raven & Court, 1998). The autism group scored lower on non-verbal IQ compared to the TD group (F (1, 40) = 21.04, p < .001 for the tone condition; F (1, 33) = 19.40, p < 0.001 for the vowel condition) (Table 1). 77% of the children with autism had comorbid intellectual impairment. The lower non-verbal IQ scores in the autism group were expected and consistent with reported IQ profiles in the literature (Dawson *et al.*, 2007).

[Table 1]

Stimuli and Procedure

This study included a pure tone condition and a vowel condition. The durations of the standard and deviant sounds for the oddball experiments in both conditions were 250 ms and 350 ms, respectively. Sound intensity was normalized to be equal for all stimuli. The nonspeech stimuli used 295 Hz pure tones to match the average fundamental frequency of the speech stimuli. The vowel stimuli used nonsense syllables (/dü1/ as spelled in the Mandarin *pinyin* or /tý/ in international phonetic alphabet spoken with the high flat tone in Chinese). The speech syllable was selected from utterances of a female native Mandarin speaker recorded with the Neundo 4 software (Steinberg Media Technologies, Germany). For the vowel condition, the deviant sound was created by lengthening the vowel portion of the standard sound by 100 ms.

Sound creation and manipulation were completed using Praat (Boersma & Weenink, 2014) and Goldwave (http://www.goldwave.com).

There were two separate presentation blocks, one for the tone condition and the other for the vowel condition. The block presentation order was counterbalanced among the subjects. The standard and the deviant stimuli in the oddball block had a presentation ratio of 21:4 with a total of 650 trials in each block. Each block was preceded with 10 trials of standard stimuli. The following stimuli in the oddball paradigm were presented pseudo-randomly with at least two consecutive standards before each deviant. The inter-stimulus interval (ISI) was 500 ms. The stimuli were presented binaurally via AKG K518 headphones at approximately 60 dB SL (sensation level). The participants were asked to watch a muted self-chosen cartoon movie and ignore the sounds during EEG recording.

EEG Recording and Data Analysis

EEG data were recorded with a 32-channel BrainAmps DC amplifier system at a sampling rate of 500 Hz (Brain Products, Germany). The left mastoid was used as the reference electrode, and the AFz as the ground electrode. Eye blinks and movements were monitored with electrodes placed below the right eye and at the outer corner of the left eye. Electrode impedances were kept below 10 kΩ. The data was offline re-referenced to the average of left and right mastoid recordings. Epochs of 800ms including a 100 ms pre-stimulus time were averaged separately for the standards and deviants. The epochs were digitally filtered with a 1-30 Hz band-pass filter sloping at 12 dB/octave and baseline-corrected. Trials with instantaneous amplitude exceeding ±150 µV were rejected. Only the standard trial immediately preceding each deviant trial was included in averaging for the standard trials to balance the trial numbers between the standard and deviant (Zhang *et al.*, 2005; Zhang *et al.*, 2009). In the pure tone condition, the average numbers of trials for the standard and deviant sounds (with the ranges reported in parenthesis) were 96 (range: 76~144) and 98 (78~144) in the autism group, and 99 (77~124) and 99 (76~121) in the TD group. In the vowel condition, the accepted trial numbers were 89 (42~106) and 90 (43~114) in the autism group, and 94 (77~121) and 94 (74~126) in the TD group. There were no significant differences in trial number comparisons between the subject groups for either stimulus condition.

Difference waveforms were derived by subtracting the averaged standard ERP from the averaged deviant ERP for each subject in each condition. Based on the grand average ERP waveforms, the MMN and P3a were quantified using post-stimulus windows of 250-400 ms and 350-500 ms respectively. Peak detection of MMN and P3a were performedwithin respective search windows for each subject. MMN and P3a amplitudes were obtained by computing the mean voltage of samples with a 60 ms time span around the peak latency. One-way ANOVA tests examining group differences was performed separately for latency and amplitude measures for each stimulus condition. Correction for multiple comparison was applied where appropriate.

Results

MMN data

In the pure tone condition, the ANOVA tests revealed significant subject group effects in MMN amplitude ($F(1, 40) = 4.30$, $p < .05$, $\eta^2 = .10$) and latency ($F(1, 40) = 5.07$, $p < .05$, $\eta^2 = .11$). As expected, the autism group showed diminished and delayed MMN compared to the TD group for detecting the duration differences in the pure tones.

