



Lifelong Tone Language Experience does not Eliminate Deficits in Neural Encoding of Pitch in Autism Spectrum Disorder

Joseph C. Y. Lau^{1,2,5,6,7} · Carol K. S. To³ · Judy S. K. Kwan^{1,4} · Xin Kang^{1,2} · Molly Losh⁵ · Patrick C. M. Wong^{1,2} 

Accepted: 11 November 2020 / Published online: 20 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Atypical pitch processing is a feature of Autism Spectrum Disorder (ASD), which affects non-tone language speakers' communication. Lifelong auditory experience has been demonstrated to modify genetically-predisposed risks for pitch processing. We examined individuals with ASD to test the hypothesis that lifelong auditory experience in tone language may eliminate impaired pitch processing in ASD. We examined children's and adults' Frequency-following Response (FFR), a neurophysiological component indexing early neural sensory encoding of pitch. Univariate and machine-learning-based analytics suggest less robust pitch encoding and diminished pitch distinctions in the FFR from individuals with ASD. Contrary to our hypothesis, results point to a linguistic pitch encoding impairment associated with ASD that may not be eliminated even by lifelong sensory experience.

Keywords Autism Spectrum Disorder · Neural pitch encoding · Frequency-following responses · Machine-learning · Tone language

Autism Spectrum Disorder (ASD) is a genetically-based neurodevelopmental disorder which affects many individuals around the world. Although a recent explosion of research in ASD has provided many new insights into this disorder, this research has mostly focused on populations of European descent. One reason that this focus may lead to an

incomplete understanding of ASD is that language, a domain that is often impaired in ASD, is fundamentally cultural. One aspect known to be fundamentally different across languages is the use of pitch in conveying meaning. All spoken languages use pitch to convey prosodic meaning such as mood, i.e. affective prosody, which is a landmark deficit domain of ASD (Baltaxe and Simmons 1985; Losh et al. 2012; Patel et al. 2019). However, most languages (European languages like English being exceptions) *also* use pitch to convey word meaning (i.e. *lexical tone*) (Yip 2002). Languages of this type are *tone languages*. For example, in Cantonese, a tone language, the syllable /ji/ means 'to cure' when produced with a high-level pitch pattern but means 'two' when produced with a low-level pitch pattern. The accurate processing of pitch is thus crucial in the identification of lexical meaning, an important component of language processing in general. Previous studies have found that auditory processing, including the processing of pitch, varies as a function of ASD vs. non-ASD diagnoses [see Haesen et al. (2011) and O'Connor (2012) for reviews]. However, the mixed findings from the literature may only generalise to the understanding of lexical tone processing in ASD to a very limited extent, due to cross-language differences. For example, the different pitch processing ability associated with ASD may result in more pervasive downstream consequences in language

✉ Patrick C. M. Wong
p.wong@cuhk.edu.hk

¹ Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

² Brain and Mind Institute, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

³ Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong SAR, China

⁴ Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

⁵ Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

⁶ Department of Psychology, Northwestern University, Evanston, IL, USA

⁷ Institute for Policy Research, Northwestern University, Evanston, IL, USA

communication in speakers of tone languages. In order to obtain a more comprehensive understanding of ASD, particularly its language phenotypes, population- and language-specific studies are needed.

However, most studies to date on pitch processing in ASD have only considered non-linguistic pitch (e.g. pure tones and musical notes) (Bonnell et al. 2003, 2010; Cheng et al. 2017; Heaton et al. 1998, 2008; Heaton 2003, 2005; Jarvinen-Pasley and Heaton 2007; Kargas et al. 2015; Lepisto et al. 2005; Mayer et al. 2016; Mottron et al. 2000). The findings from this literature are mixed at best, as both enhanced and impaired pitch processing associated with ASD have been found. In most studies examining the processing of non-linguistic pitch patterns, especially earlier ones, enhanced pitch processing was found in individuals with ASD. Studies examining the discrimination of pure tones (Bonnell et al. 2003, 2010; Heaton et al. 1998), musical notes (Heaton 2003, 2005), synthesised non-speech sounds (Cheng et al. 2017) and spoken pitch pattern variations (e.g. a spoken syllable in a non-tone language scaled to different pitch levels, which do not carry lexical meaning) (Heaton et al. 2008; Jarvinen-Pasley and Heaton 2007; Lepisto et al. 2005; Mayer et al. 2016) have generally agreed that individuals with ASD discriminated pitch at a higher performance than those without ASD. However, studies examining the processing of *global* pitch patterns such as musical melodies have found that individuals with ASD did not outperform their typically developing peers (Heaton 2005; Mottron et al. 2000). The enhancement of simple pitch patterns is often understood within an Enhanced Perceptual Functioning (EPF) framework (Mottron et al. 2000, 2006). The EPF model postulates that perception in ASD is biased to *local* stimuli, such that individuals with ASD are more sensitive in fundamental single-dimensional processing in both auditory and visual modalities, but less sensitive to global stimulus contexts and multi-dimensional aspects of processing. However, a more recent and comprehensive study using a more stringent research design found that after controlling for factors such as age, IQ, and musical experience, individuals with ASD generally performed *worse* than those without ASD, even in a simple pure tone discrimination task (Kargas et al. 2015).

The results of Kargas et al. (2015) agree with those of most other studies in the linguistic prosody processing literature, which has mostly focused on prosodic properties of European non-tone languages (e.g. lexical stress and sentence level prosody) (Chevallier et al. 2009; Diehl et al. 2008; Grossman et al. 2010; Hesling et al. 2010; McCann et al. 2007; Paul et al. 2005). The impaired ability to process linguistic prosody in language (word- and sentence-level variations in suprasegmental features including pitch) has been more consistently found in individuals with ASD, and is thus considered a phenotype of ASD (Peppe et al.

2006; Troyb et al. 2011). Individuals with ASD performed worse in identifying lexical stress contrasts [e.g. *re.'call* (v.) vs. *'re.call* (n.) in English] (Paul et al. 2005), or did not outperform individuals without ASD in identifying questions vs. statements and emphatic stress (Chevallier et al. 2009). In a standardised test of linguistic prosody processing (e.g. determining chunking, focus, and questions vs. statements forms based on prosody) known as the Profiling Elements of Prosody in Speech-Communication (PEPS-C) test (Peppe and McCann 2003), impaired performance in individuals with ASD was demonstrated (Hesling et al. 2010; McCann et al. 2007). Hesling et al. (2010) further found that prosodic processing in individuals with ASD involved a different cortical network including the left supramarginal gyrus and the default model network (e.g. left precuneus, left mid frontal gyrus, and right anterior cingulate). Although pitch is a major feature of prosody, other acoustic features such as intensity and speech rate also contribute to the perception of stress and sentence prosody (Crystal 1969). Therefore, deficits in processing sentence prosody and stress do not necessarily point to pitch processing impairment. By contrast, lexical tones are cued *primarily* by time-varying pitch patterns: a change in pitch *alone* within a syllable may entail a change in lexical meaning (Yip 2002).

A few Mismatch Negativity (MMN) studies have directly examined the processing of native lexical tones in children with ASD speaking either Mandarin (Wang et al. 2017; Yu et al. 2015) and Cantonese (Zhang et al. 2019). The MMN is an event-related potential (ERP) which is a neural marker of auditory discrimination at the cerebral cortex (Näätänen et al. 1997, 2007). Less robust MMN responses to lexical tone contrasts were found in the ASD group than in those without ASD, and these were interpreted as reflecting impaired lexical tone discrimination associated with ASD (Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019). The MMN results converge with other findings in the non-tone language prosody processing literature, indicating that a pitch processing deficit is associated with ASD.

However, the conclusion that a pitch processing deficit is associated with ASD cannot legitimately be drawn from these MMN studies, because of a potential confounding factor. While MMN is a reliable neural marker for (structural) language impairment independent of ASD (Davids et al. 2011; Roberts et al. 2011; Rinker et al. 2007; Uwer et al. 2002), the participants' LI status was not taken into account in any of these studies. As LI is a comorbid condition in many individuals with ASD, and may even have a shared genetic etiology with ASD (Bishop 2010), the failure to ascertain the LI diagnosis of ASD participants in the prior MMN studies (Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019) could have confounded the results of the later studies.

Meanwhile, all the MMN studies focused on a very specific age population, i.e. children. Pitch processing in the

neural auditory system is robustly modulated by prior auditory experience in life (Chandrasekaran et al. 2014). The extent to which findings in respect of children with ASD can be generalised to the adult ASD population is therefore unclear, especially in the light of a behavioural study which failed to find a group difference in lexical tone processing between adults with and without ASD (Cheng et al. 2017). Also, while MMN only indexes how sounds are discriminated at the auditory cortex (Näätänen et al. 2007), studies have shown that pitch processing is in fact actively shaped by early sensory levels of the neural auditory system even before the pitch signal reaches the auditory cortex (Krishnan and Gandour 2009; Chandrasekaran et al. 2014). A recent study has also shown that early sensory encoding is related to language ability in children with and without ASD (Tecoulesco et al. 2020). Therefore, it is not clear whether pitch processing deficits found to be associated with ASD reflect impairments of long-latency neural process (e.g. as indexed by the MMN) or a cascading effect of impairments contributed by impairments in short-latency early sensory neural encoding processes.

The present study is the first to comprehensively investigate lexical tone processing in both children and adults with and without ASD, while also taking into account the LI factor. It therefore focuses on how pitch patterns in lexical tones are processed at the *early sensory encoding* in the neural auditory system. Specifically, we examined the Frequency-following Response (FFR) to investigate how fine-grained acoustic signals of time-varying pitch information are encoded at early sensory levels of pitch processing. The FFR is a short-latency electrophysiological component which faithfully indexes how frequency modulation between 70 and 2500 Hz is encoded by neurons (i.e. *phase-locking*) across the auditory pathway (Chandrasekaran and Kraus 2010; Kraus et al. 2017). This frequency range falls within the range of the fundamental frequency (f_0) of speech and its harmonics. The FFR not only indexes auditory encoding at subcortical levels of the neural auditory pathway as traditionally conceived but also “higher-level” (in terms of the ascending auditory pathway) auditory processing at the level of the cortex (Bidelman 2018; Chandrasekaran and Kraus 2010; Coffey et al. 2019). The extent to which the frequency modulation of the stimuli is encoded in the FFR reflects the integrity of the auditory pathway in encoding the stimuli (Kraus et al. 2017). Therefore, the FFR enabled us to compare how fine-grained time-varying acoustic features of pitch (i.e. the f_0) were encoded at *early sensory levels* of the neural auditory pathway in individuals with and without ASD.

Prior FFR studies investigating early sensory encoding of linguistic pitch patterns in non-tone language speakers with ASD have found that linguistic pitch patterns were encoded less accurately and robustly in the neural auditory system of

individuals with ASD (Otto-Meyer et al. 2018; Russo et al. 2008). However, an abundance of FFR research has demonstrated experience-dependent neuroplasticity, suggesting that early sensory encoding of pitch is modulated by prior auditory experience in life [see Chandrasekaran et al. (2014) for a review]. In particular, converging evidence suggests that native tone language experience enhances the neural encoding of linguistic pitch patterns (Krishnan et al. 2005, 2010). Therefore, it remains unclear how such experience-dependent neuroplasticity may interact with ASD in shaping pitch processing in individuals with the diagnosis. Previous research looking at neuroplasticity in pitch processing in typically developing individuals suggests that in general, pitch experience in one auditory domain can benefit pitch processing in other domains. For example, native tone language experience facilitated the early sensory encoding of linguistic pitch patterns of a non-native language (Krishnan et al. 2010). Experience in music (in which pitch is a crucial component) also facilitated the early sensory encoding of pitch in non-native lexical tones in non-tone language speakers (Wong et al. 2007). A behavioural study looking at music processing across speakers of multiple languages further suggested that the pitch processing enhancement attributed to native tone language experience was limited not only to linguistic pitch patterns, but also to non-linguistic pitch patterns such as musical tones (Wong et al. 2012). More relevant to ASD as a genetically-based condition, a recent study has found that musical experience may eliminate genetic risks found to be related to lower lexical tone perception performance (Wong et al. 2020). This study found that neurotypical Cantonese-speaking individuals who carry a C allele of the *ASPM* gene performed worse in lexical tone perception, while individuals who are homozygous for the T allele performed better. Intriguingly, carriers of a C allele who had life-long musical training performed as well as those who were homozygous for the T allele, regardless of musical training. This provides strong evidence to suggest that a lifelong auditory experience may interact with genetically-predisposed pitch processing ability. In the neurocognitive disorder literature, one FFR study found that early sensory encoding of pitch was not impaired in speakers of a tone language who have congenital amusia (Liu et al. 2014). Amusia is a genetically-based impairment in pitch processing and music perception (Peretz et al. 2007), and its presence has been found to impair early sensory pitch encoding in non-tone language speakers (Lehmann et al. 2015). Music therapy and musical training have also been found to improve the production of speech prosody in children with ASD (Lim 2010; Lim and Draper 2011). Together, these studies suggest that neuroplasticity induced by lifelong auditory experience enhances pitch processing including at early sensory levels. Crucially, this neuroplasticity may also interact with early sensory pitch encoding deficits caused by genetically-based

impairments to *compensate* for such deficits and to enable the individuals affected to encode pitch in the typical range.

