

# *Predictive processing of music and language in autism: evidence from Mandarin and English speakers*

Article

Accepted Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Zhao, C., Ong, J. H. ORCID: <https://orcid.org/0000-0003-1503-8311>, Veic, A., Patel, A. D., Jiang, C., Fogel, A. R., Wang, L., Hou, Q., Das, D., Crasto, C., Chakrabarti, B. ORCID: <https://orcid.org/0000-0002-6649-7895>, Williams, T. I., Loutrari, A. and Liu, F. ORCID: <https://orcid.org/0000-0002-7776-0222> (2024) Predictive processing of music and language in autism: evidence from Mandarin and English speakers. *Autism Research*, 17 (6). pp. 1230-1257. ISSN 1939-3806 doi: 10.1002/aur.3133 Available at <https://centaur.reading.ac.uk/115919/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/aur.3133>

Publisher: John Wiley

the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Predictive processing of music and language in autism: Evidence from Mandarin and English speakers

Chen Zhao<sup>1</sup> | Jia Hoong Ong<sup>1</sup>  | Anamarija Veic<sup>1</sup> | Aniruddh D. Patel<sup>2,3</sup> |  
 Cunmei Jiang<sup>4</sup>  | Allison R. Fogel<sup>2</sup> | Li Wang<sup>1</sup>  | Qingqi Hou<sup>5</sup> |  
 Dipsikha Das<sup>6</sup> | Cara Crasto<sup>1</sup> | Bhismadev Chakrabarti<sup>1</sup>  | Tim I. Williams<sup>1</sup> |  
 Ariadne Loutrari<sup>1</sup> | Fang Liu<sup>1</sup> 

<sup>1</sup>School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

<sup>2</sup>Department of Psychology, Tufts University, Medford, Massachusetts, USA

<sup>3</sup>Program in Brain, Mind, and Consciousness, Canadian Institute for Advanced Research (CIFAR), Toronto, Canada

<sup>4</sup>Music College, Shanghai Normal University, Shanghai, China

<sup>5</sup>Department of Music and Dance, Nanjing Normal University of Special Education, Nanjing, China

<sup>6</sup>School of Psychology, Keele University, Staffordshire, UK

## Correspondence

Fang Liu, School of Psychology & Clinical Language Sciences, University of Reading, Earley Gate, Reading RG6 6AL, UK.

Email: [f.liu@reading.ac.uk](mailto:f.liu@reading.ac.uk)

## Present addresses

Chen Zhao, School of Health Sciences, University of Manchester, Manchester, UK; Jia Hoong Ong, Department of Psychology, School of Social Sciences, Nottingham Trent University, Nottingham, UK; and Li Wang, Brain and Mind Institute, Shenzhen Research Institute, The Chinese University of Hong Kong, Central District, Hong Kong.

## Funding information

European Research Council

## Abstract

Atypical predictive processing has been associated with autism across multiple domains, based mainly on artificial antecedents and consequents. As structured sequences where expectations derive from implicit learning of combinatorial principles, language and music provide naturalistic stimuli for investigating predictive processing. In this study, we matched melodic and sentence stimuli in cloze probabilities and examined musical and linguistic prediction in Mandarin- (Experiment 1) and English-speaking (Experiment 2) autistic and non-autistic individuals using both production and perception tasks. In the production tasks, participants listened to unfinished melodies/sentences and then produced the final notes/words to complete these items. In the perception tasks, participants provided expectedness ratings of the completed melodies/sentences based on the most frequent notes/words in the norms. While Experiment 1 showed intact musical prediction but atypical linguistic prediction in autism in the Mandarin sample that demonstrated imbalanced musical training experience and receptive vocabulary skills between groups, the group difference disappeared in a more closely matched sample of English speakers in Experiment 2. These findings suggest the importance of taking an individual differences approach when investigating predictive processing in music and language in autism, as the difficulty in prediction in autism may not be due to generalized problems with prediction in any type of complex sequence processing.

## Lay Summary

Predictive processing plays a central role in everyday life. While autism has been associated with atypical prediction across multiple domains, evidence has been mixed and mainly based on artificial rather than naturalistic associations. In two experiments, this study examined predictive processing in Mandarin and English speakers with autism across music and language domains using naturalistic melodic and sentence stimuli through production and perception tasks. Our results reveal musical training experience and receptive vocabulary as confounding factors influencing predictive processing of music and language in autism, suggesting that autism may not be linked to generalized problems with prediction in any type of complex sequence processing.

## KEY WORDS

autism, cloze probability, language, music, prediction

## INTRODUCTION

The human brain actively makes predictions of upcoming events through their associations with the current context (Bar, 2007; Bendixen, 2014). This predictive process enables efficient adaptation to a dynamically changing world (Clark, 2013; Sinha et al., 2014). In daily interactions, prediction facilitates language comprehension and speech communication through rapid and accurate anticipation of upcoming words in a sentence based on existing knowledge and experience (Kuperberg & Jaeger, 2016; Miller & Selfridge, 1950). Similar to language, music also has a hierarchical structure that unfolds rapidly in time following tonal and syntactic rules (Krumhansl & Kessler, 1982; Krumhansl & Shepard, 1979; Patel, 2013; Patel & Morgan, 2017). Consequently, music perception and appreciation also require predictive processing in deriving anticipation and gaining enjoyment/pleasure (Cheung et al., 2019; Gold et al., 2019; Huron, 2006). The similarities between language and music in hierarchical predictive processing make them excellent candidates to investigate the mechanisms of prediction in comparable domains (Fogel et al., 2015; Patel & Morgan, 2017).

Regardless of domain, predictive processing involves learning the regularities between antecedents and consequents, and detecting and applying the learned associations to a similar situation (Cannon et al., 2021; Perruchet & Pacton, 2006). Apart from top-down mechanisms, prediction in language and music also builds on implicit statistical learning, a bottom-up process (Emberson et al., 2013) where the probability of occurrence of an upcoming event is predictable based on a given context (Conway et al., 2010; Fogel et al., 2015; Kuperberg & Jaeger, 2016; Miller & Selfridge, 1950; Morgan et al., 2019). While predictive processing is evident in early infancy as a learning mechanism (Háden et al., 2015; Köster et al., 2020; Trainor, 2012), extensive research has shown that autism spectrum disorder (autism hereafter) is associated with atypical predictive skills, including predictive learning and predictive response (Cannon et al., 2021; Lawson et al., 2014; Pellicano & Burr, 2012; Sinha et al., 2014).

Autism is a neurodevelopmental condition characterized by reduced social communication and social interaction skills, repetitive and restricted behaviors and interests, and atypical sensory processing (DSM-5, American Psychiatric Association, 2013). Despite mixed results in the literature (Cannon et al., 2021), autistic individuals generally demonstrate atypical predictive skills in social functioning (Greene et al., 2019; Kinard et al., 2020; Palumbo et al., 2015), visual processing (Karaminis et al., 2016; Sheppard et al., 2016), auditory processing (Font-Alaminos et al., 2020; Goris et al., 2018), theory of mind (Balsters et al., 2017), recognizing emotions (Leung et al., 2022, but see Leung et al., 2023; Zhang et al., 2022), and action prediction (Amoruso et al., 2019;

Schuwerk et al., 2016). According to the most recent computational theoretical accounts, prediction difficulties in autism may arise due to imbalanced precision at higher (e.g., less precise prior beliefs) versus lower (e.g., more precise sensory inputs) levels (Brock, 2012; Lawson et al., 2014; Pellicano & Burr, 2012), atypical contextual modulation of this balance (Van de Cruys et al., 2014), difficulties in learning regularities (i.e., statistical learning ability; Sinha et al., 2014), imbalanced processing of global and local regularities (Xu et al., 2022), and the reduced speed of integrating new information to guide behavior—the “slow-updating” hypothesis (Lieder et al., 2019; Vishne et al., 2021).

In the language domain, autistic individuals show difficulties in semantic prediction and/or language comprehension (Booth & Happé, 2010; Frith & Snowling, 1983; Happé, 1997). They have a greater tendency than neurotypical individuals to complete sentences like “In the sea there are fish and ...” in a local manner (“chips”) than a global manner (“sharks”) (Booth & Happé, 2010). Autistic children also tend not to adjust the pronunciation of homographs based on their semantic/syntactic context, e.g., by pronouncing “BOW” (/bəʊ/ vs. /baʊ/) incorrectly in a certain context such as “He had a pink BOW” versus “He made a deep BOW” (Frith & Snowling, 1983). Based on this evidence, weak central coherence (WCC) theory proposes that autistic individuals have reduced central coherence due to their preference for the parts over the whole (Frith & Happé, 1994; Happé, 1997). However, other studies suggest that irrespective of an autism diagnosis, individuals’ language abilities affect behavioral performance during language processing, since poorer performance (i.e., reduced sensitivity to sentence context) is associated with poorer language scores (Brock et al., 2008; Norbury, 2005). Mixed results have also been reported in neuroimaging studies. Some studies indicate atypical N400 responses (an index of semantic processing) (Ring et al., 2007) and restricted neural networks in autism when processing sentences with semantically congruent versus incongruent endings (Catarino et al., 2011). Other studies show that semantic processing may be preserved in autism, although atypical processing is observed at a later stage due to reduced top-down control (Henderson et al., 2011), or is associated with delayed processing speed (DiStefano et al., 2019).

In the music domain, autism has been associated with typical or extraordinary musical skills, including perception of pitch, melody, and musical emotions (Chen et al., 2022; Janzen & Thaut, 2018; O’Connor, 2012; Ouimet et al., 2012; Quintin, 2019; Wang, Ong, et al., 2023), as well as enculturation to the pitch structure of Western music (DePape et al., 2012). However, reduced performance was also observed in pitch, emotion, and melodic processing (Bhatara et al., 2010; Ong et al., 2023; Sota et al., 2018), beat synchronization (Kasten et al., 2023; Morimoto et al., 2018; Vishne et al., 2021), active rhythmic engagement (Steinberg et al., 2021),

and metrical enculturation in autism (DePape et al., 2012). In a musical imagery task involving predictive processing, despite having impaired language abilities, autistic children showed comparable or better performance than non-autistic children in judging pitch and tempo manipulations of the continuations of familiar song excerpts (Heaton et al., 2018). Furthermore, autistic individuals showed intact predictive processing of rhythmic tones when presented with standard and deviant rhythmic tone sequences (Knight et al., 2020). Additionally, autistic children, adolescents, and adults were able to identify positive and negative emotions in music (Gebauer et al., 2014; Heaton et al. 1999; Quintin et al., 2011), using neural networks in cortical and subcortical brain areas that are typically implicated for emotion processing and reward (Fröhholz et al., 2016), such as the amygdala, ventral striatum, medial orbitofrontal cortex, ventral tegmental area, and caudate nucleus (Caria et al., 2011; Gebauer et al., 2014). Since both the perception/appreciation of music and musical emotion identification require prediction (Cheung et al., 2019; Gold et al., 2019), it could be assumed that autistic individuals have largely intact predictive processing of music based on the above evidence.

Taken together, previous literature on linguistic and musical prediction in autism seems to suggest atypical performance in the language domain but not in the music domain, even though predictive processing involves implicit statistical models in both domains (Conway et al., 2010; Fogel et al., 2015; Miller & Selfridge, 1950; Morgan et al., 2019). However, there has been no direct comparative investigation of prediction based on naturally learned associations across language and music in autism. It remains unclear whether autistic individuals would show difficulties with prediction as “future-directed information processing” in hierarchically structured sequences as constrained by temporal, syntactic, or semantic rules across the two domains (Ferreira & Chantavarin, 2018; Koelsch et al., 2019; Kuperberg & Jaeger, 2016; Patel, 2003; Rohrmeier & Koelsch, 2012; Slevc, 2012). In addition, most studies used artificial and arbitrary cue-outcome associations rather than naturalistic antecedents and consequents to examine predictive skills in autism (Cannon et al., 2021). Using ecologically valid musical and linguistic stimuli, the current study examined naturally formed statistical predictive behavior across domains in autism, adopting the approach by recent studies of musical and linguistic prediction (Fogel et al., 2015; Patel & Morgan, 2017).