In the vowel condition, there was no significant subject group effect in either MMN amplitude ($F(1, 33) = 0.09$, $p = .761$, $\eta^2 = .00$) or latency ($F(1, 33) = 2.39$, $p = .132$, $\eta^2 = .07$) (Table 2; Fig. 1).

[Table 2]
[Figure 1]

P3a data

In the pure tone condition, the two subject groups showed no significant differences in either P3a amplitude ($F(1, 40) = 0.269$, $p = .607$, $\eta^2 = .01$) or latency ($F(1, 40) = 3.46$, $p = .070$, $\eta^2 = .08$) (Table 3).

In the vowel condition, there was a significant group effect in P3a amplitude ($F(1, 33) = 8.55$, $p < .01$, $\eta^2 = .21$). Specifically, the autism group had diminished P3a compared to the TD group. No significant group effect was found in P3a latency ($F(1, 33) = 1.09$, $p = .304$, $\eta^2 = .03$).

[Table 3]

Discussion

The present study employed a passive listening oddball paradigm to examine the perception of duration differences in pure tones and vowels in school-age Mandarin-speaking children with autism and TD controls. The two stimulus conditions exhibited distinct patterns of MMN and P3a responses. First, the autism group showed diminished MMN responses in the pure tone condition but not in the vowel condition compared with the TD group. Second, the autism group showed diminished P3a responses in the vowel condition but not in the pure tone condition.

The MMN results for the pure tone stimuli are consistent with previous work that reported auditory temporal processing deficits in autism. For example, high-functioning children with autism were less accurate in duration reproduction task across various duration levels compared with TD children (Maister & Plaisted-Grant, 2011). Other behavioral paradigms including direct and indirect measures of temporal binding window (Foss-Feig *et al.*, 2010), and temporal order judgment (Shtyrov *et al.*, 1999; Kwakye *et al.*, 2011) also documented reduced temporal resolution in children with autism. Furthermore, neurophysiological studies revealed that children with autism had diminished MMN responses not only to duration contrasts in tonal stimuli but also to vowel duration changes (Lepistö *et al.*, 2005; Lepistö *et al.*, 2006) as well as subtle temporal cues such as duration of formant transition (Kuhl *et al.*, 2005). A recent fMRI study further showed that the primary auditory cortex of individuals with autism was involved in processing temporally complex sounds whereas non-primary auditory cortex was recruited for the same task in the TD individuals; moreover, such group differences were not present for spectrally complex sounds (Samson *et al.*, 2011). Thus, the current finding of MMN deficit in autism in the acoustically simple pure tone condition may be associated with a basic auditory insensitivity to temporal cues in the auditory association cortex at pre-attentive level.

Unlike the pure tone condition, the MMN data in the vowel condition do not support a simple language-universal account of duration perception deficit in autism. The lack of between-group MMN differences in the vowel condition is in sharp contrast with studies where Finnish-speaking children with autism showed diminished MMN for vowel duration discrimination in comparison with typically developing children (Lepistö *et al.*, 2005; Lepistö *et al.*, 2006). These diverging language-dependent MMN patterns support the "allophonic perception" theory for autism (You et al., 2017). According to this theory, unlike typically developing children who demonstrate categorical perception for speech sounds, children with ASD may exhibit a deficit only when discriminating phonemic contrasts. Moreover, this phonemic discrimination deficit is coupled with enhanced discrimination of acoustic differences within a phonemic category. In this perspective, the categorical perception deficit in autism is a byproduct of the allophonic mode of perception. In the case of Finnish children with autism, they showed diminished MMN for vowel duration differences because the duration contrast in Finnish involves a phonemic difference. In the Chinese children with autism, however, the vowel length difference is within-category allophonic variation for which the autism group may exhibit enhanced discrimination. In the typically developing children who demonstrate categorical perception for speech sounds, the discriminatory sensitivity is reduced for within-category differences and enhanced for across-category differences. As the allophonic mode of perception does not apply to the pure tone condition, we were able to replicate the duration processing deficit in autism for the pure tone condition. In the speech context, however, the enhancement from the allophonic mode of perception in the autism group and the reduction due to categorical mode of perception in the control group effectively pulled even the MMN responses in the two subject groups for detecting the within-category vowel duration differences.