As speaking a tone language is one of most robust types of life-long auditory experience known to induce neuroplasticity in pitch processing, native tone language experience is a prime candidate of auditory experience which may compensate for early sensory pitch encoding deficits caused by genetically-based impairments like ASD. We therefore hypothesise that life-long native tone language experience may induce a compensatory effect for pitch processing deficits in ASD: pitch processing deficits in native tone language speakers with ASD will be protected against, or alleviated, by their native tone language experience. It must be noted that this notion of *compensatory effect* of language in this hypothesis does not refer to the genetically based protective effects commonly referred to in the ASD literature (Gockley et al. 2015), but rather the neuroplasticity induced by experience that compensates for impairments caused by neurodevelopmental disorders (Ingvalson and Wong 2013; Voss et al. 2017). Given the evidence of auditory-experience dependent neuroplasticity in FFR studies, we further hypothesise that this compensatory effect pertains to the early sensory levels of the neural auditory system.

One element that is known to interact with auditory experience in inducing plasticity is time, i.e. the duration of said experience. For example, the robustness of early sensory encoding of linguistic pitch patterns is positively correlated with the duration of musical training (Wong et al. 2007). Likewise, age has been reported as a factor in modulating pitch processing. In a pitch discrimination task, children's performance was not on par with adults before puberty (Antonioni et al. 2015). Therefore, we also hypothesise that the tone language-related compensatory effect may also be contingent upon the duration of tone language experience. Linguistic pitch processing of lexical tones in children with ASD might therefore be impaired [i.e. in Wang et al. (2017); Yu et al. (2015); Zhang et al. (2019)], but *additional* language-auditory experience in adolescence might compensate for this deficit when the individuals reached adulthood. The compensatory effect might have resulted in linguistic pitch processing in adults with ASD being on par with those without ASD [e.g. in Cheng et al. (2017)].

To test this hypothesis, we elicited FFRs from the pitch patterns of three respective Cantonese lexical tones embedded in a single syllable in both children and adults with ASD and those without (non-ASD). We examined how robust pitch information in the stimuli was encoded in the FFR and how robustly the brain responded to the lexical tone stimuli. We predicted that FFRs to the pitch patterns elicited from individuals with ASD would be less robust than those without ASD in general, consistent with previous findings which suggest that *linguistic* pitch processing deficits are associated with ASD. Crucially, while we predicted less robust

FFRs in children with ASD, FFRs of adults with ASD were predicted to be on par with or less different from FFRs from non-ASD adults as compared to the children group, due to the hypothesised compensatory effect.

As an additional analysis to control for language impairment, we also classified children with ASD into those with language impairment (ASD + LI) and those without (ASD -LI). FFRs across the three groups (ASD + LI, ASD -LI, and non-ASD) were compared to delineate the contribution of LI in pitch processing in ASD. By ascertaining the LI status of our participants and testing both children and adult populations, we have controlled for LI as a possible confounding variable, and our study therefore has the potential to provide a more accurate picture of early neural sensory encoding of pitch in ASD.

In a follow-up analysis, a machine-learning- (ML) based approach was used to decode the FFRs elicited from individuals with and without ASD. ML based decoding allowed us to examine the extent to which FFRs evoked by the three lexical tone stimuli contained relevant information for discriminating across the three lexical tone categories (Xie et al. 2018, 2019). Importantly, ML based decoding allowed us to examine whether decodability of lexical tone contrasts represented in FFRs varied as a function of ASD diagnosis. As FFR provides a faithful representation of how speech stimuli are encoded in the brain (Kraus et al. 2017), this ML approach allowed us to further infer the extent to which contrastiveness across lexical tones, which are crucial in the identification of different lexical meaning in tone languages, can be maintained in neural processing across ASD and non-ASD populations at early sensory levels.

Methods

Design

We elicited FFRs from individuals with ASD and control subjects without ASD (non-ASD) to examine whether and how neural encoding of linguistic pitch patterns in lexical tones varied as a function of ASD diagnosis. FFRs from two groups of subjects, namely children and adults, were collected to examine the extent to which additional language experience presumably available in adults would interact with ASD diagnosis in the neural encoding of lexical tones. FFR to three Cantonese lexical tones, all produced with the same syllable, namely Tone 1 (T1, high-level pitch pattern), Tone 4 (T4, low-falling pitch pattern), and Tone 6 (T6, low-level pitch pattern), were elicited. Together, this study employs a 2 (DIAGNOSIS, ASD vs. non-ASD) × 2 (GROUP: children vs. adults) × 3 (TONE: three lexical tones) factorial design.

Participants

All participants were native speakers of Hong Kong Cantonese, with no history of fragile X, tuberous sclerosis, birth complications, CNS injury, hearing impairment, mental disorders, or behavioural disorders (other than ASD), as reported by the parents of the children participants, or as self-reported by adult participants. Informed assent from each child participant's parent or legal guardian or informed consent from each adult participant approved by The Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee was obtained before the study commenced, according to the requirements of the Declaration of Helsinki. All participants passed a pure-tone audiometry test by demonstrating pure-tone air conduction thresholds of 25 dB or better at frequencies of 500, 1,000, 2,000, and 4,000 Hz.

A total of 101 native Cantonese-speaking school-age children aged 8–12 completed the experiment. Of these 101 children, 63 subjects had been diagnosed with ASD (ASD group). The other 38 were control subjects who, as reported by their parents, were typically developing and did not have first- or second-degree relatives with ASD (non-ASD group). These participants were recruited using advertisements either posted on social media platforms (e.g. Facebook) or directly sent to schools and organizations with existing populations of individuals with ASD. The Test of Nonverbal Intelligence, Fourth Edition (TONI-4) was administered to all child participants. The Childhood Autism Spectrum Test (CAST) was also administered to all these participants, primarily to ensure that self-reported non-ASD participants did not meet the ASD screening criteria. ASD participants who demonstrated a non-verbal IQ ≤ 85 [in the Test of Nonverbal Intelligence, Fourth Edition (TONI-4)] ($N=1$), were reported to have attention deficit hyperactivity disorder ($N=12$), or failed the hearing test ($N=1$), were excluded. Seven non-ASD children participants were also excluded because their CAST score was ≥ 15 , a cut-off score which indicates possible ASD or related social-communication

difficulties. Only data from the remaining non-ASD children participants ($N=31$, 18 females) were included.

Child participants' ASD status was confirmed through the administration of the Autism Diagnostic Observation Schedule (ADOS-2) Module 3. ADOS-2 was administered and coded by an experimenter (the third author) who had achieved clinical and research reliability for administration and coding. The ADOS-2 administration was conducted in Cantonese. Those who scored in the range of "non-autism-spectrum" (comparison score < 4) ($N=19$) were excluded. Only data from the remaining children ASD participants ($N=30$, four females) were included.

A total of 17 native Cantonese-speaking adults with ASD aged 22–49 were recruited from employment programs particularly designed for adults previously diagnosed with Asperger syndrome or high-functioning Autism. Their current states of ASD were verified through the administration of the ADOS-2 Module 4, conducted in Cantonese. ADOS-2 on adults was also administered and coded by an experimenter who had achieved clinical and research reliability for administration and coding. Five participants scored in the range of "non-autism-spectrum" (comparison score < 2), and were hence excluded. Data from the remaining 12 ASD participants (1 female) were included. In addition, 16 participants (all male) from our laboratory's existing pool of participants were invited back to participate in this study as a control group. All 16 participants in the control group self-reported that they did not have ASD (i.e. non-ASD). The experimenter who had achieved clinical and research reliability for administration and coding in ADOS-2 did not report any trait of ASD observed in these non-ASD adult participants. TONI-4 was also administered to all adult participants.

Demographic data (chronological age, gender, musical experience, and IQ) of all participants are presented in Table 1. Differences in chronological age, musical experience, and IQ across the ASD and non-ASD groups were not statistically significant across the ASD vs. non-ASD groups (Table 1), although the non-ASD group had

Table 1 Demographic information

	ASD (children) M (SD)	Non-ASD (Children) M (SD)	ASD (Adults) M (SD)	Non-ASD (Adults) M (SD)	ASD (combined) M (SD)	Non-ASD (combined) M (SD)
Males: females	25:5	13:18	11:1	16:0	36:6	29:18
Age	9.50 (1.36)	9.71 (1.24)	27.08 (3.06)	25.25 (3.80)	14.52 (8.27)	15 (7.82)
Music (years)	2.27 (2.35)	3.47 (2.74)	2.08 (5.70)	3.94 (4.82)	2.21 (3.55)	3.63 (3.54)
Education (Years)	6.55 (2.06)	7 (2.08)	15.08 (2.27)	16.63 (2.25)	8.99 (4.48)	10.28 (5.07)
IQ	111.27 (11.89)	112.97 (10.52)	107.50 (12.94)	111.19 (12.23)	110.19 (12.16)	112.36 (11.03)

Marginal differences as per t-tests ($0.05 > p < 0.1$) between the ASD and non-ASD groups within the respective age are indicated with italics and underlined

ASD Autism Spectrum Disorder, M mean, SD standard deviation

marginally more musical experience than the ASD group [$t(87) = -1.8781, p = 0.0637$].

Stimuli

Speech stimuli used for electrophysiological testing consisted of three Hong Kong Cantonese lexical tones, namely Tone 1 (T1, high-level pitch pattern), Tone 4 (T4, low-falling pitch pattern), and Tone 6 (T6, low-level pitch pattern). The three tones had the same syllable /ji/, which in combination with the lexical tones, produced three different Cantonese words: /ji1/ (T1, ‘doctor’), /ji4/ (T4, ‘son’), and /ji6/ (T6, ‘two’). These three lexical tones were chosen because their phonemic distinctions have been reported to be stable in the language, i.e. these distinctions do not collapse under diachronic effects like sound change (Mok et al. 2013), or synchronic processing mechanisms such as talker adaptation (Wong and Diehl 2003) across the population. The stimuli were identical to those used in a series of previous studies (Lau et al. 2017; Liu et al. 2014; Maggu et al. 2016, 2018). The stimuli were produced by a male native speaker of Cantonese, and were normalised for duration (175 ms) and intensity (74 dB SPL). As such, f0 (fundamental frequency) contour is the main acoustic feature that differs across the stimuli: the f0 contours for T1, T4, and T6 range from 141–143 Hz, 87–99 Hz, and 96–106 Hz respectively. The waveforms and spectrograms of the stimuli are shown in Fig. 1. Native speakers of Cantonese (the first author and the corresponding author) confirmed the stimuli to be natural exemplars of their respective lexical tone categories.