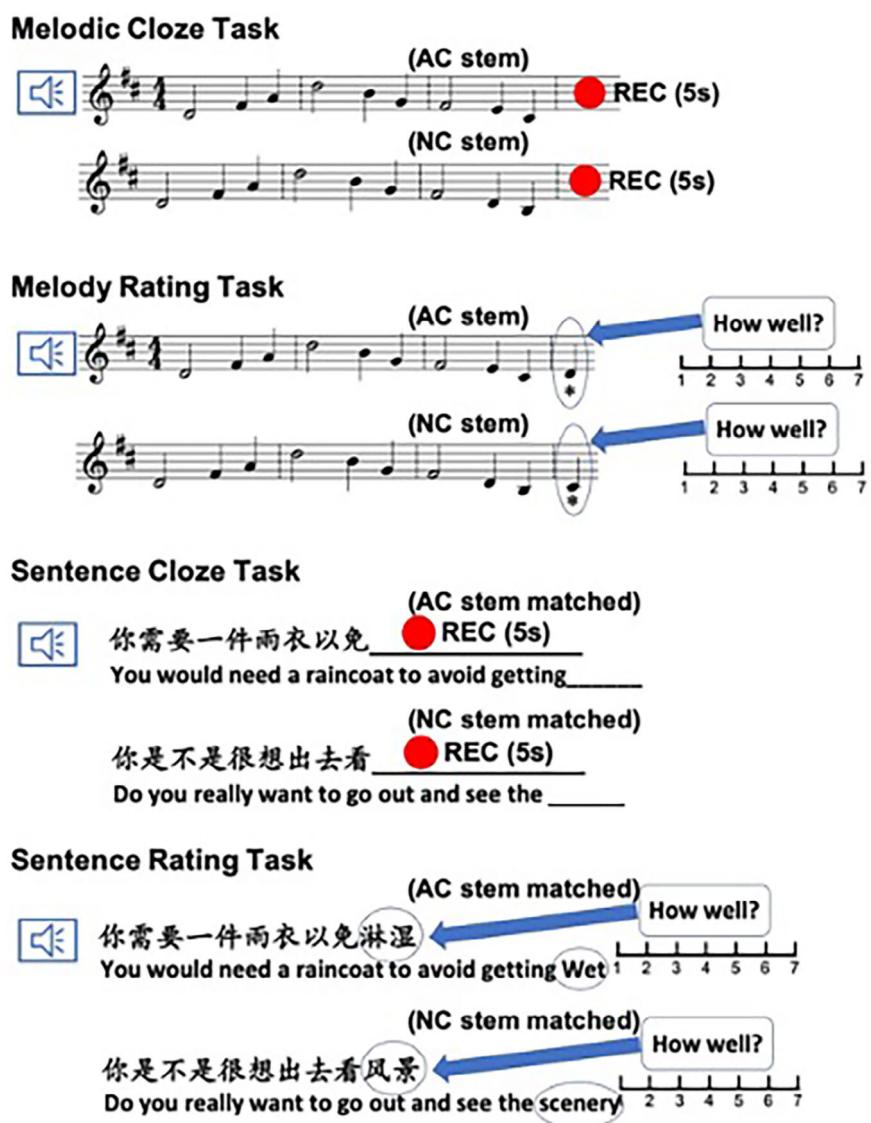
Specifically, prediction in language can be examined through sentence processing tasks, during which the brain pre-activates forthcoming words following a probabilistic approach based on the current context (DeLong et al., 2005; Kutas & Hillyard, 1984; Nieuwland et al., 2020). Sentence cloze tasks have been widely used to assess linguistic predictive ability, where participants are required to complete a context/sentence with an

appropriate word/phrase (Chik et al., 2012; Di Vesta et al., 1979; Neville & Pugh, 1976; Taylor, 1953). The production of a word/phrase when completing a sentence in a cloze task involves contextual semantic processing and a degree of expectation (DeLong et al., 2005; Nieuwland et al., 2020). Based on the context, highly expected words are more likely to be produced than less expected words. For example, the highly expected word “wet” is more frequently produced than the word “cold” when completing the sentence “You need a raincoat to avoid getting \_\_\_,” even though both words are semantically correct in this context. The degree of expectation for a particular word/phrase in a sentence is usually determined by calculating the proportion of people (cloze probability) who have completed the sentence using the word/phrase (Taylor, 1953). Behaviourally, words with high cloze probabilities are activated more rapidly and read with faster speeds than low probability words (Smith & Levy, 2013; Staub et al., 2015). Linguistic prediction can thus be assessed through production of final words/phrases in sentence cloze tasks (Staub et al., 2015), or through perception of final words/phrases in semantic congruency tasks where participants judge the semantic predictability of a final word, which can either be congruent, neutral, or anomalous in a sentence context (Stringer & Iverson, 2020).

Like language processing, music perception also involves multiple levels of prediction (Koelsch et al., 2019; Patel & Morgan, 2017; Vuust et al., 2022) and engages probabilistic predictive processing (Egermann et al., 2013; Morgan et al., 2019; Pearce & Wiggins, 2006). Using perceptual rating tasks, studies of music expectation showed that notes/chords that violate harmonic or tonal structure are deemed unexpected by both musicians and non-musicians (Egermann et al., 2013; Jiang et al., 2016; Marmel et al., 2008; Schellenberg, 1996; Steinbeis et al., 2006). Recently, Fogel et al. (2015) developed a novel melodic cloze task asking participants to sing the next note after hearing the opening of an unfamiliar, naturalistic tonal melody. Responses demonstrated that musical expectancy is influenced by statistical learning of note transition probabilities, gestalt principles of auditory processing, and the tonal hierarchy and implied harmonic structure (Fogel et al., 2015; Morgan et al., 2019; Verosky & Morgan, 2021). Thus, both the melodic cloze task and the sentence cloze task allow the direct comparison of predictive processing through production tasks across music and language (Fogel et al., 2015).

Capitalizing on these recent advances in the field we employed the melodic cloze task from Fogel et al. (2015) and created a matched sentence cloze task to compare predictive production in music and language in autism (Figure 1). Specifically, we matched the items (melodic or sentence stems) in the number of notes/syllables and in the cloze probabilities of the most frequently produced final notes/words, based on the norms. The tasks required

**FIGURE 1** An example of the melodic and sentence cloze production tasks and perceptual rating tasks of the whole melodies/sentences. REC shows the point where participants were prompted to produce the next note or word, and were recorded for 5 seconds. AC (authentic cadence) and NC (non-cadence) melodic pairs are matched for number of notes and rhythm but differ in their implied harmonic structure in a way that leads to distinct melodic expectations. The sentence stems were matched with the melodic stems based on the cloze probabilities and number of syllables (see Methods [Section 3.2] for details).



participants to either sing the note or say the word/phrase they expected to come next in a melody/sentence. We included a pitch imitation task to assess participants' pitch matching abilities because musically untrained participants may be less able to sing accurately (Dalla Bella & Berkowska, 2009; Pfördresher et al., 2010). We also included perceptual rating tasks of completed versions of the melodies and sentences to examine whether participants' perceptual ratings of the melodies and sentences correlated with their production performance.

Finally, evidence suggests that pitch, music, and language processing abilities in autism interact with each other (Eigsti & Fein, 2013; Germain et al., 2019; Globerson et al., 2015; Jones et al., 2009; Wang et al., 2022), and at the same time they are also modulated by cognitive abilities such as non-verbal IQ, receptive vocabulary, and memory (Chowdhury et al., 2017; Jamey et al., 2019; Kargas et al., 2015; Ong et al., 2023) and impacted by age (Jamey et al., 2019; Mayer et al., 2016; Ong et al., 2023; Wang et al., 2021)

and language background (Wang, Xiao, et al., 2023; Yu et al., 2015). Put simply, differences in music and/or language processing between autistic and non-autistic individuals may be due to differences in cognitive abilities and individual factors rather than autism per se. Thus, we reported data from a Mandarin-speaking sample (Experiment 1) and an English-speaking sample (Experiment 2) and compared group performance with and without including cognitive factors, age, pitch matching ability, and musical training experience in the models. We hypothesized that, across both samples, the autistic group would show intact prediction in the music domain but atypical prediction in the language domain, with participants' production performance associated with their perceptual ratings in both domains. We also predicted that background measures including age, pitch, musical, and cognitive abilities would impact predictive processing of music and language across both Mandarin and English samples of autistic and non-autistic individuals.

## EXPERIMENT 1

### Method

#### Participants

Thirty-one autistic participants (4 females) were recruited from autism centres in Nanchang and Nanjing, China, and 32 age-matched non-autistic participants (5 females) were recruited from local mainstream schools and the University of Nanchang. All participants in the autistic group had a professional clinical diagnosis of autism, which was further confirmed using the Autism Diagnostic Observation Schedule, second edition (ADOS-2, Lord et al., 2012) by author LW (with clinical and research reliability for administration and scoring). Participants in the non-autistic group reported no neurodevelopmental/psychiatric disorders or a family history of autism. Two participants in the autistic group did not complete the melody rating task, due to fatigue, loss of interest, or difficulty in concentrating on the task, and their data are marked as "NA" in the dataset. All participants had normal hearing with pure-tone air conduction thresholds of 25 dB HL or better at frequencies of 0.5, 1, 2, and 4 kHz. The study protocol was approved by the research ethics committees of University of Reading and Shanghai Normal University. Parents provided written informed consent for their children's participation.

#### Background measurements

Participants' verbal ability was estimated using the Peabody Picture Vocabulary Test, revised edition (PPVT-R, Dunn & Dunn, 1981), and nonverbal intelligence assessed with Raven's Standard Progressive Matrices (RSPM, Raven et al., 1998). Given that the Chinese norms for PPVT-R only included ages from 3.5 to 9 (Sang & Miao, 1990), standardized scores were calculated based on American norms (Dunn & Dunn, 1981). Correlation analysis revealed a significant positive relationship between the standardized scores obtained based on the Chinese norms and those based on the American norms ( $r = 0.95$ ) for participants at or below 9 years old, thus confirming the validity of this approach. RSPM scores were normalized using the means and standard deviations across different age ranges based on a Chinese sample (Zhang & Wang, 1989). Participants' short-term memory was tested using the forward Digit Span task (Wechsler, 2008), implemented via the Psychology Experiment Building Language (PEBL) test battery (Piper et al., 2016), where digit span was calculated as the maximum number of digits correctly recalled at least once among two trials of the same length of digits. Demographic information and musical training experience of participants were collected through a questionnaire, where musical training in years was calculated by

summing across experience with all instruments including voice (Pfordresher & Halpern, 2013). As can be seen from Table 1, the two groups were matched on age, gender, musical training experience, non-verbal IQ, and digit span. Although both groups showed advanced receptive vocabulary skills, the non-autistic group scored significantly higher than the autistic group on this measure.

#### Cloze production and pitch matching tasks

##### Stimuli

Stimuli in the melodic cloze task were from Fogel et al. (2015), which included 45 pairs of melodic stems in 12 major keys, three meters (3/4, 4/4, and 6/8), and with a tempo of 120 beats per minute. Containing 5–9 notes, the stems in each pair had the same length, rhythm, and melodic contour, but differed in the underlying harmonic structure as influenced by the pitch of some of the notes. In each pair, one stem implied an authentic cadence (AC) at the end, and the other did not (non-cadence, NC). Whereas each AC stem elicited a strong expectation for a particular subsequent note (the tonic, or central tone of the prevailing key), the NC stem did not create a strong expectation for any particular subsequent note. Thus, AC stems were "high constraint" and NC stems were "low constraint" in terms of how they constrained expectations for the subsequent note. Two versions of the stems were created, with the latter one octave lower than the former, for use with males vs. females (Fogel et al., 2015). Although Western music and traditional Chinese music have different systems, owing to globalization Chinese participants are widely exposed to Western music and its tonal system (Huang, 2012). Previous studies have reported that Mandarin speakers who were not musicians could differentiate tonal regularities from irregularities and were sensitive to tonality and emotions in Western music (Fang et al., 2017; Jiang et al., 2016, 2017; Sun et al., 2020; Zhou et al., 2019). Thus, the use of Western tonal melodies from Fogel et al. (2015) was ecologically valid for our Mandarin-speaking participants.