In another study on duration-cued speech contrasts (Kuhl et al., 2005), toddlers with autism were found to show diminished MMN for the /ba-wa/ contrast (i.e., short vs. long formant transition) in comparison with an age-matched TD group. As the /ba-wa/ contrast involves two different consonant phonemes, the results are also consistent with the prediction from the "allophonic perception" theory for children with autism in that the speech perception deficit in autism applies only to the between-category discrimination. Together, the existing data from studies of Finnish, Japanese, English and Chinese subjects consistently showed impaired MMN response in autism for duration discrimination in the non-speech condition. When it comes to the speech stimuli, children with autism consistently showed impaired MMN response for detecting duration differences involving a phonemic contrast but not for detecting a within-category duration variation, which supports the notion of an altered perceptual weighting system for the allophonic mode of speech perception in autism.

The P3a results revealed a different aspect of duration processing at the cortical level. P3a reflects attentional orientation or shift towards stimulus change, which becomes prominent when the acoustic difference between the standard and deviant stimuli is large (Polich, 2007). In comparison with the TD group, the autism group showed diminished P3a responses of attention switch to change detection in the vowel condition, whereas no significant group difference in the pure tone condition. These results are consistent with the previous findings of speech-specific deficit in attentional switch in children with autism (Čeponienė *et al.*, 2003; Lepistö *et al.*, 2005). Another factor that may have affected P3a is the social significance of the speech stimuli. Similar results were observed in our previous study using hummed version of lexical tones (Yu et al., 2015). It is widely documented that individuals with autism demonstrate impaired orienting to social signals and speech sounds but not as severely to nonspeech sounds that do not carry linguistic or social significance (Lepistö *et al.*, 2005).

To our knowledge, this is the first study that has examined duration perception inautism in the context of a language in which duration contrast is not phonemic. Combined with evidence from lexical tone studies (Yu *et al.*, 2015; Wang *et al.*, 2017), our results do not support domain-general advantage or disadvantage accounts based on auditory sensitivity to basic spectral or temporal features alone. Rather, a plausible explanation is that the atypical perceptual weighting systems in children with autism are not only influenced by differential basic auditory sensitivity to spectral and temporal cues but also impairment in speech perception in line with an allophonic mode of perception (You et al., 2017) accompanied with impairment in social orienting (Lepistö *et al.*, 2005). In a typically developing child, language acquisition is mediated by social factors and ambient language, involving a language-general-to-language-specific perceptual reorganization early in life (Kuhl, 2010). Disruptions in perceptual weighting of the basic spectral and temporal cues for speech learning could be detrimental for children with autism, as it will not only affect learning of higher level phonological and linguistic structures, but also may prevent proper processing and learning of socially important information. One important venue for future autism treatment and intervention as well as applied autism research is to design training protocols to inhibit the allophonic mode of speech perception and promote categorical perception of speech sounds in service of the native-language phonological system.

While our data clearly demonstrated distinct patterns of discrimination and orienting for processing duration contrasts in speech and pure tone stimuli in Mandarin-speaking children with autism, our study is not without its limitations. First, our experimental design did not include control for acoustic complexity in the pure tone vs. vowel comparison. For future work, it is desirable to add a nonspeech condition of comparable acoustic complexity in order to account for and tease apart simple vs. complex processing in contrast with speech vs. nonspeech processing. Second, our subject sample is limited to only one age group. It is important to recognize that perceptual weighting strategies for sound discrimination and categorization may change from childhood to adulthood in individuals with autism. For instance, while Finnish children with autism had impaired sound duration discrimination (Lepistö *et al.*, 2005; Lepistö *et al.*, 2006), Finnish and Japanese adults with autism were found to show enhanced or equivalent MMN responses to the duration contrasts compared to TDs (Kujala *et al.*, 2007; Lepistö *et al.*, 2007). While our study highlights the importance of cross-linguistic autism research to understand the sensory and perceptual atypicalities in autism in association with a deviant trajectory of early native language neural commitment, future research is needed to investigate how these basic auditory sensitivity advantages and disadvantages interact with phonological development and social learning across the life span. Third, we need to acknowledge that the current design did not directly assess behavioral performance of duration discrimination for the speech and nonspeech stimuli. As the MMN measure has been highlighted as a potential clinical biomarker (Kasai *et al.*, 2005; Gomot *et al.*, 2011; Roberts *et al.*, 2011; Näätänen *et al.*, 2012), further autism studies in developmental speech perception are needed to investigate the brain-behavioral relationship across individual subjects in order to have a better understanding of individual differences and develop customizable and age-appropriate training protocols for optimizing intervention outcomes.