EEG Recording and Pre-Processing

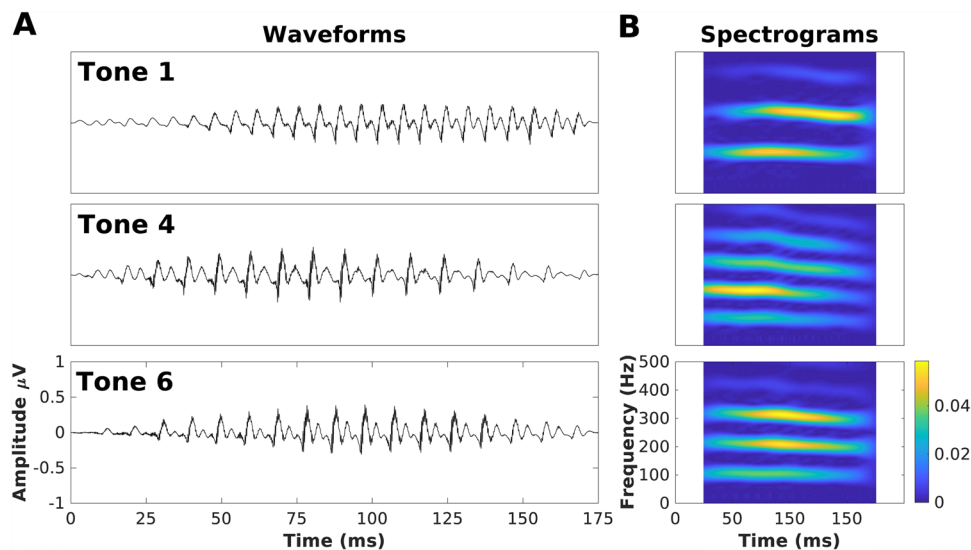
FFRs to the three lexical tone stimuli were elicited in EEG from all participants in three separate blocks. In each block,

2000 sweeps of the stimulus were presented in alternating polarity with an inter-stimulus interval (ISI) that jittered between 74 and 114 ms. The order of presentation of the three blocks was counterbalanced across participants using a Latin square design (with three blocks resulting in three orders), so as to minimise the potential effect of presentation order on the FFRs of the three tones. This highly repetitive presentation context is crucial in our study as no linguistically-relevant contextual information is available for the presentation of each individual tone.

The stimuli were presented to the participant’s right ear through electromagnetically-shielded insert earphones (ER-3A, Etymotic Research, Elk Grove Village, IL, USA) at 80 dB SPL. During the recordings, participants were encouraged to rest or sleep in a reclining chair, consistent with prior FFR recording protocols (Krishnan et al. 2004; Skoe and Kraus 2010). Stimuli were presented via the presentation software Neuroscan Stim2 (Compumedics, El Paso, TX, USA).

Electrophysiological recording took place in an acoustically and electrically shielded booth. Electrophysiological responses were recorded using a SynAmps2 Neuroscan system (Compumedics, El Paso, TX, USA) with Ag–AgCl scalp electrodes, and were digitised at a sampling rate of 20,000 Hz using CURRY Scan 7 Neuroimaging Suite (Compumedics, El Paso, TX, USA). We used a vertical electrode montage (Skoe and Kraus 2010) that differentially recorded electrophysiological responses from the vertex (Cz, active) to bilateral linked mastoids (M1 + M2, references), with a ground electrode placed on the lower forehead. Contact impedance was less than 2 k Ω for all electrodes. The recording of each experimental session lasted around 30 min. Each participant’s electrophysiological data was pre-processed offline using the EEGLAB toolbox (Delorme and Makeig 2004) with the

Fig. 1 Stimulus Characteristics: Waveforms and Spectrograms of the stimuli: syllables /ji/ with a high-level (Tone 1, T1), a low-falling (Tone 4, T4), and a low-level (Tone 6, T6) linguistic pitch patterns in Cantonese



ERPLAB plug-in (Lopez-Calderon and Luck 2014) on MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States). Responses were bandpass filtered from 80 to 2500 Hz (12 dB/octave) to isolate subcortical activity from cortical contamination, and to mimic the phase-locking limit of the subcortical auditory system (Skoe and Kraus 2010). Trials with activities greater than $\pm 35 \mu\text{V}$ were considered artefacts and rejected. Responses to trials of each tone were averaged respectively, all with a 275 ms epoching window encompassing -50 ms before stimulus onset, the 175 ms of the stimulus, and 50 ms after stimulus offset.

FFR Metrics: Peak Autocorrelation and Signal-to-Noise Ratio Measurements

To assess how robust the stimulus was encoded in the FFR, two metrics were derived from each FFR.

To assess the strength of pitch encoded in the FFR, the metric of *peak autocorrelation*, which is a measure of periodicity and phase locking (Wong et al. 2007; Liu et al. 2014), was derived from each FFR from all tones and participants. Using a short-time running autocorrelation technique, the waveform of each FFR (the 175 ms portion excluding the pre-stimulus and post-stimulus portions) was divided into 125 bins, each of 50 ms (49 ms overlap between adjacent time bins). Each of the 125 bins was then time-shifted in 1 ms steps with a delayed version of itself, and a Pearson's r was calculated at each 1 ms interval. For each bin, the maximum autocorrelation value was recorded, with higher values indicating more periodic time frames (Liu et al. 2014). The *peak autocorrelation* of each FFR was computed by taking the average of the autocorrelation peaks (r -values) from the 125 bins.

To assess the overall magnitude of neural activation over the entire FFR period (relative to the pre-stimulus baseline) (Russo et al. 2004), the *signal-to-noise ratio (SNR)* of each FFR was also derived. To do this, the root mean square (RMS) amplitudes of the FFR period (*neural lag* to *neural lag* + 175 ms) and the pre-stimulus baseline period (-50 to *neural lag*) of the waveform were first recorded. The RMS amplitudes were taken as the mean absolute values of all sample points of the waveform within the respective time windows, in μV . The quotient of the FFR RMS amplitude and the pre-stimulus RMS amplitude was taken as the *SNR* value (Russo et al. 2004).

Computation of both *peak autocorrelation* and *SNR* was performed using the Brainstem Toolbox (Skoe and Kraus 2010) on MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States).

Statistical Analyses

To test if lexical tone encoding in FFRs varied as a function of ASD diagnosis across both children and adult groups, as well as tone categories given our 2 (DIAGNOSIS) \times 2 (GROUP) \times 3 (TONE) factorial study design, two linear mixed-effects models were fitted on *peak autocorrelation* and *SNR* metrics respectively. The linear mixed-effects models were fitted using the *lme4* package (Bates et al. 2015) in R (R Core Team 2020).

Given that the linear mixed-effects model is a parametric statistical test, Fisher transformation was first applied to the *peak autocorrelation* values, since Pearson's correlation coefficients do not comprise a normal distribution (Wong et al. 2007).

Each model included fixed effects for DIAGNOSIS, GROUP, and TONE. To take into account the contributions of GENDER, IQ, chronological AGE, years of musical experience (MUSIC), and years of EDUCATION in FFRs, these predictors were also included as fixed effects in each model. To test if the effect of ASD diagnosis on FFRs was modulated by children vs. adults groups as well as tone categories, fixed effects of DIAGNOSIS \times GROUP and DIAGNOSIS \times TONE interactions were included. As random effects, intercepts of SUBJECT and stimulus presentation ORDER in the Latin square counterbalancing design for each subject, as well as by-SUBJECT and by-ORDER random slopes, were included in each model. Minimal intercepts-only random effects structures were used on all models. The p -value of each fixed effect and random effect was computed by a likelihood ratio test comparing a model constructed without the effect in question with the full model. *Post-hoc* pairwise comparisons were conducted using Tukey's honestly significant difference tests (Tukey's HSD) using the *lsmeans* package (Lenth 2016) on R.

Results

FFRs of Cantonese Lexical Tones were Less Robust for the ASD Diagnosis Group

Figure 2 shows the grand averaged FFR waveforms and spectrograms of the ASD and non-ASD diagnosis groups. Mean FFR *peak autocorrelation* and *SNR* of all tones in children and adult groups computed from the FFRs are presented in Fig. 3.

A significant effect for DIAGNOSIS can be found for *peak autocorrelation* [$\chi^2(4) = 14.781, p = 0.0052$], with the Non-ASD group's fisher-transformed *peak autocorrelation* being estimated as 0.134 ± 0.0788 (standard error) higher than the ASD group. In contrast, the effect for DIAGNOSIS for *SNR* was not significant [$\chi^2(4) = 7.0104, p = 0.1353$].

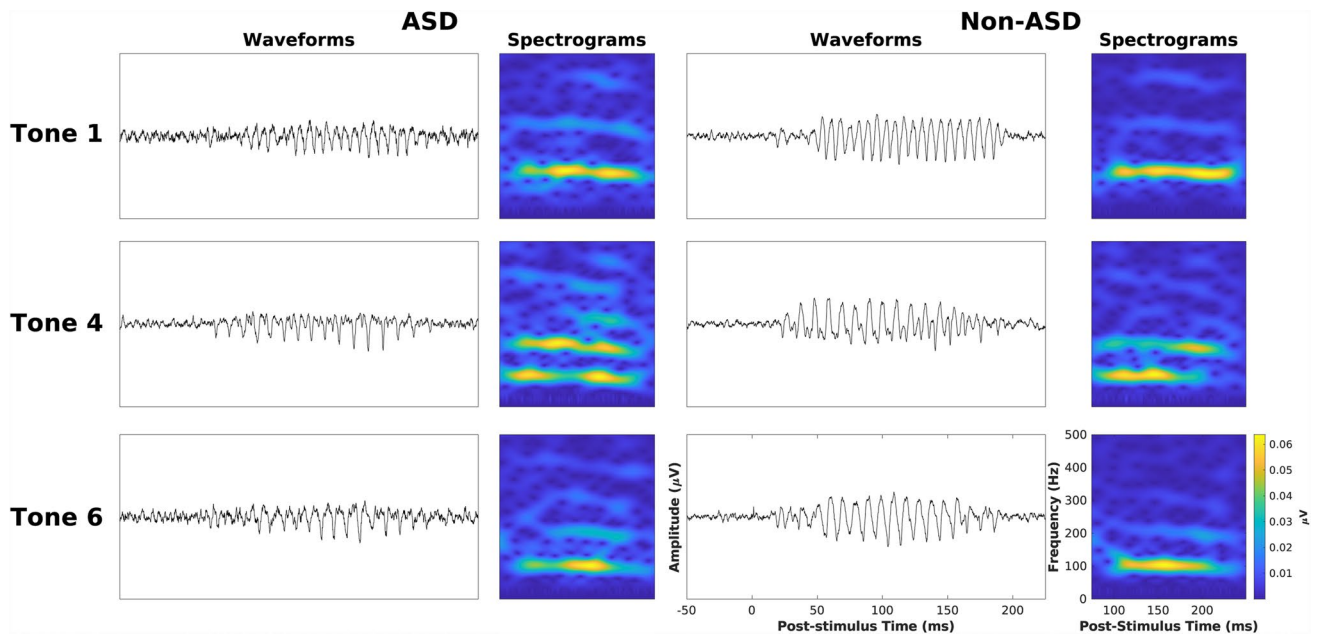
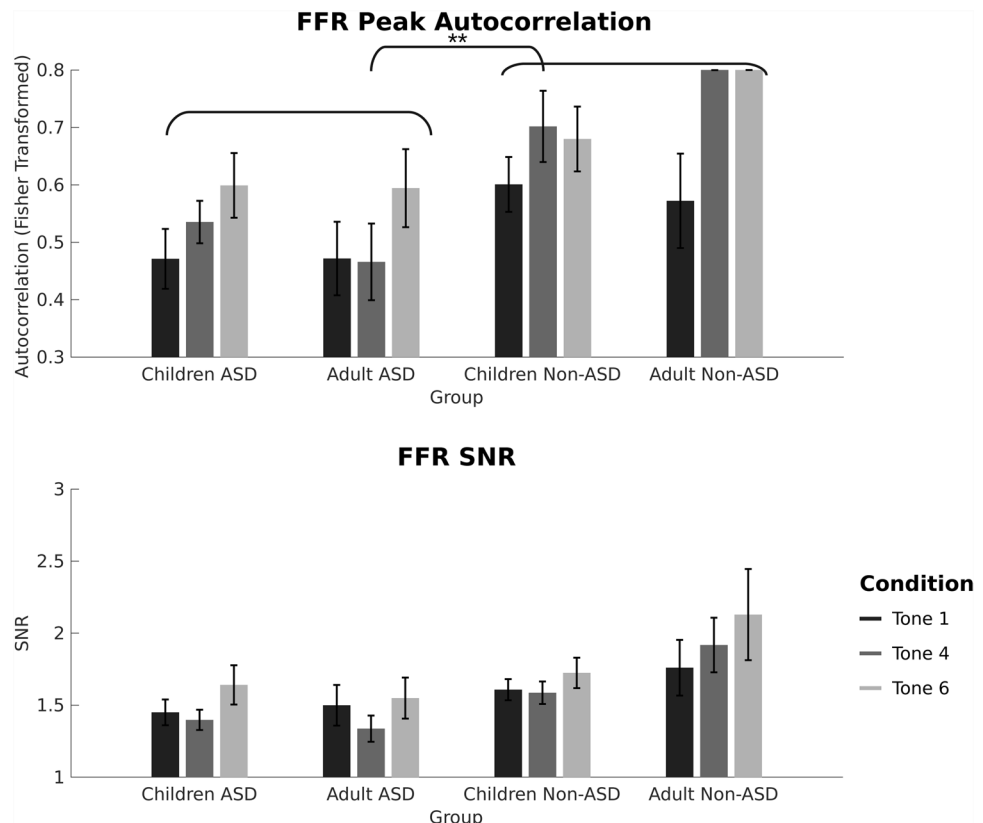