Matching the cloze probabilities of the melodic stems, a list of 204 sentence stems were selected based on the norms established from previous studies in English (Arcuri et al., 2001; Block & Baldwin, 2010) and translated into Chinese. To establish the cloze probabilities of these sentence stems in Chinese, a validation study was conducted with a group of Mandarin-speaking neurotypical adults ( $n = 34$ ) via an online survey (<https://www.onlinesurveys.ac.uk>). Using the same instruction as in Block and Baldwin (2010), these volunteers were asked to provide a word/phrase that they thought would best complete each of the sentence stems. The cloze probability for each sentence was then calculated according to the participants' responses. A final battery of 90 cloze sentences was selected based on their cloze probabilities and numbers of syllables, individually matched with those of the

**TABLE 1** Demographic characteristics of the autistic and non-autistic groups in Experiment 1 on Mandarin speakers.

| Measures             | Autistic group | Non-autistic group | t or chi-squared test between groups |
|----------------------|----------------|--------------------|--------------------------------------|
| n                    | 31             | 32                 | NA                                   |
| Age                  | 10.49 (2.52)   | 11.47 (2.71)       | $t(61) = -1.46, p = 0.150$           |
| Age range            | 7.00–15.91     | 7.55–15.69         | NA                                   |
| Sex                  | F = 4, M = 27  | F = 5, M = 27      | $\chi^2(1) = 0.00, p = 1$            |
| Musical training     | 0.81 (1.21)    | 0.47 (1.06)        | $t(61) = 1.16, p = 0.250$            |
| NVIQ                 | 0.72 (0.96)    | 0.86 (0.65)        | $t(61) = -0.68, p = 0.500$           |
| Receptive vocabulary | 129.06 (24.28) | 141.41 (12.63)     | $t(61) = -2.50, p = 0.015$           |
| Digit span           | 8.35 (1.03)    | 8.13 (1.08)        | $t(61) = 0.85, p = 0.400$            |

Note: F = female, M = male. NVIQ: standard score of Raven's standard progressive matrices; Receptive vocabulary: standard score of The Peabody Picture Vocabulary Test, revised edition (PPVT-R).

**TABLE 2** Characteristics of the melodic and sentence stems used in Experiment 1.

|   | Melodic stem |       |       | Sentence stem |       |       |
|---|--------------|-------|-------|---------------|-------|-------|
|   | Range        | Mean  | SD    | Range         | Mean  | SD    |
| Cloze probability (%)                       | 20.00–100.00 | 55.37 | 21.32 | 14.71–97.06   | 55.78 | 21.59 |
| Length <sup>a</sup>                         | 6–9          | 8.40  | 0.83  | 6–9           | 8.36  | 0.87  |
| Pitch range (st <sup>b</sup> ) <sup>c</sup> | 5.20–29.11   | 14.83 | 6.01  | 5.80–28.91    | 13.47 | 6.43  |
| Duration (s)                                | 2.50–8.66    | 5.02  | 1.23  | 1.52–2.74     | 2.23  | 0.25  |

<sup>a</sup>The number of notes/syllables.

<sup>b</sup>Semitone.

<sup>c</sup>The mean distance between the highest and lowest pitch in the stem.

melodic stems (within  $\pm 5\%$  difference in cloze probability and within  $\pm 2$  difference in the number of notes/syllables). These sentence stems were then recorded by a native female speaker of Mandarin using Praat (Boersma & Weenink, 2001), with 44.1 kHz sampling rate and 16-bit amplitude resolution. The details of the melodic and sentence stems are shown in Table 2.

Using a Latin square design, the melodic stems were pseudorandomised into 8 lists of 45 items, among which 22/23 were AC stems and the rest were NC stems (Fogel et al., 2015). The stems from the same pair appeared in different lists, which were assigned to participants in a counterbalanced order. Using the same randomization method as in the melodic cloze task, the 90 matched sentence stems were also grouped into 8 lists of 45 items and presented in counterbalanced order across participants.

### Procedure

The experiment was carried out in classrooms in local autism centres in Nanchang and Nanjing, China. Cloze stems were presented with PsychoPy (version 1.9.1) through Sennheiser HD280 pro headphones connected to a laptop via a Roland RUBIX22 USB Audio Interface. Prior to the melodic cloze task, participants' note production accuracy was evaluated using a pitch matching task. Eight notes were played one at a time: F4, A4, B3, G#4, A#3, D4, C#4, and Eb4, corresponding to 349.2, 440.0, 246.9, 415.3, 233.1, 293.7, 277.2, 311.1 Hz, respectively,

for female participants. Notes that were one octave lower were used for male participants, with fundamental frequencies at 174.6, 220.0, 123.5, 207.7, 116.6, 146.9, 138.6, and 155.6 Hz. Participants were instructed to imitate the pitch of the notes as closely as possible. In the melodic cloze task, after hearing a melodic stem, participants were instructed to "sing the note you think comes next" by humming or on a syllable of their own choice (e.g., "la", "da", etc.) within a 5-sec recording window. In the sentence cloze task, after hearing a sentence stem, participants were instructed to "say the word/phrase you think best completes the sentence" within a 5-sec recording window (Figure 1). A short practice session was presented to familiarize participants with the task procedure and stimuli, using different cloze melodies/sentences from the actual task. From the practice sessions, all participants understood the 5-sec response window as required when performing the tasks.

### Perceptual rating tasks

#### Stimuli

In the rating tasks, the stimuli included the melodies and sentences completed/produced in full, including the last notes/words that had the highest cloze probabilities based on the norms. The norms for the notes were from Fogel et al. (2015), and the norms for the words were from the validation study (described above).

### Procedure

In the rating tasks, which followed the production task, participants were told that “In the previous production tasks, you have tried to complete the melodies/sentences yourself. In this task, another person has completed the melodies/sentences using their preferred notes/words. On a scale of 1 to 7, please rate how well you think the last note/word(s) continues/completes the melody/sentence (1: very badly; 7: very well). Would you use the same note/word(s) to continue/complete the melody/sentence?” To familiarize participants with the rating tasks, a short practice session was presented, using the same melodies/sentences as in the cloze practice sessions but with the last note/word(s) added.

### Data analysis

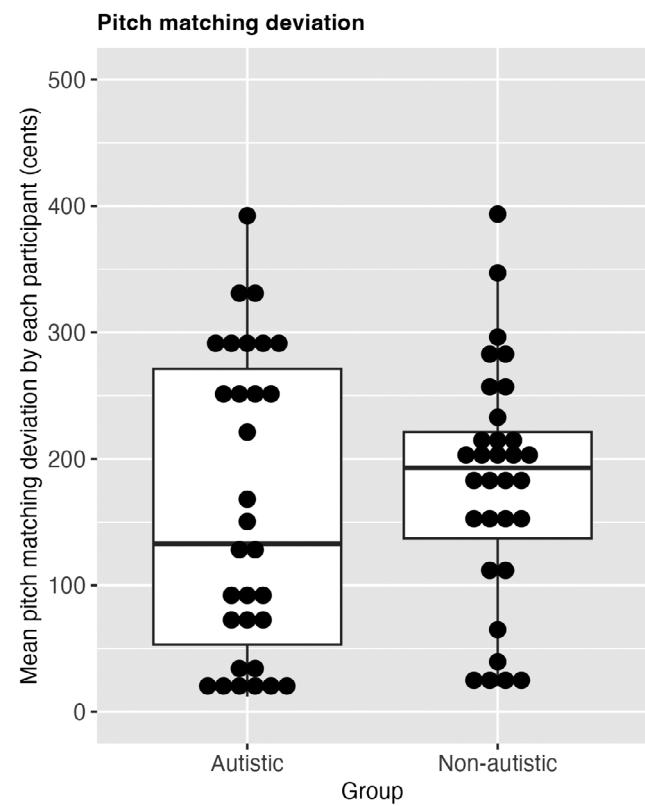
Participants’ production data were analyzed offline. The sung notes in the melodic cloze and pitch matching tasks were manually labeled, and their fundamental frequencies ( $F_0$ ) were extracted using ProsodyPro (Xu, 2013) in Praat (Boersma & Weenink, 2001). For the pitch matching task, the accuracy of note production was assessed individually, and the deviations (in cents; 100 cents = 1 semitone) in pitch from the actual notes were averaged across the eight notes for each participant. The sung notes produced in the melodic cloze task were also assessed individually for each participant. The  $F_0$  of each sung note was matched to the closest semitone (within 50 cents deviation) in the Western chromatic scale (e.g., A4 = 440 Hz). When analyzing the pitch matching and melodic cloze production data, we allowed octave jumps by the participants due to their different vocal ranges so that the final  $F_0$  values were adjusted (by  $\pm 12$  semitones) before comparing them against the expected notes.

Using custom-written Matlab scripts (The MathWorks Inc., 2022), participants’ sung notes were categorized into four categories based on the norms from Fogel et al. (2015): 1. no response; 2. a note that has not been reported in the norms; 3. a less frequent note from the norms; 4. the most frequent note from the norms. Participants’ sentence cloze production was transcribed and categorized offline by author CZ, a Mandarin speaker, into four categories: 1. no response; 2. a grammatically/semantically incorrect word/phrase that has not been reported in the norms; 3. a less frequent word/phrase from the norms, or a grammatically and semantically correct word/phrase not from the norms; 4. the most frequent word/phrase from the norms. An independent Mandarin-speaking research assistant also coded 33 out of a total of 63 datasets, with high inter-rater reliability ( $\kappa = 0.965$ ,  $p < 0.001$ ). Disagreements were resolved through discussions between the two coders. No co-coding was done for the melodic cloze categories since categorization was done automatically by Matlab scripts

(The MathWorks Inc., 2022). In addition, reaction time (RT) during the cloze production tasks was measured as the time between the offset of a stem and the onset of a vocalization using Praat (Boersma & Weenink, 2001).

For statistical analyses,  $t$  tests were used to compare group performance on pitch matching ability. Counts (in percentage) of the four response categories in the music and language tasks were summed and tabulated for each group, and chi-squared tests were used to evaluate if there were any group differences. The four response categories were then simplified into two categories and converted to binomial data as 1 (correct: the most frequent responses in the norms) or 0 (incorrect: no or all other responses).

Participants’ binomial production responses (1 or 0) and their corresponding RT for producing the most frequent responses, as well as the perceptual rating data, were then analyzed using generalized linear mixed-effects models (i.e., logistic regression for production responses) or linear mixed-effects models (log-transformed RT data and rating) in R (version R-3.6.0; R: A Language and Environment for Statistical Computing, 2019), with packages of *lme4* (Bates et al., 2015), *car* (Fox & Weisberg, 2018), and *lmerTest* (Kuznetsova et al., 2017). For each measure, we modeled a “simple model” across both music and language tasks, which consisted only of group (autistic versus non-autistic), task (music versus language), constraint (high versus low), and all interactions to examine whether there were any group differences in the measure when other background measures were not considered. Then, for the same measure, we modeled a “full model” to account for background measures and stimulus properties. Each full model consisted of the following predictors: group (autistic versus non-autistic), task (music versus language), constraint (high versus low), and all possible interactions between the three, as well as age (a continuous predictor), years of musical training (a continuous predictor), non-verbal IQ (a continuous predictor), receptive vocabulary (a continuous predictor), digit span (a continuous predictor), pitch matching deviation (a continuous predictor), and stimulus duration (a continuous predictor). To examine the relationship between production and perception, perceptual ratings of the items were added as a predictor for the production categories, and vice versa. In all the models, categorical predictors were effect-coded and continuous predictors were mean-centered. Mean centring for stimulus duration was done by task (i.e., separately for music and language) as the stimulus duration for the musical stimuli were longer than the linguistic stimuli (see Table 2). The variance inflation factor (VIF) was calculated to check multi-collinearity among the independent variables. Given that the VIF values were all smaller than 5, multi-collinearity among these independent variables was low (O’Brien, 2007). In each model, the subject and item random intercepts were included as random effects. We tried modeling the random effects as maximal



**FIGURE 2** Mean pitch matching deviations among Mandarin speakers in Experiment 1.

as possible (Barr et al., 2013), but due to convergence issues in some of the models, by-item and/or by-subject slopes were removed.