Acknowledgements
This work was supported by grants from the Natural Science Foundation of China (NSFC 31571136), Key Project of National Social Science Foundation of China (15AZD048), as well as Key Project of National Natural Science Foundation of Guangdong Province, China (2014A030311016) to Suiping Wang. Zhang was additionally supported by Natural Science Foundation of China (NSFC 31728009), the University of Minnesota's Grand Challenges Exploratory Research Grant Award and Brain Imaging Research Award from the College of Liberal Arts. We thank Yang Fan, Guiwen He, Xiaoyun Wu, and Kai Fan for their assistance.

Conflict of Interest Statement
The authors declare no conflict of interest.

Author Contributions
XW and YF collected and analyzed data. XW and LY prepared the figures. DH, LY, SW and YZ completed the manuscript. All authors have reviewed and approved the submission.

Ethical Approval
All procedures performed in studies involving human participants were in accordance with the ethical standards of South China Normal University and the 1964 Helsinki declaration and its later amendments of ethical standards.

Data Accessibility
Anonymised data and supporting materials for the manuscript are available upon request to share with the research community. Please contact the corresponding authors for access.

Figure caption

Figure 1. ERP waveforms and topographical maps in the pure tone condition **(Panel A)** and the vowel condition **(Panel B)** for the autism group and typically developing (TD) group. The left column in each panel shows the ERP waveforms at Cz evoked by the standard and deviant stimuli. The right column presents the deviant-minus-standard difference waveforms at Cz and the scalp topography maps of the MMN and P3a responses.