Fig. 2 Frequency-following responses: Waveforms and spectrograms of grand-averaged Tone 1, Tone 4, and Tone 6 frequency-following responses (FFRs) from ASD and non-ASD groups

Fig. 3 Results: Mean *peak autocorrelation* (top) and *signal-to-noise ratio (SNR)* (bottom) of all three tones' FFRs from ASD and non-ASD diagnosis groups, grouped by age groups. Error bars denote \pm one standard error from the mean. $**p < 0.01$ of the effect of DIAGNOSIS in linear mixed-effects model



The effect of GROUP was not significant for either *peak autocorrelation* [$\chi^2(2) = 1.358, p = 0.5071$] or *SNR* [$\chi^2(2) = 2.1697, p = 0.3380$].

Crucially, the DIAGNOSIS \times GROUP interaction was neither significant for *peak autocorrelation* [$\chi^2(1) = 1.1279, p = 0.2882$] nor *SNR* [$\chi^2(1) = 1.5361, p = 0.2152$].

A significant effect of TONE was also observed for both *peak autocorrelation* [$\chi^2(4) = 32.775, p < 0.0001$] and *SNR* [$\chi^2(4) = 13.783, p = 0.0081$]. Post-hoc comparisons showed that *peak autocorrelation* differed between T1 and T4 ($p = 0.004$), T1 and T6 ($p < 0.001$), but not between T4 and T6 ($p = 0.4171$). For *SNR*, T1 and T6 ($p = 0.0109$), T4 and T6 ($p = 0.0034$), but not T1 and T4 ($p = 0.9288$), differed significantly.

The DIAGNOSIS \times TONE interaction was marginally significant for *peak autocorrelation* [$\chi^2(2) = 5.7012, p = 0.0578$]. Post-hoc comparisons revealed that for the non-ASD diagnosis group, *peak autocorrelation* for T1 were 0.1749 lower (± 0.00389 , standard error) when compared to T4 ($p = 0.0002$), and 0.1642 lower (± 0.00389 , standard error) when compared to T6 ($p = 0.0005$). For the ASD diagnosis group, *peak autocorrelation* for T1 were

0.1263 lower (± 0.0411 , standard error) when compared to T6 ($p = 0.0291$).

There was also a marginal effect of EDUCATION [$\chi^2(1) = 3.5276, p = 0.0604$] for *peak autocorrelation* ($\beta = 0.027$).

Detailed results of all effects in the linear mixed-effects models are shown in Table 2.

Discussion

The results suggest that FFRs were less robust for individuals with ASD than for those without ASD. Specifically, the fine grained encoding (indexed by *peak autocorrelation*) of the primary acoustic correlate to Cantonese lexical tones (e.g. the f0), as opposed to the overall neural activation (indexed by *SNR*) per se, was less robust in ASD groups.

The results also suggest that FFRs did not differ across our children and adult groups. Crucially, the lack of a DIAGNOSIS \times GROUP interaction suggests that the effect of ASD diagnosis is not modulated by children vs. adult groups.

As expected, the results show differences across FFRs elicited by different lexical tones, consistent with results found in previous studies suggesting that variability of FFRs

Table 2 Linear-mixed effects models results

Effect	Peak autocorrelation	SNR
Diagnosis	$\chi^2(4) = 14.781, p = 0.0052^{**}$	$\chi^2(4) = 7.0104, p = 0.1353$
Group	$\chi^2(2) = 1.358, p = 0.5071$	$\chi^2(2) = 2.1697, p = 0.3380$
Tone	$\chi^2(4) = 32.775, p < 0.0001^{****}$	$\chi^2(4) = 13.783, p = 0.0081^{**}$
Gender	$\chi^2(1) = 0.0291, p = 0.8645$	$\chi^2(1) = 0.0106, p = 0.9179$
Age	$\chi^2(1) = 0.0475, p = 0.8275$	$\chi^2(1) = 0.1434, p = 0.7050$
IQ	$\chi^2(1) = 0.0210, p = 0.8847$	$\chi^2(1) = 0.0016, p = 0.9676$
Music	$\chi^2(1) = 1.5203, p = 0.2176$	$\chi^2(1) = 0.7886, p = 0.3745$
Education	$\chi^2(1) = 3.5276, p = 0.0604$	$\chi^2(1) = 2.3188, p = 0.1278$
Diagnosis \times group	$\chi^2(1) = 1.1279, p = 0.2882$	$\chi^2(1) = 1.5361, p = 0.2152$
Diagnosis \times tone	$\chi^2(2) = 5.7012, p = 0.0578$	$\chi^2(2) = 1.0726, p = 0.5849$
Random effect: subject	$\chi^2(2) = 87.18, p < 0.0001^{****}$	$\chi^2(2) = 76.20, p < 0.0001^{****}$
Random effect: order	$\chi^2(1) = 0, p = 1$	$\chi^2(1) = 0, p = 1$
Effect: <i>FFR metric</i>	Post-hoc comparisons	
Diagnosis: <i>peak auto correlation</i>	Non-ASD—ASD: LSMEANS: $0.187 \pm 0.065, p = 0.0049^{**}$	
Tone: <i>peak auto correlation</i>	T1—T4: LSMEANS: $0.1095 \pm 0.0283, p = 0.0004^{***}$ T1—T6: LSMEANS: $0.1452 \pm 0.0283, p < 0.0001^{***}$ T4—T6: LSMEANS: $0.0358 \pm 0.0283, p = 0.4171$	
Tone: <i>SNR</i>	T1—T4: LSMEANS: $0.0221 \pm 0.0604, p = 0.9288$ T1—T6: LSMEANS: $0.1765 \pm 0.0604, p = 0.0109^*$ T4—T6: LSMEANS: $0.1986 \pm 0.0604, p = 0.0034^{**}$	
Diagnosis \times tone: <i>peak auto correlation</i>	T1: Non-ASD—ASD: LSMEANS: $0.1310 \pm 0.0727, p = 0.4679$ T4: Non-ASD—ASD: LSMEANS: $0.2618 \pm 0.0727, p = 0.0056^{**}$ T6: Non-ASD—ASD: LSMEANS: $0.1688 \pm 0.0727, p = 0.1914$	

Likelihood-ratio test statistics for the linear-mixed effects models on FFR *peak autocorrelation* and *signal-to-noise ratio (SNR)* (top) and results for *post-hoc* pairwise comparisons conducted using Tukey's HSD, with estimates from the *lsmeans* R package (bottom)

Significant effects are indicated with $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$

exists across different stimulus conditions (Liu et al. 2014; Maggu et al. 2018).

We also found a marginal DIAGNOSIS \times TONE interaction for *peak autocorrelation*, with *post-hoc* tests showing that the difference between ASD and non-ASD were only significant for T4 ($p = 0.0056$), but not for T1 ($p = 0.4979$) and T6 ($p = 0.1914$).

The marginal effect of EDUCATION, with a positive β value (albeit a weak one), suggests that the number of years of formal education may be positively related to the FFR. Education is a correlate to socio-economic status (SES). In the FFR literature, the association between lower SES and less robust neural encoding has already been demonstrated (Skoe et al. 2013). It is also likely that in the current study, SES, represented by years of formal education, indexes genetic and experience-based factors that modulate the individual variability found in FFRs in our participants.

Examining the Role of Language Impairment in FFR

Cantonese-speaking Children with and Without Language impairment

Since LI is a common co-morbid condition of ASD, we also planned for an additional analysis to directly examine the role of LI in modulating early sensory encoding of pitch indexed by the FFR. The analysis was performed on our children group, whose LI diagnoses could be determined through the administration of the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS), available for children aged 5–12 (T'sou et al. 2006). This analysis tests the extent to which FFR *peak autocorrelation* and *SNR* varied as a function of LI diagnosis (Table 3).

Methods

In the children group, ASD participants were divided into two experimental groups based on their language

Table 3 LI analysis: demographic information of participants

	ASD + LI M (SD)	ASD – LI M (SD)	Non-ASD M (SD)
Males: females	12:4	13:1	13:18
Age	9.81 (1.52)	9.14 (1.10)	9.71 (1.24)
Music (years)	1.94 (2.14)	2.64 (2.59)	3.47 (2.74)
Education (years)	5.97 (2.55)	7.21 (1.05)	7 (2.08)
IQ	106.06 (7.50)*	117.21 (13.38)*	112.97 (10.52)*

Significant one-way analysis of variance ($p < 0.05$) across the three groups are indicated with*

ASD Autism Spectrum Disorder, LI language impairment, M mean, SD standard deviation

performance as assessed by HKCOLAS, namely 1) ASD children with language impairment (ASD + LI); and 2) ASD children without language impairment (ASD -LI). Participants in the ASD + LI group (N = 16) all scored below the 10th percentile on any two out of five subtests in the HKCOLAS for their age, the cut-off criteria for LI diagnosis in HKCOLAS (T'sou et al. 2006). Participants in the ASD -LI group (N = 14) scored above the 20th percentile on at least four out of five subtests in the HKCOLAS for their age. All non-ASD participants (N = 31) were not diagnosed to have LI based on the same criteria of HKCOLAS. Demographic information on these three diagnosis groups (ASD + LI, ASD -LI, non-ASD) is presented in Table 4.

To test if FFRs varied as a function of these three diagnosis groups (ASD + LI, ASD -LI, non-ASD), two linear mixed-effects models were fitted on *peak autocorrelation* and *SNR* metrics respectively. Each model included fixed effects for DIAGNOSIS, TONE, and their interaction. Covariates of GENDER, IQ, chronological AGE, years of musical experience (MUSIC), and years of EDUCATION were also included as fixed effects. As random effects, intercepts of SUBJECT and stimulus presentation ORDER in the Latin square counterbalancing design for each subject, as well as by-SUBJECT and by-ORDER random slopes, were included in each model. The p-value of each fixed effect and random effect was computed by a likelihood ratio test comparing a model constructed without the effect in question with the full model.

Planned comparisons were also conducted to compare if *peak autocorrelation* (which showed a significant DIAGNOSIS effect in the main analysis) was different among the ASD + LI, ASD -LI, and non-ASD diagnosis groups. The planned comparisons were conducted using Tukey's honestly significant difference tests (Tukey's HSD) using the *lsmeans* package (Lenth 2016) on R.