Statistical significance of the fixed effects was tested using the *Anova()* function from the *car* package for linear models (Type III Analysis of Variance Wald chisquare tests). Post-hoc analyses of the interaction effects were investigated with the *emmeans* package with *p* values adjusted using the *holm* method (Lenth et al., 2020) in R. Effect sizes for each predictor in the binomial models (production) were estimated using odds ratio whereas those in the linear models (log-transformed RT and rating) were estimated using  $R^2$  using the *r2beta()* function from the *r2glmm* package (Jaeger et al., 2017). In the interest of space, only statistically significant effects/interactions are reported in the Results section below, with the entire model outputs displayed in Supplementary Tables S1.1–S1.6.

## Results

### Pitch matching

Figure 2 shows the boxplots of pitch matching deviations by each participant in each group. There was no significant difference in pitch matching deviation between the

two groups ( $t(61) = -0.71$ ,  $p = 0.482$ ; autistic mean (SD) = 161.69 (116.29); non-autistic mean (SD) = 180.65 (91.94)).

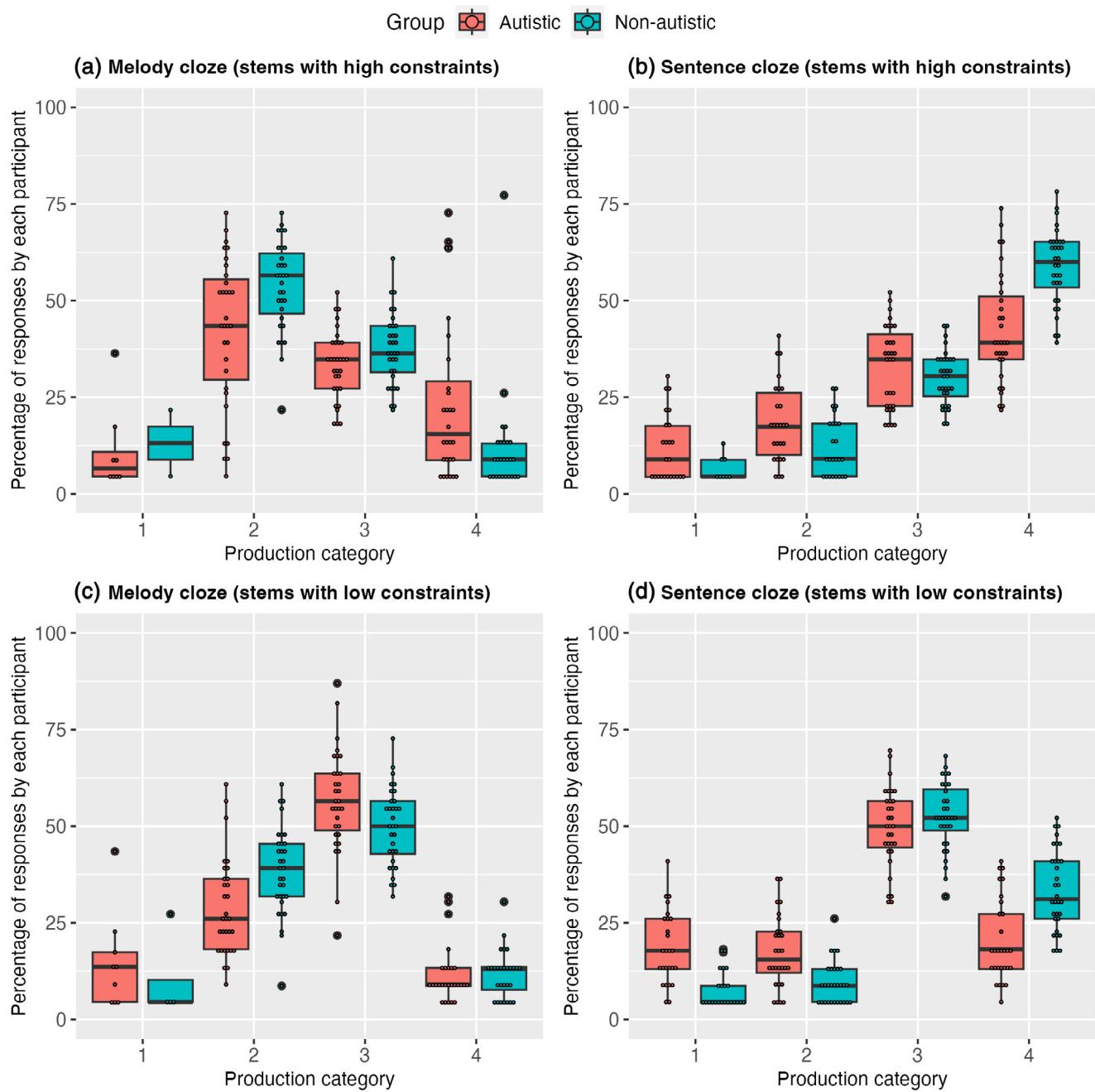
### Melodic and sentence cloze production tasks

Figure 3 shows the boxplots of percentages of the responses from the four categories by each participant in each group separated by the task (music vs. language) and constraint (high vs. low) conditions, and Table 3 displays the response counts in percentage from the four categories and chi-squared tests comparing the distribution of the percentages by groups. The two groups differed significantly only in their response categories for the language task, but not the music task, regardless of whether the items had high-constraints or low-constraints.

### Production of the most frequent responses

In the simple model on participants' binomial production responses (1: the most frequent responses based on the norms; 0: no or other responses; see Figure 4), significant effects of Constraint ( $\chi^2(1) = 24.21$ ,  $p < 0.001$ , Odds Ratio (OR) = 1.51) and of Task ( $\chi^2(1) = 79.47$ ,  $p < 0.001$ , OR = 2.33) were observed, as well as all the two-way interactions between Group, Constraint, and Task (Group  $\times$  Task:  $\chi^2(1) = 29.00$ ,  $p < 0.001$ , OR = 0.71; Group  $\times$  Constraint:  $\chi^2(1) = 9.58$ ,  $p = 0.002$ , OR = 1.15; Task  $\times$  Constraint:  $\chi^2(1) = 10.34$ ,  $p = 0.001$ , OR = 1.29). Importantly, all these effects were qualified by a significant three-way interaction of Group  $\times$  Constraint  $\times$  Task ( $\chi^2(1) = 11.43$ ,  $p < 0.001$ , OR = 0.88). Pairwise comparisons revealed that production of the most frequent responses according to the norms was less common among autistic participants than non-autistic participants in the Language task for both high and low constraints (High: Autistic M (SD) = 42.55% (14.20%) vs. Non-autistic M (SD) = 58.70% (9.81%),  $z = 5.42$ ,  $p < 0.001$ ; Low: Autistic M (SD) = 20.39% (10.58%) vs. Non-autistic M (SD) = 33.55% (10.38%),  $z = 5.42$ ,  $p < 0.001$ ). Conversely, production of the most frequent responses was more common among autistic participants for the music task under high constraint (Autistic M (SD) = 23.54% (20.93%) vs. Non-autistic M (SD) = 11.99% (14.37%),  $z = 3.41$ ,  $p < 0.001$ ), but there was no group difference under low constraint (Autistic M (SD) = 11.59% (7.39%) vs. Non-autistic M (SD) = 11.72% (6.15%),  $z = 0.28$ ,  $p = 0.778$ ).

In the full model, that is, with the addition of other predictors involving participants' background measures and stimulus properties, there were significant effects of age ( $\chi^2(1) = 19.91$ ,  $p < 0.001$ , OR = 1.10), receptive vocabulary ( $\chi^2(1) = 13.85$ ,  $p < 0.001$ , OR = 1.01), and pitch matching deviation ( $\chi^2(1) = 5.99$ ,  $p = 0.014$ , OR = 1.0), suggesting that older participants ( $B = 0.10$ ,  $SE = 0.02$ ), participants with higher receptive vocabulary ( $B = 0.01$ ,  $SE = 0.00$ ), and participants who were better



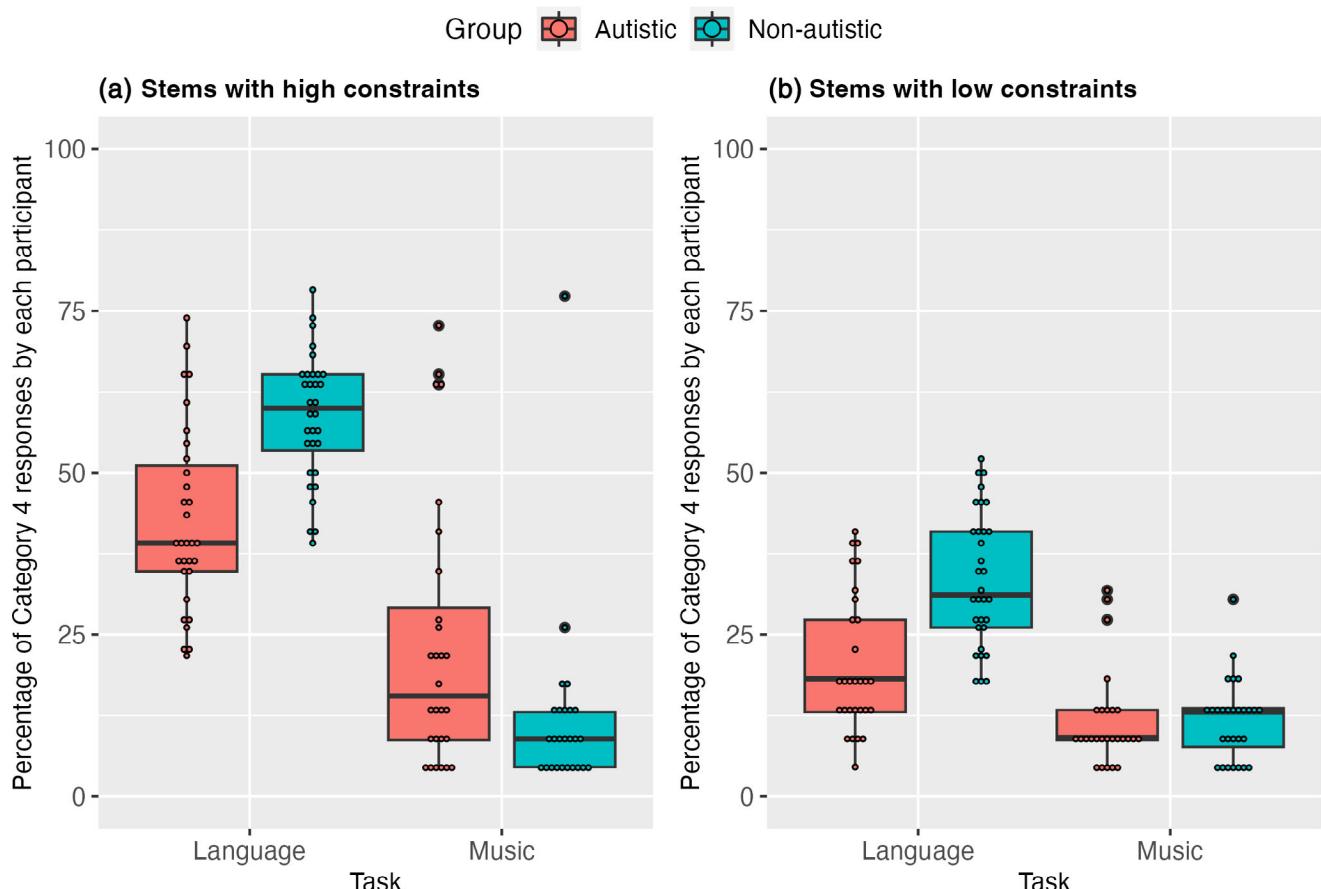
**FIGURE 3** Four production categories among Mandarin speakers in Experiment 1.

able to pitch match ( $B = -0.001$ ,  $SE = 0.001$ ) were more likely to produce the most frequent responses. There was also a significant effect of perceptual rating ( $\chi^2(1) = 141.69$ ,  $p < 0.001$ ,  $OR = 1.35$ ), suggesting that participants' perception and production performance was correlated ( $B = 0.30$ ,  $SE = 0.02$ ). Crucially, even after those factors were accounted for, similar findings as the simple model were found: there was a significant three-way interaction involving Group  $\times$  Task  $\times$  Constraint ( $\chi^2(1) = 11.13$ ,  $p < 0.001$ ,  $OR = 0.88$ ). Similar to the simple model, group differences were found under high and low

constraints in the language task, and only in the high constraint music task but not in the low constraint music task (Language High: Autistic M (SD) = 42.55% (14.20%) vs. Non-autistic M (SD) = 58.70% (9.81%),  $z = 4.87$ ,  $p < 0.001$ ; Language Low: Autistic M (SD) = 20.39% (10.58%) vs. Non-autistic M (SD) = 33.55% (10.38%),  $z = 4.36$ ,  $p < 0.001$ ; Music High: Autistic M (SD) = 23.54% (20.93%) vs. Non-autistic M (SD) = 11.99% (14.37%),  $z = 4.41$ ,  $p < 0.001$ ; Music Low: Autistic M (SD) = 11.59% (7.39%) vs. Non-autistic M (SD) = 11.72% (6.15%),  $z = 0.70$ ,  $p = 0.485$ ).