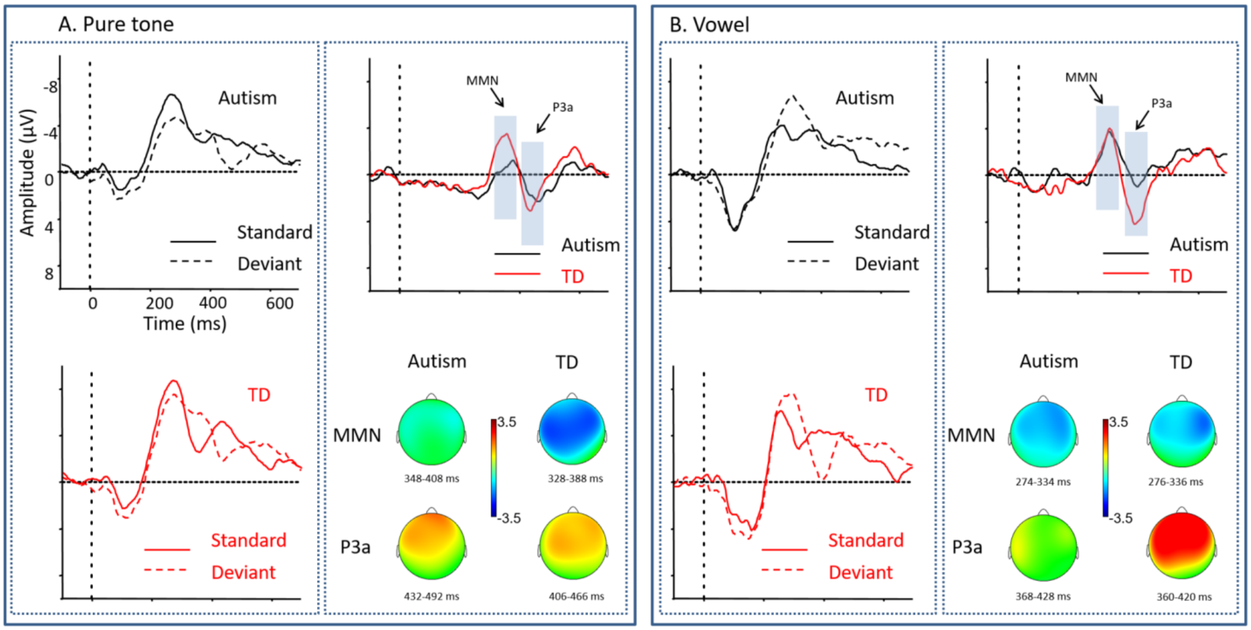

Table 1

Statistics for demographic information of the sample

Condition	*N* (male)		Age (*SD*, range)				Non-verbal IQ (*SD*)			
	Autism	TD	Autism	TD	*p*	η^2	Autism	TD	*p*	η^2
Pure tone	22 (20)	20 (16)	9.6 (1.88, 7-14)	9.4 (1.71, 7.5-12.5)	0.731	0.00	74 (23)	103 (17)	<0.001	0.34
Vowel	18 (16)	17 (15)	9.8 (2.03, 7-14)	9.4 (1.85, 7-12.5)	0.607	0.01	78 (15)	101 (16)	<0.001	0.37

SD: Standard deviation; The *p and* η^2 *represent significance and effect size of F-statistics for group differences.*

Table 2

MMN mean amplitude and latency data at Cz in the autism group and the TD group.

Condition	Amplitude (μV) (*SD*)		Latency (ms) (*SD*)	
	Autism	TD	Autism	TD
Pure tone	-1.51 (2.39)	-3.20 (2.91)	372 (42)	349 (19)
Vowel	-3.27 (3.36)	-2.93 (3.16)	309 (26)	297 (22)

SD: Standard deviation

Table 3

P3a mean amplitude and latency data at Cz in the autism group and the TD group.

Condition	Amplitude (μV) (*SD*)		Latency (ms) (*SD*)	
	Autism	TD	Autism	TD
Pure tone	2.27 (3.25)	2.71 (2.03)	467 (38)	448 (26)
Vowel	1.10 (2.90)	4.21 (3.40)	415 (45)	402 (29)

SD: Standard deviation

Supplementary Material

As Chinese versions of the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000) and the Autism Diagnostic Interview –Revised (ADI-R; Lord et al. 1994) have not been officially validated and widely adopted in China (Huang et al. 2013; Sun et al. 2013), we confirmed diagnosis of the children with autism in our study using the Chinese version of the Gilliam Autism Rating Scale–Second Edition (GARS-2; Gilliam 2006). The GARS-2 is a norm-referenced assessment tool for differentiating individuals with autism from typically developing and those with behavioral disorder. The three subtests of GARS-2—Stereotyped Behaviors, Communication, and Social Interaction, are based on the DSM-IV-TR (American Psychiatric Association 2000) and Autism Society of America (1994) criteria of autism. The normative sample for GARS-2 included 1,107 children and young adults between the ages of 3 and 22 with a diagnosis of autism. The Autism Index (AI) score assessed by GARS-2 can range from 40 to 165. An AI below 70 (69 or less) represent an individual is unlikely to have autism; an AI score between 70 and 84 represents an individual may have autism. Scores above 84 (85 or higher) represent an individual is very likely to have autism. Each child's assessment was obtained from the special education teachers who had daily contact with the child for at least 6 hours. The assessment showed an overall AI of 136 (SD 40, range 57–165) (for individual score, see Fig. S1). There were 3 out of the 23 children who participated in the study with scores (57, 66, 70) below the "very likely to have autism" cut-off (85). For these 3 children, we have obtained independent secondary confirmation of diagnoses of autism with two additional experienced pediatricians who were unrelated to the present study.