Results: FFR did not Vary as a Function of LI Diagnosis

Mean FFR *peak autocorrelation* and *SNR* of all tones in children and adult groups computed from the FFRs are presented in Fig. 4. Detailed results of all effects in the linear mixed-effects models are shown in Table 2.

Consistent with the main analysis, results of the linear mixed-effects models in the current analysis showed a significant effect of TONE in both *peak autocorrelation* [$\chi^2(6) = 14.874$, $p = 0.0213$] and *SNR* [$\chi^2(6) = 13.517$, $p = 0.0355$] models. The effect of EDUCATION, which was marginally significant in the main analysis, was significant in both *peak autocorrelation* [$\chi^2(1) = 5.4962$, $p = 0.0191$] and *SNR* [$\chi^2(1) = 4.7520$, $p = 0.0291$] models.

Crucially, unlike the main analysis examining the effect of ASD vs. non-ASD diagnoses, the effect of DIAGNOSIS was not significant for both *peak autocorrelation* [$\chi^2(6) = 8.1219$,

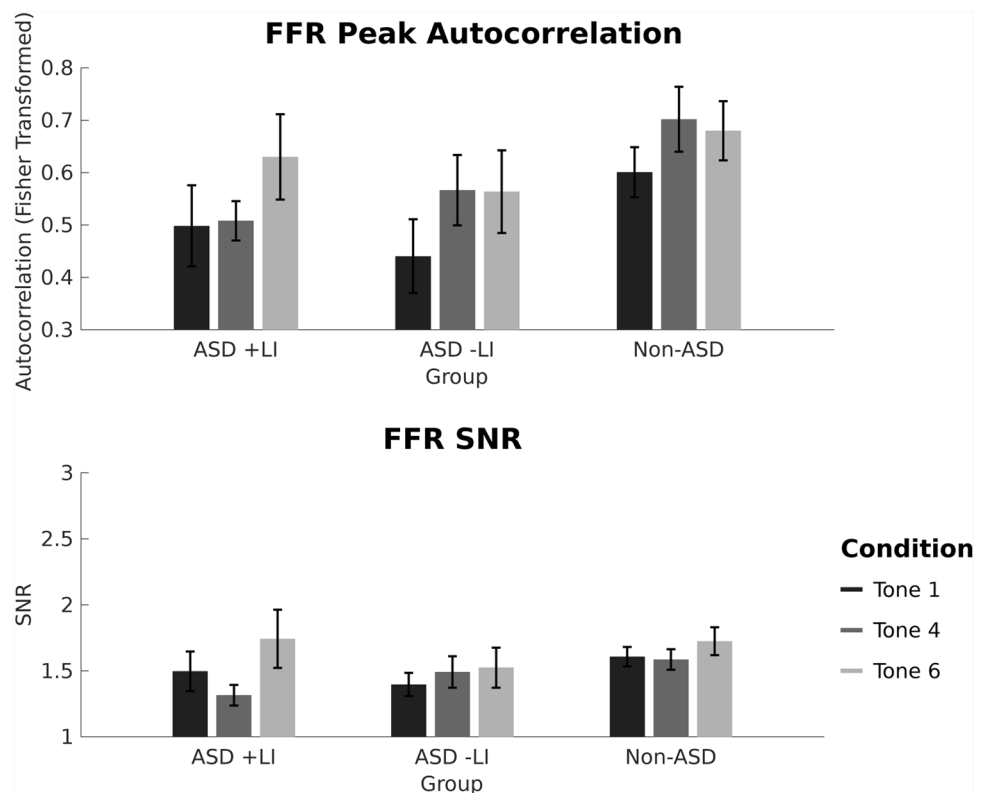
Table 4 Linear-mixed effects models results of LI analysis

Effect	Peak autocorrelation	SNR
Diagnosis	$\chi^2(6) = 8.1219, p = 0.2293$	$\chi^2(6) = 7.4273, p = 0.2831$
Tone	$\chi^2(6) = 14.874, p = 0.0213^*$	$\chi^2(6) = 13.517, p = 0.0355^*$
Gender	$\chi^2(1) = 0.0086, p = 0.9263$	$\chi^2(1) = 0.1277, p = 0.7209$
Age	$\chi^2(1) = 0.5320, p = 0.4658$	$\chi^2(1) = 0.00004, p = 0.9841$
IQ	$\chi^2(1) = 1.0461, p = 0.3064$	$\chi^2(1) = 0.0016, p = 0.9676$
Music	$\chi^2(1) = 0.1743, p = 0.1743$	$\chi^2(1) = 0.7886, p = 0.3745$
Education	$\chi^2(1) = 5.4962, p = 0.0191^*$	$\chi^2(1) = 4.7520, p = 0.0291^*$
Diagnosis \times Tone	$\chi^2(4) = 4.0724, p = 0.3963$	$\chi^2(4) = 5.0611, p = 0.2811$
Random effect: Subject	$\chi^2(2) = 59.551, p < 0.0001^{****}$	$\chi^2(2) = 39.919, p < 0.0001^{****}$
Random effect: Order	$\chi^2(1) = 0, p = 1$	$\chi^2(1) = 0, p = 1$
Effect: <i>FFR metric</i>	Planned comparison	
Diagnosis: <i>peak auto correlation</i>	ASD+LI—Non-ASD: LSMEANS: $0.1109 \pm 0.0868, p = 0.4127$ ASD-LI—Non-ASD: LSMEANS: $0.1579 \pm 0.0930, p = 0.2136$ ASD+LI—ASD-LI: LSMEANS: $0.0469 \pm 0.1077, p = 0.9008$	

Likelihood-ratio test statistics for the linear-mixed effects models on *FFR peak autocorrelation* and *signal-to-noise ratio (SNR)* (top) and results for the planned comparison conducted using Tukey’s HSD, with estimates from the *lsmeans* R package (bottom)

Significant effects are indicated with * $p < 0.05$, **** $p < 0.0001$

Fig. 4 Results: Mean *peak auto-correlation* (top) and *signal-to-noise ratio (SNR)* (bottom) of all three tones’ FFRs from ASD+LI, ASD-LI, and non-ASD diagnosis groups. Error bars denote \pm one standard error from the mean



$p = 0.2293$] and *SNR* [$\chi^2(6) = 7.4273, p = 0.2831$] models in the current analysis.

Planned comparisons showed that *peak autocorrelation* did not differ between the ASD+LI and non-ASD groups

($p = 0.4147$), the ASD-LI and non-ASD groups ($p = 0.2136$), and the ASD+LI and ASD-LI groups ($p = 0.9008$).

The marginally significant *DIAGNOSIS \times TONE* interaction for *peak autocorrelation* in the main analysis was

not significant here in the current analysis [$\chi^2(4) = 4.0724$, $p = 0.3963$].

Discussion

The results of this analysis do not present any evidence to suggest that FFRs vary as a function of LI diagnosis.

In the planned comparisons, while lack of difference for ASD + LI vs. non-ASD (LSMEANS = -0.1109) and ASD-LI vs. non-ASD (LSMEANS = -0.1579) comparisons may be due to the lower sample size after the subgroup division, the direct comparison between ASD + LI vs. ASD—LI (-0.0469 ± 0.1077 , $p = 0.9008$) gives us confidence to infer that FFR *peak autocorrelation* LI does not vary as a function of LI diagnosis *independent of ASD*.

These results, considered alongside those of the main analysis, provide evidence which suggests that less robust pitch encoding (as indexed by FFRs) is associated with ASD *per se*, but not with its common co-morbid LI condition.

Follow-up Analysis: Machine Learning-Based Decoding of Frequency-following Responses (FFR) to Cantonese Lexical Tones

Univariate analyses of FFR metrics suggest less robust pitch encoding of lexical tones in FFRs is associated with ASD. As lexical tones serve to contrast lexical meanings, from a developmental perspective, learning how lexical tones contrast with each other is crucial in the learning of words in a tone language, which could then scaffold into the learning of other structural aspects of language (Singh and Fu 2016). To facilitate the learning of how lexical tones maximally contrast with each other, caregivers implicitly tend to hyperarticulate aspects of pitch patterns that would maximally contrast lexical tones (Tang et al. 2017). As the results of the main analysis show that less robust pitch encoding is consistent across the children and adult groups, it is likely that this pitch encoding deficit may emerge early in life. Impairment in linguistic pitch processing may therefore prevent an individual with ASD from fully acquiring aspects of pitch patterns that could maximally contrast different lexical tones categories. We further hypothesise that the impaired pitch encoding of lexical tones in ASD would lead to the encoding of different lexical tones which were less contrastive from each other. In FFR research, machine-learning-based modelling is a novel approach which allows for the decoding of how the acoustic distinctions of different stimuli are represented in the neural auditory pathway (Xie et al. 2019). Here, we used a machine-learning approach to examine whether lexical tone categories can be decoded from FFRs evoked by the three stimuli, and crucially, the extent to which such decoding is different for FFRs elicited from our ASD and non-ASD groups.

Methods: Support Vector Machine Classification

In this follow-up analysis, we aimed to construct machine-learning-based models to decode the tone category from FFRs elicited from ASD and non-ASD subjects.

Decoding was performed using a supervised machine-learning approach, in which we trained a machine-learning classifier to classify features of each FFR into the three tone categories (T1, T4, T6) provided to the model. Features of each FFR consisted of the time series of the raw waveform. The machine-learning classifier we used was the support vector machine (SVM). SVM performs classification by finding a hyperplane or a set of hyperplanes in a high-dimensional space to separate out the data according to pre-specified labels. Its high-dimensional nature makes the SVM especially powerful in making classifications in neurophysiological responses such as the FFR, in which each data sample may consist of thousands of features (Xie et al. 2019). To ensure the internal validity of the classification, a standard cross-validation procedure was performed. Cross-validation is an iterative process in which the classifier is trained by using only a subset of the data, while the performance of the trained classifier is evaluated by its classification performance on a held-out subset which is not used to train the classifier. In this study, decoding performance is defined as the cross-validation accuracy of the classification of FFRs into the type of stimuli the FFRs were elicited from (i.e. T1, T4, or T6).

Two sub-models were used, namely an *ASD sub-model* and a *non-ASD sub-model* to decode FFRs elicited by ASD and non-ASD subjects respectively. Decoding performance from the *ASD sub-model* and *non-ASD sub-model* was compared.

Dataset

The dataset consisted of FFRs of all three tones (T1, T4, T6) of all participants included in the main univariate analysis.