**TABLE 3** Counts of responses from the four categories and chi-squared tests comparing the distribution of the counts in percentage by groups in Experiment 1 on Mandarin speakers.

| Task     | Constraint | Group        | Category 1 | Category 2 | Category 3 | Category 4 | Chi-squared test    |
|----------|------------|--------------|------------|------------|------------|------------|---------------------|
| Language | High       | Autistic     | 9          | 15         | 33         | 43         | $\chi^2(3) = 8.10$  |
| Language | High       | Non-autistic | 2          | 10         | 30         | 59         | $p = 0.044$         |
| Language | Low        | Autistic     | 16         | 16         | 50         | 19         | $\chi^2(3) = 11.51$ |
| Language | Low        | Non-autistic | 5          | 9          | 52         | 33         | $p = 0.009$         |
| Music    | High       | Autistic     | 3          | 42         | 34         | 21         | $\chi^2(3) = 6.18$  |
| Music    | High       | Non-autistic | 1          | 52         | 38         | 10         | $p = 0.103$         |
| Music    | Low        | Autistic     | 4          | 29         | 57         | 10         | $\chi^2(3) = 3.73$  |
| Music    | Low        | Non-autistic | 1          | 39         | 50         | 10         | $p = 0.292$         |



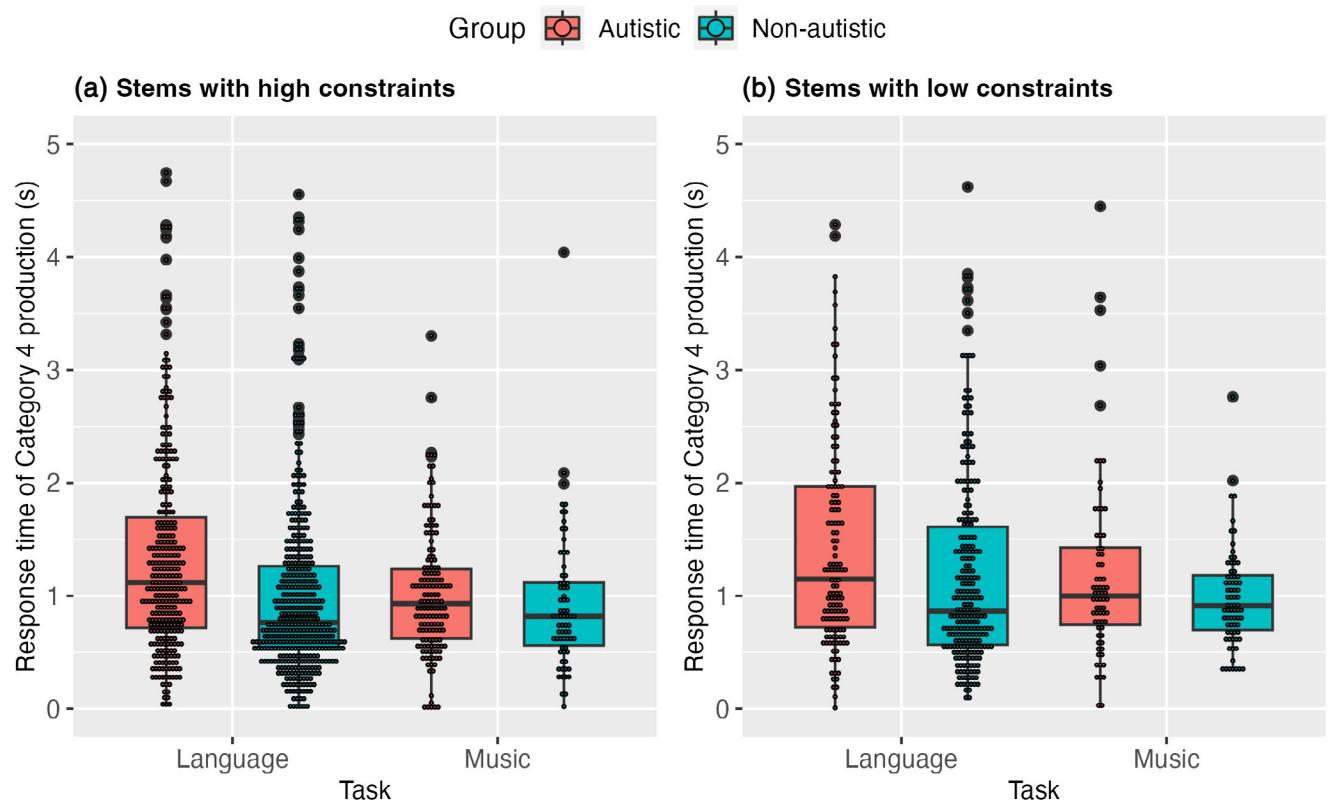
**FIGURE 4** Category 4 production among Mandarin speakers in Experiment 1.

### Reaction time

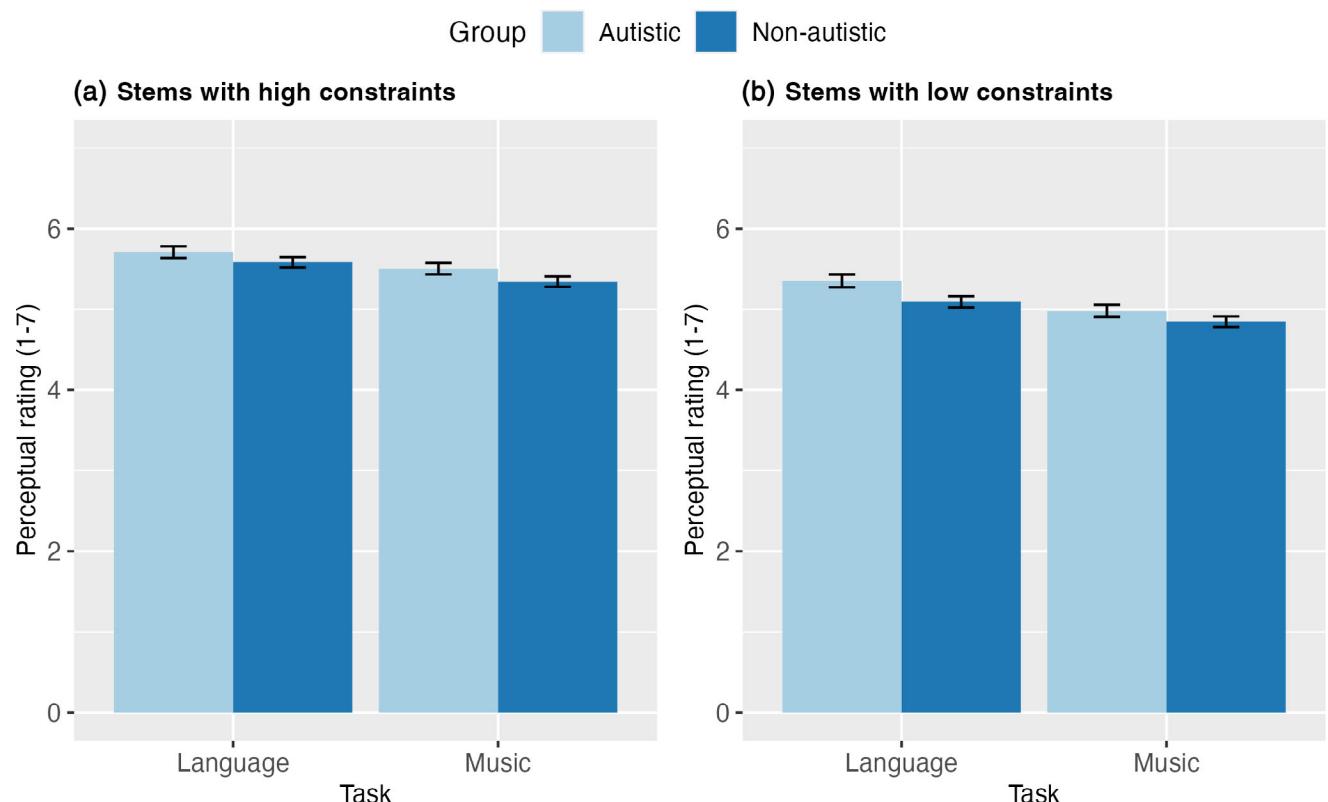
Regarding reaction times (see Figure 5), the simple model revealed a significant effect of Constraint ( $\chi^2(1) = 6.78$ ,  $p = 0.009$ ,  $R^2 = 0.011$ ), with reaction time being faster for high constraint items than for low constraints items (High M (SD) = 1.11 (0.80) vs. Low M (SD) = 1.21 (0.83),  $t(209) = 2.60$ ,  $p = 0.010$ ). There was also a significant effect of Task ( $\chi^2(1) = 5.53$ ,  $p = 0.019$ ,  $R^2 = 0.018$ ), which was qualified by a Group  $\times$  Task interaction ( $\chi^2(1) = 5.39$ ,  $p = 0.020$ ,  $R^2 = 0.014$ ). Pairwise comparisons revealed that group differences were only evident in

the language task, in which autistic participants had longer reaction times than non-autistic participants (Autistic M (SD) = 1.37 (0.92) vs. Non-autistic M (SD) = 1.07 (0.81),  $t(65.9) = 3.26$ ,  $p = 0.002$ ), whereas no group differences were found in the music task (Autistic M (SD) = 1.07 (0.65) vs. Non-autistic M (SD) = 0.95 (0.52),  $t(57.5) = 0.63$ ,  $p = 0.530$ ).

In the full model, the same findings as the simple model were found: there were significant effects of Constraint ( $\chi^2(1) = 6.80$ ,  $p = 0.009$ ,  $R^2 = 0.010$ ), with reaction time being faster for high constraint items than for



**FIGURE 5** Response times of Category 4 production among Mandarin speakers in Experiment 1.



**FIGURE 6** Perceptual ratings with means and standard errors across all individual items among Mandarin speakers in Experiment 1.

low constraints items (High M (SD) = 1.11 (0.80) vs. Low M (SD) = 1.21 (0.83),  $t(212) = 2.60, p = 0.010$ ), Task ( $\chi^2(1) = 5.49, p = 0.019, R^2 = 0.017$ ), which was qualified by a Group  $\times$  Task interaction ( $\chi^2(1) = 4.37, p = 0.037, R^2 = 0.011$ ). Pairwise comparisons revealed that group differences were only evident in the language task, in which autistic participants had longer reaction times than non-autistic participants (Autistic M (SD) = 1.37 (0.92) vs. Non-autistic M (SD) = 1.07 (0.81),  $t(58.2) = 2.40, p = 0.002$ ), whereas no group differences were found in the music task (Autistic M (SD) = 1.07 (0.65) vs. Non-autistic M (SD) = 0.95 (0.52),  $t(58.6) = 0.61, p = 0.542$ ). No other predictors were significant in the model.