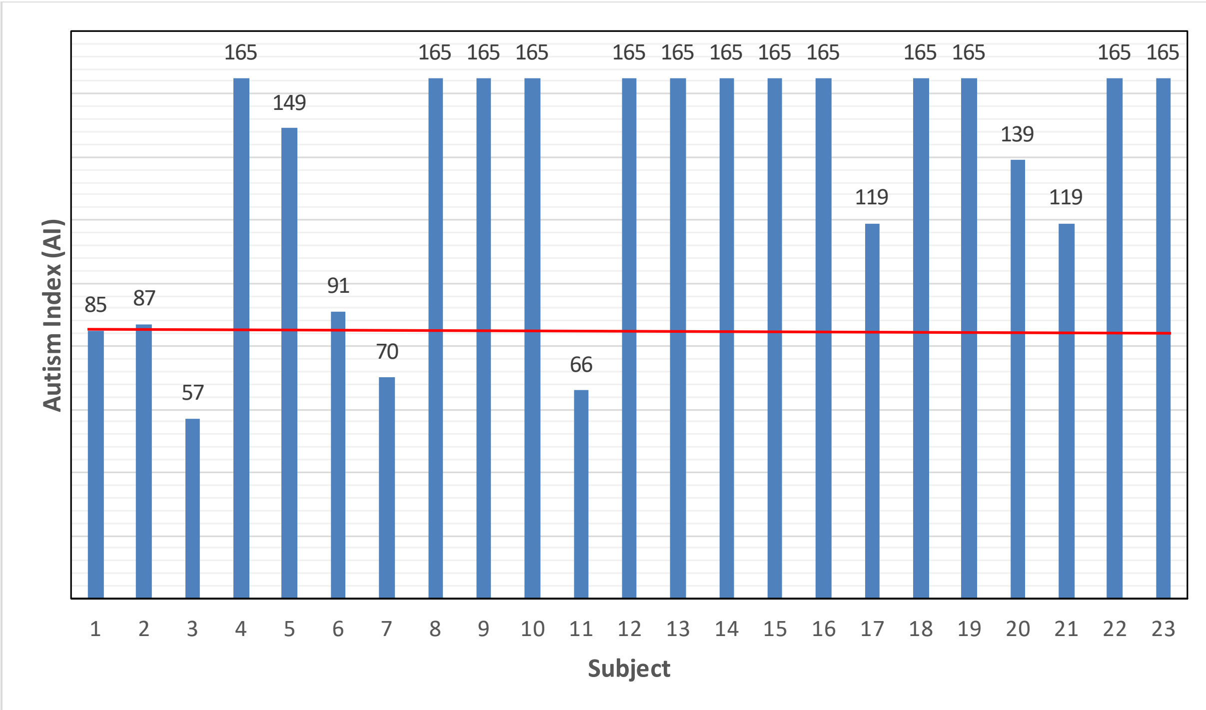

Figure S1. Bar graph showing the Autism Index (AI) of each child in the autism group measured by GARS-2. The red horizontal line indicates the "very likely to have autism" cut-off score (85).

References

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. *Journal of child psychology and psychiatry, and allied disciplines*, **45**, 1107-1114.

American Psychiatric Association (2000) *Diagnostic and statistical manual of mental disorders (4th ed. Text Revision)* American Psychiatric Association, Washington, DC.

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

Boersma, P. & Weenink, D. (2014) Praat: doing phonetics by computer (Version 5. 3. 79)[Computer program]. Retrieved from http://www.praat.org/.

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E. & Bonnel, A.M. (2003) Enhanced pitch sensitivity in individuals with autism: a signal detection analysis. *J Cogn Neurosci*, **15**, 226-235.

Brodeur, D.A., Gordon Green, C., Flores, H. & Burack, J.A. (2014) Time estimation among low-functioning individuals with autism spectrum disorders: evidence of poor sensitivity to variability of short durations. *Autism research : official journal of the International Society for Autism Research*, **7**, 237-244.

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R. & Yaguchi, K. (2003) Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc Natl Acad Sci U S A*, **100**, 5567-5572.

Chen, Y., Norton, D.J., McBain, R., Gold, J., Frazier, J.A. & Coyle, J.T. (2012) Enhanced local processing of dynamic visual information in autism: evidence from speed discrimination. *Neuropsychologia*, **50**, 733-739.

Dawson, M., Soulieres, I., Gernsbacher, M.A. & Mottron, L. (2007) The level and nature of autistic intelligence. *Psychol Sci*, **18**, 657-662.

Falter, C.M., Noreika, V., Wearden, J.H. & Bailey, A.J. (2012) More consistent, yet less sensitive: interval timing in autism spectrum disorders. *Quarterly journal of experimental psychology*, **65**, 2093-2107.

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S.A. & Pennisi, G. (2003) The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, **114**, 1671-1680.

Foss-Feig, J.H., Kwakye, L.D., Cascio, C.J., Burnette, C.P., Kadivar, H., Stone, W.L. & Wallace, M.T. (2010) An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res*, **203**, 381-389.