Support Vector Machine Procedure

The SVM procedures are illustrated in Fig. 5. All SVM procedures were implemented using the LIBSVM library (Chang and Lin 2011) adapted onto MATLAB. The dataset for each sub-model contains the three FFRs elicited by the three tones from each subject in *ASD* and *non-ASD* subjects respectively. FFR features of each sample consist of the 5500 amplitude values from the whole 275 ms of the raw FFR waveform (including 50 ms pre-stimulus and 50 ms of post-stimulus time periods) recorded with a 20,000 Hz sampling rate. In each sub-model, an SVM using a linear kernel constructed three classifiers using a “one-against-one” approach to test the FFR features from all the pairwise combinations

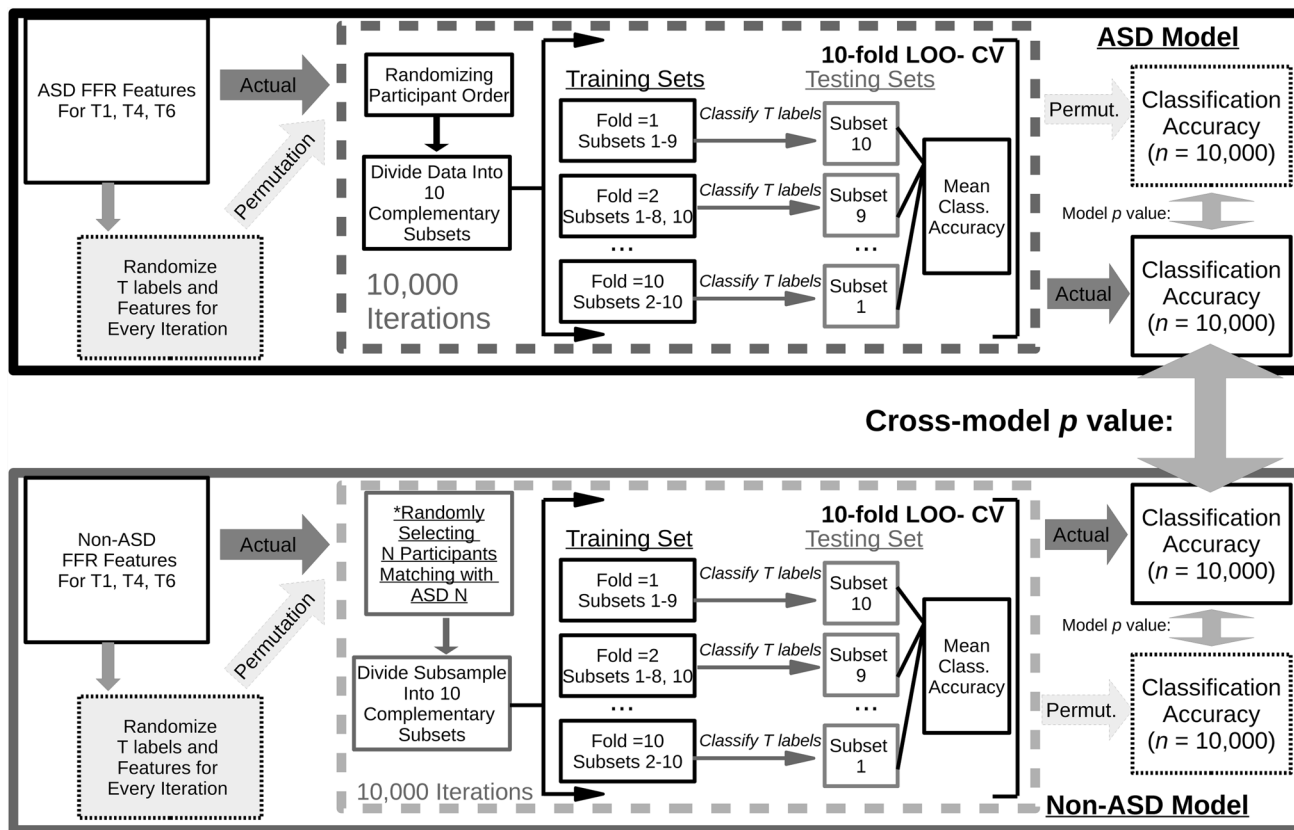


Fig. 5 Machine-learning based decoding models: Procedures to implement linear support vector machines (SVMs) to classify frequency following responses (FFRs) elicited by Cantonese lexical tones [high-level (Tone 1, T1), low-falling (Tone 4, T4), and a low-level (Tone 6, T6) pitch patterns] in ASD subjects and Non-ASD subjects. Leave-one-out cross-validation (LOO-CV): the linear SVM classifier is trained with nine out of ten subsets of the dataset to classify FFRs into one of the three tone categories, while this classi-

fier is then validated to test how well it can generalise to FFR data in the held-out subset. In a total of ten folds, each subset takes turns to be held-out for validation. Permutation: the permutation models to derive a null distribution of classification accuracy are identical with the model on the actual dataset, except that the tone labels and features (time series) are first randomised. Cross-model *p* values are computed by comparing the distributions of accuracies of the actual ASD and non-ASD models

of the three tones. Each classifier assigned one vote for its preferred tone label, and the label with the highest votes across all three classifiers was taken as the classified tone. To objectively evaluate the performance of SVM classification, a cross-validation procedure (ten-fold leave-one-fold-out cross-validation) was performed. The cross-validation procedure started with a randomisation of the order of the list of FFR features (i.e. time series data points). The randomised list of FFR features was then divided into ten consecutive folds. An SVM classifier was trained with the FFRs of nine of the ten folds, and this training was validated by generalizing to the held-out fold (i.e. to label the tone category of the FFRs in the held-out fold). This training-validation process was repeated 10 times until all ten folds had been tested against each other. The accuracy of classification was the percentage of correctly labelled tone category averaged across all ten folds of cross-validation. The cross-validation procedure was repeated for 10,000 iterations, resulting in a distribution containing 10,000 classification accuracy

values, which represents the SVC model’s performance in decoding tone categories from FFRs. The significance of each sub-model was computed using a permutation approach, i.e. by comparing the distribution of the 10,000 accuracy values from the *actual model* (whose data were not permuted) with a null distribution of accuracy values computed by the same cross-validated SVC procedures with labels and features permuted 10,000 times. The percentage of accuracy values from the permuted model that were equal to or higher than the median of the distribution of accuracy values from the *actual model* was taken as the *p*-value of the sub-model (Xie et al. 2018) (Fig. 6).

The procedures of the *ASD sub-model* were identical to the *non-ASD sub-model* except for an additional bootstrapping procedure in the *non-ASD sub-model*. Since there were fewer ASD subjects than non-ASD subjects (ASD $N = 42$; Non-ASD $N = 47$), a balanced bootstrapping technique was adopted to avoid any model performance differences possibly attributable to the larger sample size of the Non-ASD

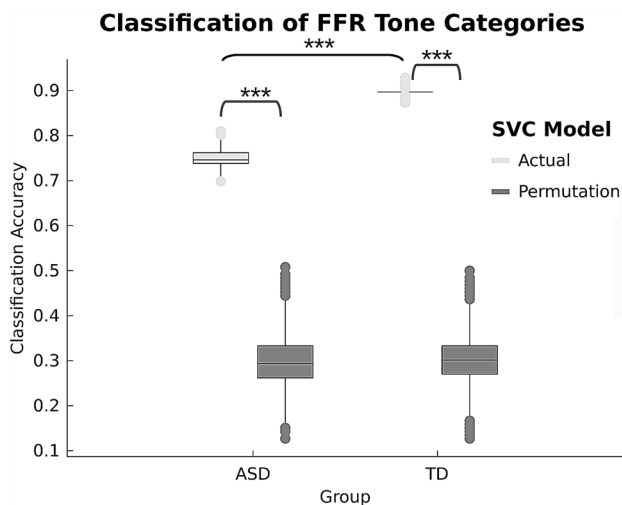


Fig. 6 Machine-learning based decoding results: Boxplot of support vector classification accuracy of actual and permuted models of FFRs from ASD and non-ASD groups. Note that accuracies in both actual models are significantly higher than the null distribution from the respective permuted models. Crucially, accuracies for both TD models are significantly higher than the ASD models. *** $p < 0.001$

group. In each iteration of classification in the *non-ASD sub-model*, all FFR data samples from five subjects (47–42: the number of Non-ASD subjects which exceeded the total number of ASD subjects) were randomly discarded to achieve undersampling with replacement of the dataset (i.e. bootstrapping).

Classification performance difference between the *ASD* and *Non-ASD* sub-models was evaluated by comparing the distribution of 10,000 accuracy values from the *actual model* in the *ASD* sub-model to that in the *Non-ASD* sub-model. The percentage of accuracy values from the sub-model (which has a lower mean accuracy) that was equal to or higher than the median of the distribution of accuracy values from the other sub-model (which has a higher mean accuracy) was taken as the p -value of the model. This p -value estimates the extent to which FFRs of the three tones were decoded more successfully in one group than in the other group.

Results and Discussion: Decoding Performance of Cantonese Lexical Tones were Worse for FFRs of Individuals with ASD

The distributions of classification accuracy of FFR tone categories from SVC models are presented in the box plots in Figs. 3. Overall, good decoding performance was obtained, with the median accuracy of the *ASD* submodel at 0.746 and that of the *non-ASD* sub-model at 0.8968. Permutation tests showed that classification accuracies for both ASD and non-ASD sub-models were both significantly higher than the

null (permutation) distribution (both $ps < 0.001$). Crucially, the SVC model shows that the classification of FFRs from the TD group was significantly more accurate than the ASD group ($p < 0.001$). Results from this ML-based analytics using SVM models suggest that the decodability of FFRs elicited by different Cantonese lexical tones was worse for native-speaking individuals with ASD than for those without ASD.

General Discussion

The present study examined the FFRs elicited by three Cantonese lexical tones from native-speaking children and adults either with or without ASD. Using a well-established measurement of pitch strength on FFRs (i.e. *peak autocorrelation*), results revealed that the f_0 contours encoded in the FFRs of individuals with ASD were less robust than those without ASD.

To the best of our knowledge, this is the first study to examine the early sensory encoding of linguistically-relevant pitch patterns of lexical tones in FFR from individuals with ASD. As far as we are aware, only two other studies have investigated neural pitch encoding in ASD with FFR (Otto-Meyer et al. 2018; Russo et al. 2008). The coverage of these studies was only confined to English speakers, to which the pitch pattern associated with the syllable was not linguistically relevant. These studies also documented poorer neural pitch encoding in speech stimuli in individuals with ASD. Unlike the English speakers from the two studies, all ASD participants in the present study have lifelong tone-language experience. As our results are consistent with the two prior studies of English speakers, our tone language compensatory effect hypothesis is therefore challenged.

Our tone language compensatory effect hypothesis postulated that neuroplasticity in early sensory encoding of pitch induced by lifelong tone language experience may compensate for pitch processing deficits associated with ASD. Prior research has demonstrated the fluidity of experience-dependent neuroplasticity in pitch processing. Neuroplasticity induced by lifelong auditory experience can modulate other domains of pitch processing. For example, lifelong experience in tone language not only may enhance the processing of non-native lexical tones (Krishnan et al. 2010 and musical tones (Wong et al. 2012), but also compensate for some aspects of pitch processing deficits caused by a genetically-based condition (Peretz et al. 2007), namely amusia (Liu et al. 2014). Lifelong experience in music enhances the neural encoding of linguistic pitch patterns (Wong et al. 2007) and may even eliminate genetic risks found to lower lexical tone processing performance (Wong et al. 2020). This evidence strongly suggests that life-long pitch experience may not only modulate other domains of pitch

processing, but also compensate for poorer pitch processing ability (which is genetically-predisposed). However, the results of the current study provide no evidence to suggest that life-long tone language experience could compensate for the impaired linguistic pitch processing associated with ASD, a genetically-based disorder. We posit that this result speaks to the biological depth of linguistic pitch processing impairment in ASD, which affects linguistic pitch processing as fundamental as at the early sensory encoding level. Despite the fluidity of experience-dependent neuroplasticity in pitch processing, it may not be possible to compensate for impairment of such biological depth.

The biological depth of linguistic pitch processing deficits in ASD is also supported by the convergence of results in our children and adult subjects. The duration of auditory experience, not just its presence, is one element that has been shown to modulate pitch processing (Antoniou et al. 2015; Wong et al. 2007). Hence, we hypothesised that age is another factor that would interact with ASD diagnosis in modulating early sensory pitch encoding. Because of children's younger age, they would have experienced tone language for a shorter period of time compared to adults. *Additional* language experience only available in adults may therefore be needed to compensate for the linguistic pitch processing deficit in ASD. Thus, lexical tone processing may be impaired in children with ASD (Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019) but would be on par with people without ASD when the individuals with ASD reach adulthood (Cheng et al. 2017). In contrast, our results showed that the factor of children vs. adults groups did not interact with ASD diagnosis in modulating FFRs, and that pitch encoding in FFRs was less robust in both for ASD, regardless of age. While the results of the present study generally agree with those of the MMN literature, which found impaired lexical tone processing in children with ASD (Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019), they are at odds with the finding of Cheng et al. (2017) that lexical tone processing did not differ across adults with and without ASD on a group level. The lack of an interactive effect between age group and ASD diagnosis in the present study does not lend support to our hypothesis. Instead, it implies that, due to the biological depth of pitch processing deficits in ASD, even life-long exposure to pitch patterns would struggle to compensate for impairments so severe. However, another factor that may contribute to the lack of language experience-related compensatory effect is that linguistic pitch encoding ability may have already levelled-off before children reach teenage. Therefore, additional language experience did not further improve ASD individuals' linguistic pitch encoding. Indeed, a previous study found that ten year-old children already discriminated lexical tones as well as adults (Ciocca and Lui 2003). Our present study did not find an effect of age GROUP. Whether intensive training

focusing on pitch and lexical tone (e.g., Song et al. 2008) would provide a sufficient dose of experience to result in an improvement in pitch encoding for individuals with ASD would require further research.