### Perceptual rating tasks of the completed melodies and sentences

Figure 6 shows participants' ratings of the completed melodies and sentences with the most frequent responses based on the norms separated by items with high versus low constraints. The simple model revealed significant effects of Task ( $\chi^2(1) = 5.56, p = 0.018, R^2 = 0.007$ ), with music stimuli receiving lower ratings than language stimuli (Music M (SD) = 5.16 (1.81) vs. Language M (SD) = 5.43 (1.92),  $z = 2.36, p = 0.018$ ), and of Constraint ( $\chi^2(1) = 32.55, p < 0.001, R^2 = 0.018$ ), as stimuli with high constraints received higher ratings than those with low constraints (High constraint M (SD) = 5.53 (1.79) vs. Low constraint M (SD) = 5.07 (1.92),  $z = 5.71, p < 0.001$ ). No other effects or interactions were significant.

In the full model, there was a significant effect of Constraint ( $\chi^2(1) = 20.95, p < 0.001, R^2 = 0.010$ ), which was qualified by a Task  $\times$  Constraint interaction ( $\chi^2(1) = 4.98, p = 0.026, R^2 = 0.001$ ). Subsequent pairwise comparisons revealed that the difference between High constraint vs. Low constraint was greater in the Music task than in the Language task (Music High: M (SD) = 5.42 (1.76) vs. Music Low: M (SD) = 4.91 (1.83),  $z = 5.03, p < 0.001$ ; Language High: M (SD) = 5.65 (1.82) vs. Language Low: M (SD) = 5.22 (1.99),  $z = 2.46, p = 0.014$ ). Additionally, there was a significant effect of production category ( $\chi^2(1) = 207.18, p < 0.001, R^2 = 0.037$ ), suggesting that participants' perception and production performance was correlated (Frequent Response: M (SD) = 5.96 (1.62) vs. Infrequent Response: M (SD) = 5.07 (1.90),  $z = 14.39, p < 0.001$ ). No other predictors were significant in the model.

### Discussion

Results from Experiment 1 on Mandarin speakers suggest that across both autistic and non-autistic groups, participants' predictive production of the final notes/

words in a context and their perceptual ratings of the most frequent notes/words based on the norms mirrored each other across both domains. In the cloze production tasks, while the two groups showed similar distributions of responses classified into four categories for the music task, the non-autistic group produced more responses which were the most frequent in the norms than did the autistic group for the language task. When the production responses were examined based on two categories (the most frequent responses based on the norms vs. no/other responses), the autistic group produced more notes that were the most frequent in the norms for the music task compared to the non-autistic group for stems with high constraints only, but not for stems with low constraints. Regardless of the constraint condition of the stems, the non-autistic group produced more words that were the most frequent in the norms for the language task compared to the autistic group. While the two groups showed similar response times in the music task, the autistic group responded to sentences more slowly than the non-autistic group in the language task. In terms of expectedness ratings of the melodies and sentences that were completed with the most frequent responses in the norms, both autistic and non-autistic groups provided higher expectedness ratings for sentences than melodies, and for sentences and melodies with high constraints than those with low constraints. Finally, age, receptive vocabulary, and pitch matching ability were significant factors predicting the production of most frequent responses across both tasks. Overall, these findings suggest that autistic Mandarin speakers showed intact predictive processing of music but atypical prediction of language. However, given that the two groups differed somewhat in their background measures (see Table 1), which might partly explain the group differences observed, we repeated the experiment on a sample of English-speaking participants who were more closely matched in Experiment 2.

## EXPERIMENT 2

### Introduction

While the findings of Experiment 1 suggest a dissociation in music and language predictive processing abilities among autistic individuals, there are several caveats that need to be considered. Firstly, the findings might be influenced by potential confounding variables of the sample, such as musical training (no between-group difference but years of training in favor of the autistic group) and receptive vocabulary (significant between-group difference in favor of the non-autistic group). When comparing the simple model with the full model on production response, although the significant main effects and interactions remained in the full model after accounting for participant and stimulus characteristics, we still cannot

rule out the possibility that the imbalanced musical training experience and receptive vocabulary ability across the two groups might have contributed to the interaction effects of group, task, and constraint in the models. Previous studies have shown a link between receptive language skills and predictive language processing in young autistic children (Prescott et al., 2022). It has also been proposed that atypical predictive processing may cause language difficulties in autism (Ellis Weismer & Saffran, 2022). Therefore, it needs to be acknowledged that the current findings may not generalize to the entire autism spectrum which manifests a range of cognitive abilities across different autistic individuals (Lenroot & Yeung, 2013; Tager-Flusberg & Kasari, 2013).

Secondly, it is worth noting that the norms in the melodic cloze task were collected from musically-trained Western adults ( $9 \pm 4.8$  years) (Fogel et al., 2015), whereas we focused on Chinese children with little musical training in Experiment 1 (autistic:  $0.81 \pm 1.23$  years; non-autistic:  $0.47 \pm 1.08$  years). As a result, our participants showed reduced pitch matching abilities and produced fewer numbers of the most frequent notes across the different melodic stems as compared with the norms established by the adult musicians in Fogel et al. (2015). While Mandarin-speaking children are exposed to Western music due to globalization, they nonetheless likely differ from Western musician adults in their degree of implicit knowledge of Western tonal music (Yang & Welch, 2023). Thus, the current findings may not generalize to the entire autism spectrum which manifests a diversity of musical abilities (Ong et al., 2023; Sota et al., 2018).

Thirdly, in the current Mandarin sample, participants aged between 7 and 16, which might have introduced potential confounding factors related to developmental differences within this range (Eccles, 1999). The wide age span could impact the participants' cognitive abilities, language skills, and musical training (Nippold, 2000; Paus, 2005; Tierney et al., 2015), which in turn could influence their predictive processing abilities. Thus, it is important to acknowledge that developmental differences within the age range could influence the current findings.

To address these limitations, we conducted the same experiment on a well-matched sample of English speakers in Experiment 2 to further elucidate the relationship between predictive processing, music, and language in autism.

## Method

### Participants

A total of 79 British English-speaking individuals, aged between 12 and 57, participated in Experiment 2. Participants in the autistic group ( $n = 28$ ) received a clinical diagnosis of autism by a licensed clinician whereas

non-autistic participants ( $n = 51$ ) reported no psychiatric or neurological disorders. All participants completed the Raven's Standard Progressive Matrices (Raven et al., 1998) test as a measure of nonverbal IQ, the Receptive One-Word Picture Vocabulary Test (ROWPVT-4) (Martin & Brownell, 2011) as a measure of receptive vocabulary, and the digit span forward task as a measure of short-term memory (Piper et al., 2016). Participants' musical training background was collected through a questionnaire, where their years of musical training were summed across all instruments including voice (Pfordresher & Halpern, 2013). When comparing all participants between the two groups, no significant differences were observed in age, musical training experience, or cognitive abilities (see Table 4). However, there were more female participants in the non-autistic group ( $p = 0.055$ ), and autistic participants had shorter digit spans than non-autistic participants ( $p = 0.089$ ). To further match the two groups, a subset of 28 non-autistic participants were chosen to individually match with the 28 autistic participants on age, sex, digit span, musical training, nonverbal IQ, and receptive vocabulary. As can be seen from Table 4, the two groups of participants ( $n = 28$  each) were closely matched on all background measures (all  $ps > 0.24$ ).

### Cloze production and pitch matching tasks

#### Stimuli

Stimuli in the melodic cloze task were the same as those in Experiment 1. Ninety sentence stems were chosen to match the cloze probabilities (within  $\pm 3\%$  difference) of the 90 melodic stems based on the norms established from previous studies in English (Arcuri et al., 2001; Block & Baldwin, 2010), as well as on number of notes/syllables (within  $\pm 2$  difference). We replaced some names/pronouns in the original sentences with names of different syllables in order to match the number of syllables/notes across the melodies and sentences. For example, in the stem "For a runner Ted is rather," "Ted" was replaced with "Connor" to make the stem nine syllables long rather than eight syllables. However, we did not make changes to some of the names/pronouns, because (1) some sentences sound more natural with pronouns rather than names, (2) we tried to match the number of syllables between the sentences rather than across the melodies and sentences, and (3) we tried to use the same word categories (names or pronouns) for the high-constraint versus low-constraint sentence pairs. The final set of sentence stems and the whole sentences were then recorded by a native female speaker of British English using Praat (Boersma & Weenink, 2001), with 44.1 kHz sampling rate and 16-bit amplitude resolution.

Using the same randomization method as in Experiment 1, the 90 matched melodic stems and sentence stems

**TABLE 4** Demographic characteristics of the autistic and non-autistic groups in Experiment 2 on English speakers.

| Measures             | Group 1: Autistic participants | Group 2: All non-autistic participants | Group 3: A subset of non-autistic participants | <i>t</i> or chi-squared test between groups 1–2 | <i>t</i> or chi-squared test between groups 1–3 |
|----------------------|--------------------------------|--|--|---|---|
| <i>n</i>             | 28                             | 51                                     | 28   | NA  | NA  |
| Age                  | 28.79 (15.27)                  | 24.57 (9.43)                           | 27.04 (11.81)                                  | <i>t</i> (77) = 1.50, <i>p</i> = 0.139          | <i>t</i> (54) = 0.47, <i>p</i> = 0.639          |
| Age range            | 12–57                          | 13–55                                  | 13–55  | NA  | NA  |
| Sex                  | F = 14, M = 12, NB = 2         | F = 36, M = 15                         | F = 14, M = 14                                 | $\chi^2(2) = 5.81, p = 0.055$                   | $\chi^2(2) = 2.15, p = 0.341$                   |
| Musical training     | 4.38 (5.64)                    | 5.64 (6.01)                            | 5.29 (6.79)                                    | <i>t</i> (76) = -0.90, <i>p</i> = 0.371         | <i>t</i> (54) = -0.54, <i>p</i> = 0.594         |
| NVIQ                 | 51.79 (29.32)                  | 43.04 (24.54)                          | 48.93 (27.01)                                  | <i>t</i> (77) = 1.39, <i>p</i> = 0.167          | <i>t</i> (54) = 0.37, <i>p</i> = 0.711          |
| Receptive vocabulary | 108.93 (16.07)                 | 107.86 (13.27)                         | 109.00 (13.49)                                 | <i>t</i> (77) = 0.31, <i>p</i> = 0.756          | <i>t</i> (54) = -0.02, <i>p</i> = 0.986         |
| Digit span           | 6.63 (1.52)                    | 7.22 (1.35)                            | 7.11 (1.45)                                    | <i>t</i> (76) = -1.72, <i>p</i> = 0.089         | <i>t</i> (53) = -1.17, <i>p</i> = 0.247         |

Note: F = female, M = male, NB = non-binary. Musical training is in years. NVIQ: percentiles of Raven's standard progressive matrices; Receptive vocabulary: standard score of the receptive one word picture vocabulary test, 4th edition (ROWPVT-4). There was one participant's data missing on musical training and digit span.

were grouped into 8 lists of 45 items and presented in counterbalanced order across participants.

#### Procedure

The experiment was carried out in a soundproof room at the University of Reading, following the same procedure for the cloze production and pitch matching tasks as in Experiment 1.