Foss-Feig, J.H., Stone, W.L. & Wallace, M.T. (2012) Processing of Non-Speech Auditory Stimuli in Individuals with Autism Spectrum Disorders: The Impact of Stimulus Characteristics. *Int Rev Res Dev Disa*, **43**, 87-145.

Gage, N.M., Siegel, B., Callen, M. & Roberts, T.P. (2003) Cortical sound processing in children with autism disorder: an MEG investigation. *Neuroreport*, **14**, 2047-2051.

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

Gomot, M., Blanc, R., Clery, H., Roux, S., Barthelemy, C. & Bruneau, N. (2011) Candidate electrophysiological endophenotypes of hyper-reactivity to change in autism. *J Autism Dev Disord*, **41**, 705-714.

Gomot, M., Giard, M.H., Adrien, J.L., Barthelemy, C. & Bruneau, N. (2002) Hypersensitivity to acoustic change in children with autism: electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, **39**, 577-584.

Groen, W.B., van Orsouw, L., Huurne, N., Swinkels, S., van der Gaag, R.J., Buitelaar, J.K. & Zwiers, M.P. (2009) Intact spectral but abnormal temporal processing of auditory stimuli in autism. *J Autism Dev Disord*, **39**, 742-750.

Haesen, B., Boets, B. & Wagemans, J. (2011) A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Res Autism Spect Dis*, **5**, 701-714.

Heaton, P. (2005) Interval and contour processing in autism. *J Autism Dev Disord*, **35**, 787-793.

Hitoglou, M., Ververi, A., Antoniadis, A. & Zafeiriou, D.I. (2010) Childhood autism and auditory system abnormalities. *Pediatric neurology*, **42**, 309-314.

Huang, A.X., Jia, M. & Wheeler, J.J. (2013) Children with autism in the People's Republic of China: diagnosis, legal issues, and educational services. *J Autism Dev Disord*, **43**, 1991-2001.

Jansson-Verkasalo, E., Čeponienė, R., Kielinen, M., Suominen, K., Jäntti, V., Linna, S.-L., Moilanen, I. & Näätänen, R. (2003) Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neuroscience Letters*, **338**, 197-200.

Jarvinen-Pasley, A. & Heaton, P. (2007) Evidence for reduced domain-specificity in auditory processing in autism. *Dev Sci*, **10**, 786-793.

Järvinen-Pasley, A., Wallace, G.L., Ramus, F., Happé, F. & Heaton, P. (2008) Enhanced perceptual processing of speech in autism. *Dev Sci*, **11**, 109-121.

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**, 2067-2075.

Kasai, K., Hashimoto, O., Kawakubo, Y., Yumoto, M., Kamio, S., Itoh, K., Koshida, I., Iwanami, A., Nakagome, K., Fukuda, M., Yamasue, H., Yamada, H., Abe, O., Aoki, S. & Kato, N. (2005) Delayed automatic detection of change in speech sounds in adults with autism: a magnetoencephalographic study. *Clinical neurophysiology*, **116**, 1655-1664.

Kuhl, P.K. (2004) Early language acquisition: cracking the speech code. *Nat Rev Neurosci*, **5**, 831-843.

Kuhl, P.K. (2010) Brain mechanisms in early language acquisition. *Neuron*, **67**, 713-727.

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**, F1-F12.

Kujala, T., Aho, E., Lepistö, T., Jansson-Verkasalo, E., Nieminen-von Wendt, T., von Wendt, L. & Näätänen, R. (2007) Atypical pattern of discriminating sound features in adults with Asperger syndrome as reflected by the mismatch negativity. *Biol Psychol*, **75**, 109-114.

Kujala, T., Lepistö, T. & Näätänen, R. (2013) The neural basis of aberrant speech and audition in autism spectrum disorders. *Neuroscience and biobehavioral reviews*, **37**, 697-704.

Kwakye, L.D., Foss-Feig, J.H., Cascio, C.J., Stone, W.L. & Wallace, M.T. (2011) Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in integrative neuroscience*, **4**, 129.

Lepistö, T., Kajander, M., Vanhala, R., Alku, P., Huotilainen, M., Näätänen, R. & Kujala, T. (2008) The perception of invariant speech features in children with autism. *Biol Psychol*, **77**, 25-31.