Interestingly, the results of the present study stand in contrast to the abundance of non-linguistic literature showing enhanced processing of *local* pitch patterns in individuals with ASD (Bonnell et al. 2003, 2010; Cheng et al. 2017; Heaton et al. 1998, 2008; Heaton 2003; Jarvinen-Pasley and Heaton 2007; Lepisto et al. 2005; Mayer et al. 2016). The enhancement of the processing of *local* pitch patterns, e.g. isolated pure-tones and individual musical notes, is often understood within an Enhanced Perceptual Functioning (EPF) framework. The EPF model postulates that perception in ASD is biased to fundamental and single-dimensional local (Mottron et al. 2006). Contrary to more robust neural pitch encoding as predicted by the EPF, we instead found *less* robust encoding in individuals with ASD. One potential explanation for these surprising results is that the current study examining FFRs tapped into different aspects of pitch processing than those considered in prior studies indexed by the behavioural discrimination paradigm. However, studies investigating the relationship between behavioural pitch discrimination and pitch encoding in FFR using pure tones (Marmel et al. 2013) and musical notes (Bidelman et al. 2011) converged to find that the FFR was a reliable neural precursor to behavioural pitch discrimination. Therefore, methodological differences are not likely the sole reason behind the contrastive results between the present study and the non-linguistic pitch discrimination literature. Another major difference between the present study and the non-linguistic pitch discrimination literature is that the stimuli used in the present study were linguistic in nature.

Previous findings from the linguistic prosody literature suggest that the processing of sentence prosody and lexical stress was impaired for individuals with ASD (Chevallier et al. 2009; Diehl et al. 2008; Grossman et al. 2010; Hesling et al. 2010; McCann et al. 2007; Paul et al. 2005). Indeed, these previous findings could not rule out the EPF because deficits in sentence prosody and lexical stress processing could potentially be due to the impaired ability to attend to *global* contextual cues, as opposed to a fundamental *local* pitch processing deficit per se. In contrast, pitch patterns in lexical tones in the present study can be defined as *local* in a sense that they can be linguistically-relevant (i.e. contrasting lexical meaning) at a single-syllable level without other linguistic contexts (Cheng et al. 2017). Therefore, the results of the current study suggest that even for linguistic pitch patterns as *local* as those in lexical tones, pitch processing is impaired as fundamentally as at the early sensory encoding level. Specifically, our results showed less robust FFRs in our ASD group only in terms of *peak autocorrelation* but not *SNR*. This suggests that the deficit in neural encoding

associated with ASD is specific to the encoding of linguistic-relevant pitch patterns (as *peak autocorrelation* indexes) but not overall neural activation to the auditory signal in general (as indexed by *SNR*). Such impairment in the processing of pitch patterns that are *local* and linguistically relevant speaks to the domain-specific nature of pitch processing impairment associated with ASD. Although the results of the current study have contributed partly to the understanding of EPF (specifically its lack thereof in linguistic pitch processing), future studies which test the processing of both global and local linguistic and non-linguistic pitch patterns in a hierarchy of neural processing levels are needed to lend further support to our interpretation.

Intriguingly, we found that FFR was also modulated by the interaction between ASD diagnosis and lexical tone category of the stimuli. Less robust pitch encoding associated with ASD was more evident in FFRs of T4 than in T1 and T6. While FFR indexes phase-locking of both subcortical and cortical neuronal ensembles (Coffey et al. 2019), phase-locking in lower frequency signals has more cortical contributions than that in higher-frequency signals (Bidelman 2018). Our T4 stimulus has the lowest f_0 out of the three stimuli. One explanation of the larger ASD vs. non-ASD group difference found in T4 is that pitch encoding deficits associated with ASD at the cortical level had more contributions to the less robust FFRs in T4 than in T1 and T6.

One important contribution of the present study to the literature is that our analysis revealed that FFR did not vary as a function of structural language impairment. Prior MMN studies on lexical tone processing which found less robust MMN in ASD subjects (Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019) may have been confounded by their unascertained LI. It could be that, although ASD elevated the risk factors of LI, it was LI rather than ASD that directly leads to the less robust linguistic pitch processing found in individuals with ASD. Indeed, a previous study on a non-tone language population found that when children with ASD were classified into language impaired and non-language impaired groups based on linguistic structural deficits, neural responses to pure tone distinctions were most impaired in those with language impairment (Roberts et al. 2011). Our results suggest a dissociation between LI status and FFR responses, and challenge this interpretation. Instead, they suggest that impaired early sensory encoding of pitch can indeed be attributed to ASD rather than LI in a tone language-speaking population.

The results of the ML-based analyses provide further insights into the nature of this impaired neural pitch processing associated with ASD, as well as its implications as a precursor to more general language deficits. Previous studies of the neural processing of lexical tones have found poorer discrimination among lexical tones categories (Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019), but the

neurophysiological bases and behavioural consequences of such poorer discrimination remain elusive. The ML model on FFRs revealed that the decodability of FFRs elicited by different Cantonese lexical tones was worse for individuals with ASD. One interpretation of this result is that it was more difficult for the model to classify the encoded tone category from FFRs from individuals with ASD, as the relevant acoustic features of the stimuli were encoded in their FFRs less congruently and robustly. The FFR indexes how the congruent and robust auditory stimuli are encoded at early sensory levels (subcortical and early cortical levels) of the neural auditory system (Kraus et al. 2017). One intriguing implication is that the less classifiable (i.e. less distinguishable) FFRs to the model might indicate that the auditory stimuli of different lexical tone categories were also less distinguishable when encoded in the brain by the individuals with ASD. This suggests that in general, poorer pitch encoding may result in an elision in the acoustic distinctions across meaning-contrastive lexical tone categories when encoded by individuals with ASD. This elision of acoustic distinctions in early sensory encoding (including at subcortical levels) may then be the precursor to the poorer linguistic pitch pattern discrimination among lexical tones categories found in previous studies (Roberts et al. 2011; Wang et al. 2017; Yu et al. 2015; Zhang et al. 2019) which examined the MMN, a neurophysiological component known to be of cortical origin (Näätänen et al. 2007). Together, this pitch processing deficit, which leads to more collapsed lexical tone distinctions, may even further contribute to general language processing deficits associated with ASD. Impaired neural pitch encoding at the syllable level may lead to a negative cascade effect that affects syntactic processing: impaired pitch encoding at the syllable level may affect lexical processing, which may then affect processing at the sentence level. Syntactic processing may also be affected by pitch processing at the syntax-prosody interface (e.g. focus and prosodic cues that resolve structural ambiguity). Future work is required to elucidate how pitch processing deficits contribute to language processing deficits that underlie social communication deficits associated with ASD. A promising area of enquiry would be to examine how pitch at the syntax-prosody interface is processed by individuals with ASD, along with the relationship between the neural processing of pitch and pragmatics in ASD.

Our study has several limitations that we must note. First, it represents only an initial attempt to test a preliminary hypothesis on tone language compensatory effect. While the tone language compensatory effect hypothesis is not supported here, our results showing impaired early sensory pitch encoding in individuals with ASD were only relative to those without ASD speaking the same language. Future studies examining both tone language and non-tone language speakers are warranted to address the question of whether

tone language experience provides at least some level of compensatory effect relatively to non-tone language experience. Second, since there is no known standardised test for the measurement of structural language ability and diagnosis of language impairment of Cantonese for adults (the HKCOLAS was only designed for children of five to 12 years of age), the present study was only able to perform an analysis testing for the contribution of LI in FFRs on our children group. One factor that may have contributed to the less robust FFRs was that individuals with ASD are transactionally exposed to lower frequency and lower quality language experiences across their lifespan (Naigles 2013). One factor that may have contributed to the less robust pitch encoding in individuals with ASD, especially adults, is that they have had less language exposure than those who do not have ASD. However, our subgroup analysis on children found that FFR did not vary as a function of language ability (as measured by HKCOLAS's criteria of language impairment). To distinguish the contributions of lower frequency and lower quality language experience from the direct effects of pitch processing deficits associated with ASD, future studies are needed on tone languages in which standardised tests are available for *both* children and adults. Also, gender differences in our ASD and non-ASD groups may have partly contributed to our FFR results, as a recent large-scale study found gender differences in auditory processing (Krizman et al. 2019). Since we recruited participants randomly from the general population instead of actively matching demographics across the ASD and non-ASD groups, there were more non-ASD female children ($N=18$) than ASD female children ($N=4$). ASD is a condition that is disproportionately more common in males more than in females (Loomes et al. 2017). However, gender was not found to be a statistically significant factor in modulating FFRs in the present study. Gender differences in auditory processing were found only in individuals older than 14–15 years of age, which is above our child participants' age range. Nevertheless, future studies that compare FFRs of females and males with ASD would be needed to examine the role of gender differences in ASD in modulating linguistic pitch processing. Another limitation of the present study is the wide age range represented in both our children (aged 8–12) and adult (aged 22–49) groups. In particular, the age range for our children group includes children who either have or have not reached puberty. Although age was not found to modulate our FFR results statistically, larger scale studies dividing participants into more precise age groups could better delineate the effect of age and its potential interaction with ASD in linguistic pitch processing, especially considering the relatively small sample size of our adult group. Lastly, we also acknowledge that no standardised tests were administered to non-ASD adults in our study to confirm their self-reported status, and we also acknowledge that age was used merely

as a proxy for language input. Future research should more systematically measure self-reported (non)-ASD status as well as carefully quantify language input, potentially with a longitudinal design.

The current study provides the first evidence that tone language experience is not sufficient to enable early sensory encoding of pitch to be in the typical range for individuals with ASD, as far as we are aware. From a clinical perspective, our results may improve our understanding of the malleable and less malleable factors that contribute directly to language problems associated with ASD, thereby facilitating the fine-tuning of treatment regimens to adapt to different linguistic environments. From a theoretical perspective, we further posit that this pitch processing deficit is subserved by general and fundamental deficits in the auditory system, as this fundamental deficit seems to be independent of environmental exposure such as language experience and its duration, or structural language problems. We conclude that this independence speaks to the biological depth of this pitch encoding deficit associated with ASD. Because of its biological depth, we further propose the possibility that this pitch encoding deficit may be an endophenotype candidate of ASD. The identification of such a potential endophenotype candidate may provide a narrowed scope for the exploration of the genetic underpinnings of ASD (Losh et al. 2008). In particular, this hypothesised endophenotype cuts across cultures and affects different linguistic domains for speakers of different languages. While in English speakers it affects domains such as sentence prosody, in tone language speakers it affects lexical meaning as well. To test this hypothesis and establish the endophenotype status of this pitch processing deficit, future studies could usefully study whether a pitch processing deficit is a subclinical marker that is present among individuals with ASD and their family members who do not have ASD, and seek to identify the genetic markers shared among the family that contribute to this pitch processing deficit (Losh et al. 2017).

Acknowledgements We thank CUHK-NTU-WSU Joint Laboratory for Infant Research and CUHK-UU Joint Centre for Language, Mind and Brain for supporting our research. We also thank Gangyi Feng for his advice on Support Vector Machine. We would also like to thank families who participated in this research study.

Author Contributions PW designed the study; JK, JL, CT, and XK performed research; JL and XK analysed data; JL and PW wrote the paper; CT, JK, XK, and ML edited the paper and provided general comments on the study.