#### Perceptual rating tasks

##### Stimuli

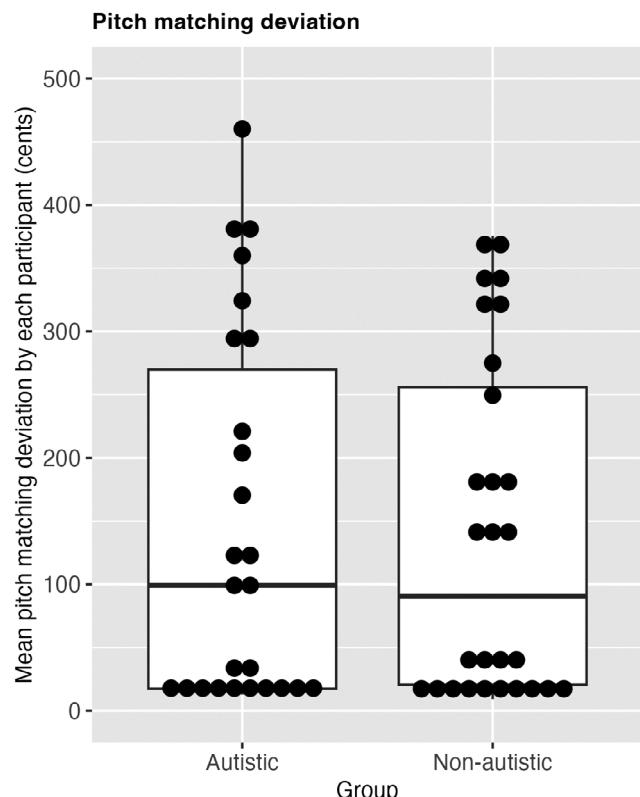
In the rating tasks, the stimuli included the melodies and sentences completed/produced in full, including the last notes/words that had the highest cloze probabilities based on the norms. The norms for the notes were from Fogel et al. (2015), and the norms for the words were from previous studies (Arcuri et al., 2001; Block & Baldwin, 2010) as described above.

##### Procedure

The same procedure was used as in the rating tasks in Experiment 1.

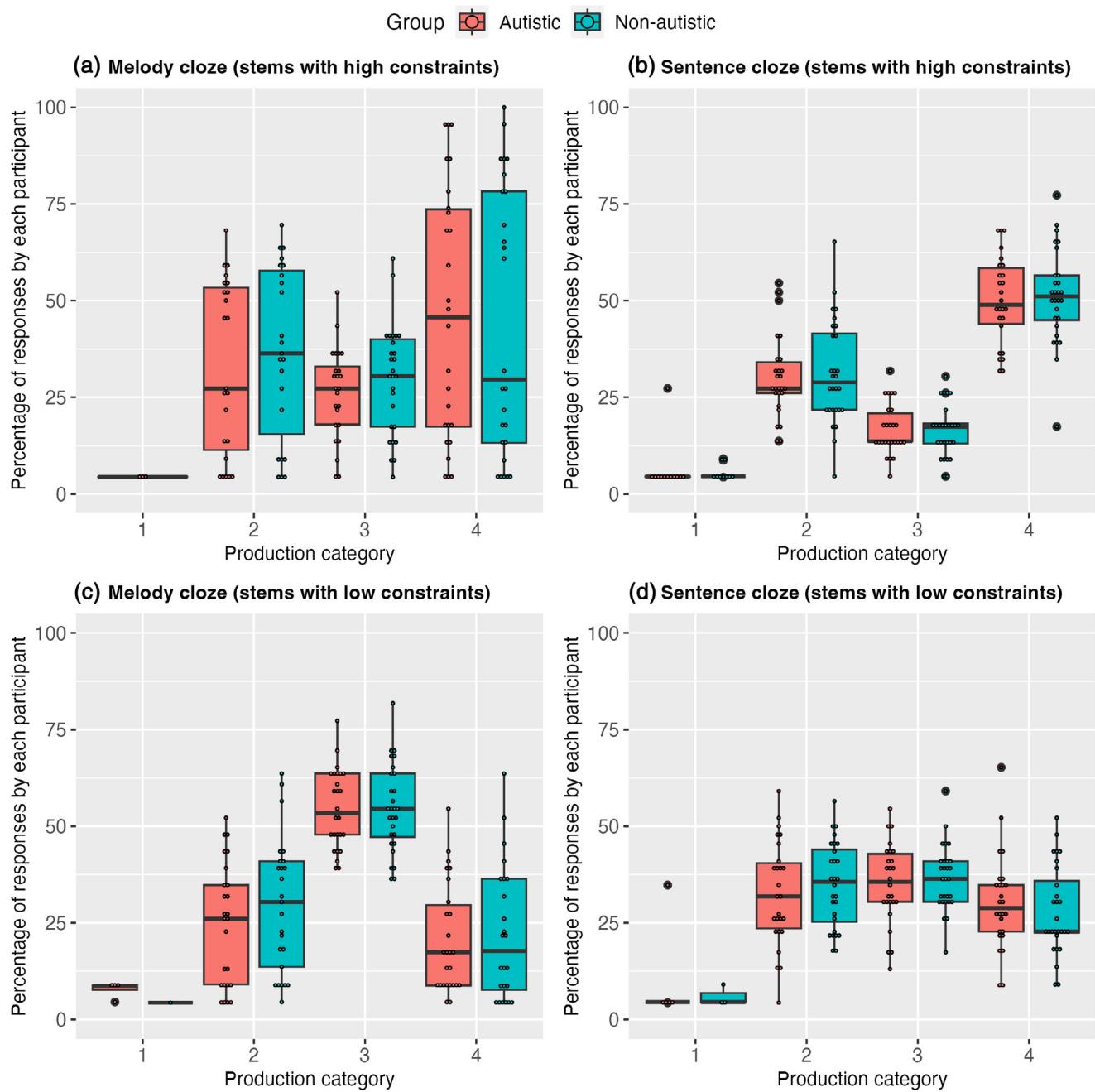
#### Data analysis

Analysis of the pitch matching and melodic cloze production data followed the same procedure as in Experiment 1. The transcription of the sentence cloze responses was done independently and then cross-checked by two research assistants. Using custom-written Matlab and R scripts, participants' sung notes and spoken words were categorized into four categories based on the norms from Fogel et al. (2015), Arcuri et al. (2001), and Block and Baldwin (2010): 1. no response; 2. a note/word that



**FIGURE 7** Mean pitch matching deviations among English speakers in Experiment 2.

has not been reported in the norms; 3. a less frequent note/word from the norms; 4. the most frequent note/word from the norms. Statistical analysis was performed using the same approach as in Experiment 1. In the interest of space, only statistically significant effects/interactions are reported in the Results section below, with the entire model outputs displayed in Supplementary Tables S2.1–S2.6.



**FIGURE 8** Four production categories among matched English speakers ( $n = 28$  in each group) in Experiment 2.

## Results

### Pitch matching

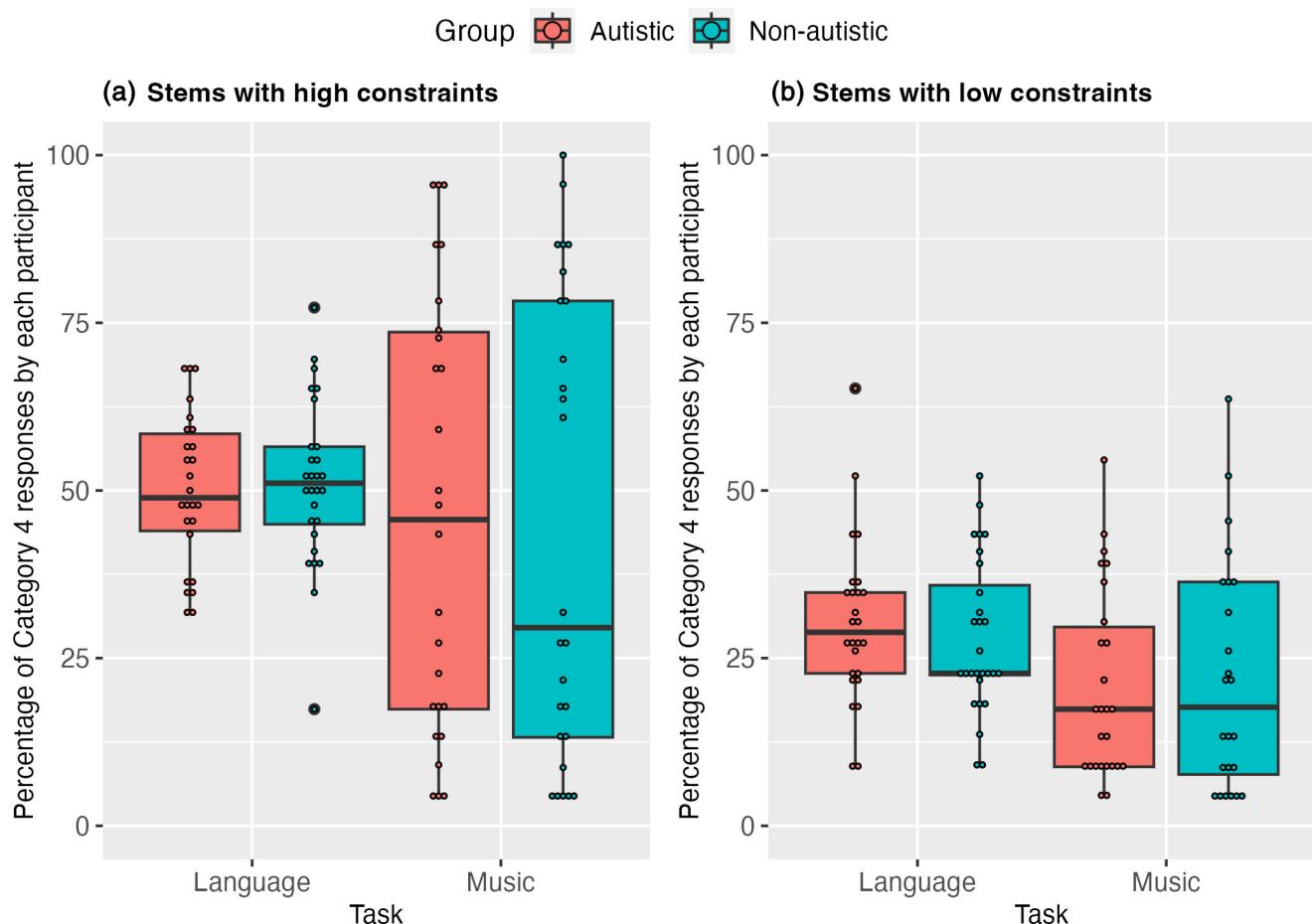
Figure 7 shows the boxplots of pitch matching deviations by each participant in each group. There was no significant difference in pitch matching deviation between the two groups ( $t(52) = 0.18$ ,  $p = 0.860$ ; autistic mean (SD) = 144.82 (144.09); non-autistic mean (SD) = 138.08 (131.08)).

### Melodic and sentence cloze production tasks

Figure 8 shows the boxplots of percentages of the responses from the four categories by each participant in each group separated by the task (music vs. language) and constraint (high vs. low) conditions, and Table 5 displays the responses counts in percentage from the four categories and chi-squared tests comparing the distribution of the counts by groups. The two groups did not differ significantly in their response categories in any of the conditions.