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M. &Näätänen, R. (2005) The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Res*, **1066**, 147-157.

Lepistö, T., Nieminen-von Wendt, T., von Wendt, L., Näätänen, R. & Kujala, T. (2007) Auditory cortical change detection in adults with Asperger syndrome. *Neurosci Lett*, **414**, 136-140.

Lepistö, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Näätänen, R. & Kujala, T. (2006) Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical neurophysiology*, **117**, 2161-2171.

Lord, C., Rutter, M. & Le Couteur, A. (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*, **24**, 659-685.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H. Jr, Leventhal, B.L., DiLavore, P.C., Pickles, A. & Rutter, M. (2000) The Autism Diagnostic Observation Schedule—Generic: a standard measure of social and communication defi- cits associated with the spectrum of autism. *J. Autism Dev. Disord.*, 30, 205–223.

Maister, L. & Plaisted-Grant, K.C. (2011) Time perception and its relationship to memory in Autism Spectrum Conditions. *Dev Sci*, **14**, 1311-1322.

Martin, J.S., Poirier, M. & Bowler, D.M. (2010) Brief report: Impaired temporal reproduction performance in adults with autism spectrum disorder. *J Autism Dev Disord*, **40**, 640-646.

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T.,Kreegipuu, K., Carlson, S. & Ponton, C. (2012) The mismatch negativity (MMN)--a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clinical neurophysiology*, **123**, 424-458.

Näätänen, R., Kujala, T. & Winkler, I. (2011) Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, **48**, 4-22.

Pallett, P.M., Cohen, S.J. & Dobkins, K.R. (2014) Face and object discrimination in autism, and relationship to IQ and age. *J. Autism Dev. Disord.*, **44**, 1039-1054.

Polich, J. (2007) Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, **118**, 2128-2148.

Raven, J.C. & Court, J.H. (1998) *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press Oxford, UK.

Roberts, T.P., Cannon, K.M., Tavabi, K., Blaskey, L., Khan, S.Y., Monroe, J.F., Qasmieh, S., Levy, S.E. & Edgar, J.C. (2011) Auditory magnetic mismatch field latency: a biomarker for language impairment in autism. *Biol Psychiatry*, **70**, 263-269.
Samson, F., Hyde, K.L., Bertone, A., Soulieres, I., Mendrek, A., Ahad, P., Mottron, L. & Zeffiro, T.A. (2011) Atypical processing of auditory temporal complexity in autistics. *Neuropsychologia*, **49**, 546-555.
Samson, F., Mottron, L., Jemel, B., Belin, P. & Ciocca, V. (2006) Can spectro-temporal complexity explain the autistic pattern of performance on auditory tasks? *J Autism Dev Disord*, **36**, 65-76.
Shtyrov, Y., Kujala, T., Ilmoniemi, R.J. & Naatanen, R. (1999) Noise affects speech-signal processing differently in the cerebral hemispheres. *Neuroreport*, **10**, 2189-2192.
Sun, X., Allison, C., Matthews, F.E., Sharp, S.J., Auyeung, B., Baron-Cohen, S. & Brayne, C. (2013) Prevalence of autism in mainland China, Hong Kong and Taiwan: a systematic review and meta-analysis. *Mol Autism*, **4**, 7.
Szelag, E., Kowalska, J., Galkowski, T. & Poppel, E. (2004) Temporal processing deficits in high-functioning children with autism. *Br J Psychol*, **95**, 269-282.
Wang, X., Wang, S., Fan, Y., Huang, D. & Zhang, Y. (2017) Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific reports*, **7**, 43254.
Yip, M. (2002) *Tone*. Cambridge University Press. Cambridge, UK.
You, R.S., Serniclaes, W., Rider, D. & Chabane, N. (2017) On the nature of the speech perception deficits in children with autism spectrum disorders. *Research in developmental disabilities*, **61**, 158-171.
Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S. & Zhang, Y. (2015) Pitch Processing in Tonal-Language-Speaking Children with Autism: An Event-Related Potential Study. *J Autism Dev Disord*, **45**, 3656-3667.
Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, J., Stevens, E.B., Kawakatsu, M., Tohkura, Y. & Nemoto, I. (2009) Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *NeuroImage*, **46**, 226-240.
Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M. & Tohkura, Y. (2005) Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, **26**, 703-720.