Funding This work is supported by funding from the Health and Medical Research Fund (HKSAR) [02130846], the University Grants Committee (HKSAR) (RGC/HSSPFS) [34000118], the Global Parent Child Resource Centre Limited, and Dr Stanley Ho Medical Development Foundation to P.W.

Data Availability The numeric data and analysis scripts are available on https://osf.io/nqpkal/?view_only=851f70775d8e4659b81cd8d0084df62f.

Compliance with Ethical Standards

Conflict of Interest P.W. and C.T. declare that they are owners of two separate startup companies supported by a Hong Kong Government technology startup scheme. The research reported here is not associated with these two companies. The other authors declare no conflict of interests.

Ethical Approval All procedures have been approved by The Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee and are in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed assent (from each child participant's parent or legal guardian) or informed consent (from each adult participant) approved by The Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee was obtained.

References

- Antoniou, M., To, C. K., & Wong, P. C. (2015). Auditory cues that drive language development are language specific: evidence from Cantonese. *Applied Psycholinguistics*, *36*(6), 1493.
- Baltaxe, C. A. and Simmons, J. Q. (1985). Prosodic development in normal and autistic children. In *Communication problems in autism*, pages 95–125. Springer.
- Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software Articles*, *67*(1), 1–48.
- Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage*, *175*, 56–69.
- Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience*, *33*(3), 530–538.
- Bishop, D. V. (2010). Overlaps between autism and language impairment: phenomimicry or shared etiology? *Behavior Genetics*, *40*(5), 618–629.
- Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., et al. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, *48*(9), 2465–2475.
- Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced pitch sensitivity in individuals with autism: a signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235.
- Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology*, *47*(2), 236–246.
- Chandrasekaran, B., Skoe, E., & Kraus, N. (2014). An integrative model of subcortical auditory plasticity. *Brain Topography*, *27*(4), 539–552.
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.
- Cheng, S. T., Lam, G. Y., & To, C. K. (2017). Pitch perception in tone language-speaking adults with and without autism spectrum disorders. *i-Perception*, *8*(3), 2041669517711200.
- Chevallier, C., Noveck, I., Happe, 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.
- Ciocca, V., & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders*, *1*(2), 141–147.
- Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 1–10.
- Crystal, D. (1969). *Prosodic systems and intonation in English* (Vol. 1). Cambridge: Cambridge University Press.
- Davids, N., Segers, E., Van den Brink, D., Mitterer, H., van Balkom, H., Hagoort, P., & Verhoeven, L. (2011). The nature of auditory discrimination problems in children with specific language impairment: An MMN study. *Neuropsychologia*, *49*(1), 19–28.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: a psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*(2), 144–152.
- Gockley, J., Willsey, A. J., Dong, S., Dougherty, J. D., Constantino, J. N., & Sanders, S. J. (2015). The female protective effect in autism spectrum disorder is not mediated by a single genetic locus. *Molecular Autism*, *6*(1), 25.
- Grossman, R. B., Bemis, R. H., Skwerer, D. P., and Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*.
- Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(2), 701–714.
- Heaton, P. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 543–551.
- Heaton, P. (2005). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 787.
- Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception: An Interdisciplinary Journal*, *15*(3), 291–305.
- Heaton, P., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology*, *25*(6), 771–782.
- Hesling, I., Dilharreguy, B., Peppe, S., Amirault, M., Bouvard, M., & Allard, M. (2010). The integration of prosodic speech in high functioning autism: a preliminary fMRI study. *PLoS ONE*, *5*(7), e11571.
- Ingvalson, E. M., & Wong, P. (2013). Training to improve language outcomes in cochlear implant recipients. *Frontiers in Psychology*, *4*, 263.
- Jarvinen-Pasley, A., & Heaton, P. (2007). Evidence for reduced domain-specificity in auditory processing in autism. *Developmental Science*, *10*(6), 786–793.
- Kargas, N., Lopez, B., Reddy, V., & Morris, P. (2015). The relationship between auditory processing and restricted, repetitive behaviors in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *45*(3), 658–668.

- Kraus, N., Anderson, S., and White-Schwoch, T. (2017). The frequency-following response: a window into human communication. In *The Frequency-Following Response*, pages 1–15. Springer.
- Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically relevant pitch patterns. *Brain and Language*, *110*(3), 135–148.
- Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*(1), 81–95.
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Research. Cognitive Brain Research*, *25*(1), 161–168.
- Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, *189*(1), 1–12.
- Krizman, J., Bonacina, S., & Kraus, N. (2019). Sex differences in subcortical auditory processing emerge across development. *Hearing Research*, *380*, 166–174.
- Lau, J. C., Wong, P. C., & Chandrasekaran, B. (2017). Context-dependent plasticity in the subcortical encoding of linguistic pitch patterns. *Journal of Neurophysiology*, *117*(2), 594–603.
- Lehmann, A., Skoe, E., Moreau, P., Peretz, I., & Kraus, N. (2015). Impairments in musical abilities reflected in the auditory brainstem: evidence from congenital amusia. *European Journal of Neuroscience*, *42*(1), 1644–1650.
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33.
- Lepisto, 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 Research*, *1066*(1–2), 147–157.
- Lim, H. A. (2010). Effect of “developmental speech and language training through music” on speech production in children with autism spectrum disorders. *Journal of Music Therapy*, *47*(1), 2–26.
- Lim, H. A., & Draper, E. (2011). The effects of music therapy incorporated with applied behavior analysis verbal behavior approach for children with autism spectrum disorders. *Journal of Music Therapy*, *48*(4), 532–550.
- Liu, F., Maggu, A. R., Lau, J. C., & Wong, P. C. (2014). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Frontiers in Human Neuroscience*, *8*, 1029.
- Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466–474.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.
- Losh, M., Klusek, J., Martin, G. E., Sideris, J., Parlier, M., & Piven, J. (2012). Defining genetically meaningful language and personality traits in relatives of individuals with fragile x syndrome and relatives of individuals with autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *159*(6), 660–668.
- Losh, M., Martin, G. E., Lee, M., Klusek, J., Sideris, J., Barron, S., & Wassink, T. (2017). Developmental markers of genetic liability to autism in parents: A longitudinal, multigenerational study. *Journal of Autism and Developmental Disorders*, *47*(3), 834–845.
- Losh, M., Sullivan, P. F., Trembath, D., & Piven, J. (2008). Current developments in the genetics of autism: from phenome to genome. *Journal of Neuropathology & Experimental Neurology*, *67*(9), 829–837.
- Maggu, A. R., Liu, F., Antoniou, M., & Wong, P. C. (2016). Neural correlates of indicators of sound change in Cantonese: evidence from cortical and subcortical processes. *Frontiers in Human Neuroscience*, *10*, 652.
- Maggu, A. R., Wong, P. C., Antoniou, M., Bones, O., Liu, H., & Wong, F. C. (2018). Effects of combination of linguistic and musical pitch experience on subcortical pitch encoding. *Journal of Neurolinguistics*, *47*, 145–155.
- Marmel, F., Linley, D., Carlyon, R., Gockel, H., Hopkins, K., & Plack, C. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *Journal of the Association for Research in Otolaryngology*, *14*(5), 757–766.
- Mayer, J. L., Hannent, I., & Heaton, P. F. (2016). Mapping the developmental trajectory and correlates of enhanced pitch perception on speech processing in adults with ASD. *Journal of Autism and Developmental Disorders*, *46*(5), 1562–1573.
- McCann, J., Peppe, S., Gibbon, F. E., O’Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, *42*(6), 682–702.
- Mok, P. P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(03), 341–370.
- Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43.
- Mottron, L., Peretz, I., & Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: beyond central coherence? *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(8), 1057–1065.
- Näätänen, R., Lehtokoski, A., Lenneberg, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, *118*(12), 2544–2590.
- Naigles, L. R. (2013). Input and language development in children with autism. In *Seminars in speech and language*, volume 34, pages 237–248. Thieme Medical Publishers.
- Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research*, *236*(3), 733–743.
- O’Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854.
- Patel, S. P., Kim, J. H., Larson, C. R., & Losh, M. (2019). Mechanisms of voice control related to prosody in autism spectrum disorder and first-degree relatives. *Autism Research*, *12*(8), 1192–1210.
- Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220.
- Peppe, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the peps-c test and the revised version. *Clinical Linguistics & Phonetics*, *17*(4–5), 345–354.
- Peppe, S., McCann, J., Gibbon, F., O’Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*, *38*(10), 1776–1791.
- Peretz, I., Cummings, S., & Dube, M.-P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *The American Journal of Human Genetics*, *81*(3), 582–588.
- R Core Team. (2020). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

- Rinker, T., Kohls, G., Richter, C., Maas, V., Schulz, E., & Schecker, M. (2007). Abnormal frequency discrimination in children with SLI as indexed by mismatch negativity (MMN). *Neuroscience Letters*, *413*(2), 99–104.
- Roberts, T. P., Cannon, K. M., Tavabi, K., Blaskey, L., Khan, S. Y., Monroe, J. F., et al. (2011). Auditory magnetic mismatch field latency: a biomarker for language impairment in autism. *Biological Psychiatry*, *70*(3), 263–269.
- Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, *115*(9), 2021–2030.
- Russo, N., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, *119*(8), 1720–1731.
- Singh, L., & Fu, C. S. (2016). A new view of language development: the acquisition of lexical tone. *Child Development*, *87*(3), 834–854.
- Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302.
- Skoe, E., Krizman, J., & Kraus, N. (2013). The impoverished brain: disparities in maternal education affect the neural response to sound. *Journal of Neuroscience*, *33*(44), 17221–17231.
- Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902.
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, *142*(2), 493–503.
- Tecoulesco, L., Skoe, E., & Naigles, L. R. (2020). Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. *Brain and Language*, *208*, 104810.
- Troyb, E., Knoch, K., and Barton, M. (2011). Phenomenology of ASD: definition, syndromes, and major features. *The neuropsychology of autism*, 9–34.
- Tsou, B., Lee, T., Tung, P., Man, Y., Chan, A., To, C., & Chan, Y. (2006). *Hong Kong Cantonese oral language assessment scale*. Hong Kong: City University of Hong Kong.
- Uwer, R., Albrecht, R., & Von Suchodoletz, W. (2002). Automatic processing of tones and speech stimuli in children with specific language impairment. *Developmental Medicine and Child Neurology*, *44*(8), 527–532.
- Voss, P., Thomas, M. E., Cisneros-Franco, J. M., & de Villiers-Sidani, E. (2017). Dynamic brains and the changing rules of neuroplasticity: implications for learning and recovery. *Frontiers in Psychology*, *8*, 1657.
- 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.
- Wong, P. C., Ciocca, V., Chan, A. H., Ha, L. Y., Tan, L.-H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PLoS One*, *7*(4), e33424.
- Wong, P. C., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421.
- Wong, P. C., Kang, X., Wong, K. H., So, H.-C., Choy, K. W., & Geng, X. (2020). Aspm-lexical tone association in speakers of a tone language: direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances*, *6*(22), eaba5090.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422.
- Xie, Z., Reetzke, R., & Chandrasekaran, B. (2018). Taking attention away from the auditory modality: context-dependent effects on early sensory representation of speech. *Neuroscience*, *384*, 64–75.
- Xie, Z., Reetzke, R., & Chandrasekaran, B. (2019). Machine learning approaches to analyze speech-evoked neurophysiological responses. *Journal of Speech, Language, and Hearing Research*, *62*(3), 587–601.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S., & Zhang, Y. (2015). Pitch processing in tonal-language speaking children with autism: an event-related potential study. *Journal of Autism and Developmental Disorders*, *45*(11), 3656–3667.
- Zhang, J., Meng, Y., Wu, C., Xiang, Y.-T., & Yuan, Z. (2019). Non-speech and speech pitch perception among Cantonese-speaking children with autism spectrum disorder: an ERP study. *Neuroscience Letters*, *703*, 205–212.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.