**TABLE 5** Counts of responses from the four categories and chi-squared tests comparing the distribution of the counts in percentage between groups in Experiment 2 on English speakers.

| Task     | Constraint | Group        | Category 1 | Category 2 | Category 3 | Category 4 | Chi-squared test   |
|----------|------------|--------------|------------|------------|------------|------------|--------------------|
| Language | High       | Autistic     | 3          | 30         | 17         | 50         | $\chi^2(3) = 0.26$ |
| Language | High       | Non-autistic | 2          | 31         | 16         | 51         | $p = 0.968$        |
| Language | Low        | Autistic     | 2          | 33         | 35         | 30         | $\chi^2(3) = 0.48$ |
| Language | Low        | Non-autistic | 1          | 35         | 36         | 28         | $p = 0.924$        |
| Music    | High       | Autistic     | 1          | 29         | 24         | 46         | $\chi^2(3) = 1.61$ |
| Music    | High       | Non-autistic | 0          | 30         | 28         | 41         | $p = 0.658$        |
| Music    | Low        | Autistic     | 1          | 24         | 54         | 21         | $\chi^2(3) = 1.19$ |
| Music    | Low        | Non-autistic | 0          | 26         | 55         | 19         | $p = 0.756$        |



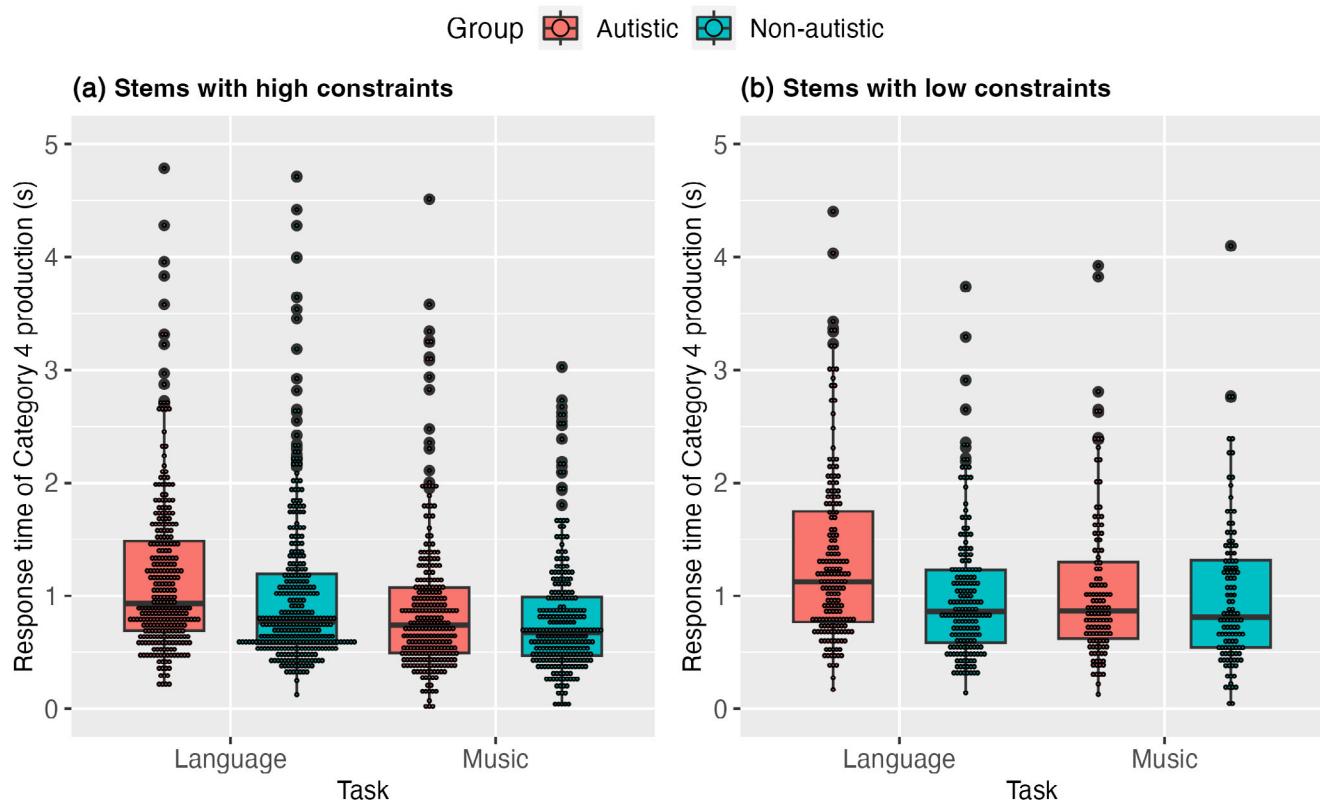
**FIGURE 9** Category 4 production among English speakers in Experiment 2.

#### Production of the most frequent responses

In the simple model on participants' binomial production responses (1: the most frequent responses based on the norms; 0: no or other responses; see Figure 9), there were significant effects of Task ( $\chi^2(1) = 10.15$ ,  $p = 0.001$ , OR = 1.50) and Constraint ( $\chi^2(1) = 54.47$ ,  $p < 0.001$ , OR = 2.04), both of which were qualified by a Task  $\times$  Constraint interaction ( $\chi^2(1) = 22.60$ ,  $p < 0.001$ , OR = 0.84). Pairwise comparisons revealed that the difference between High constraint vs. Low constraint was

greater in the Music task than in the Language task (Music High: M (SD) = 45.40% (33.43%) vs. Music Low: M (SD) = 21.30% (15.57%),  $z = 8.33$ ,  $p < 0.001$ ; Language High: M (SD) = 50.66% (11.69%) vs. Language Low: M (SD) = 29.09% (11.78%),  $z = 5.33$ ,  $p < 0.001$ ). There were no significant effects or interactions involving Group.

In the full model, there were significant effects of age ( $\chi^2(1) = 6.26$ ,  $p = 0.012$ , OR = 1.02) and pitch matching



**FIGURE 10** Response times of Category 4 responses among English speakers in Experiment 2.

deviation ( $\chi^2(1) = 8.57, p = 0.003, \text{OR} = 1.0$ ), suggesting that older participants ( $B = 0.02, \text{SE} = 0.01$ ) and participants who were better able to pitch match ( $B = -0.002, \text{SE} = 0.001$ ) were more likely to produce the most frequent responses. There was also a significant effect of perceptual rating ( $\chi^2(1) = 74.73, p < 0.001, \text{OR} = 1.36$ ), suggesting that participants' perception and production performance was correlated ( $B = 0.30, \text{SE} = 0.04$ ). Unlike in the simple model, there was only a significant effect of Constraint ( $\chi^2(1) = 38.68, p < 0.001, \text{OR} = 1.69$ ), with there being more frequent responses for the High constraint items than for the Low constraint items (High:  $M (\text{SD}) = 48.08\% (24.87\%)$  vs. Low:  $M (\text{SD}) = 25.34\% (14.21\%)$ ,  $z = 6.22, p < 0.001$ ). Like the simple model, there were no significant effects or interactions involving Group.

#### Reaction time

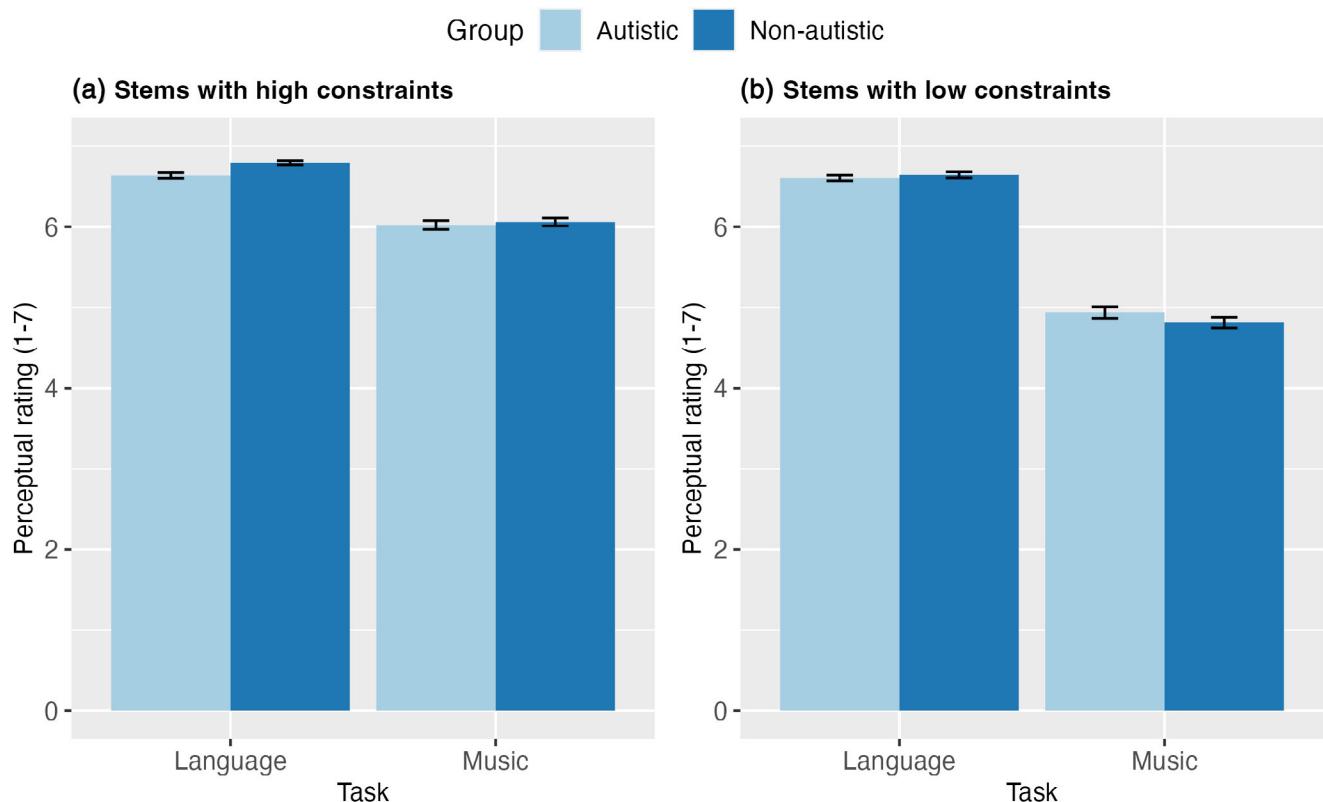
The simple model for reaction times (see Figure 10) revealed significant effects of Task ( $\chi^2(1) = 7.60, p = 0.006, R^2 = 0.033$ ), Constraint ( $\chi^2(1) = 8.31, p = 0.004, R^2 = 0.022$ ), and a significant Task  $\times$  Constraint interaction ( $\chi^2(1) = 7.78, p = 0.005, R^2 = 0.006$ ). Pairwise comparisons showed that whereas there was no significant difference in reaction times between High constraint items and Low constraint items in the Language task (High:  $M (\text{SD}) = 1.09 (0.71)$  vs. Low:  $M (\text{SD}) = 1.17 (0.70)$ ,  $t(123) = 1.30, p = 0.196$ ), reaction times

were faster in the High constraint items than in the Low constraint items in the Music task (High:  $M (\text{SD}) = 0.85 (0.59)$  vs. Low:  $M (\text{SD}) = 1.03 (0.66)$ ,  $t(151) = 3.77, p = 0.002$ ). There were no significant effects or interactions involving Group.

In the full model, similar to the simple model, there were significant effects of Task ( $\chi^2(1) = 8.59, p = 0.003, R^2 = 0.030$ ), Constraint ( $\chi^2(1) = 7.64, p = 0.006, R^2 = 0.018$ ), and a significant Task  $\times$  Constraint interaction ( $\chi^2(1) = 4.78, p = 0.029, R^2 = 0.004$ ), such that differences in constraint were only observed in the Music task but not in the Language task (Music High:  $M (\text{SD}) = 0.85 (0.59)$  vs. Music Low:  $M (\text{SD}) = 1.03 (0.66)$ ,  $t(194) = 3.33, p = 0.001$ ; Language High:  $M (\text{SD}) = 1.09 (0.71)$  vs. Language Low:  $M (\text{SD}) = 1.17 (0.70)$ ,  $t(125) = 1.38, p = 0.169$ ). There was also a significant effect of sound duration ( $\chi^2(1) = 41.89, p < 0.001, R^2 = 0.043$ ), such that participants tended to be faster for longer stimuli ( $B = -0.09, \text{SE} = 0.01$ ). Like the simple model, there were no significant effects or interactions involving Group.

#### Perceptual rating tasks of the completed melodies and sentences

Figure 11 shows participants' ratings of the completed melodies and sentences with the most frequent responses



**FIGURE 11** Perceptual ratings with means and standard errors across all individual items among English speakers in Experiment 2.

based on the norms separated by items with high versus low constraints. The simple model revealed significant effects of Task ( $\chi^2(1) = 126.46, p < 0.001, R^2 = 0.232$ ), Constraint ( $\chi^2(1) = 64.54, p < 0.001, R^2 = 0.074$ ), and a significant Task  $\times$  Constraint interaction ( $\chi^2(1) = 339.85, p < 0.001, R^2 = 0.056$ ). Pairwise comparisons showed that whereas there was no significant difference in ratings between High constraint items and Low constraint items in the Language task (High: M (SD) = 6.71 (0.76) vs. Low: M (SD) = 6.62 (0.89),  $z = 1.07, p = 0.286$ ), ratings were higher in the High constraint items than in the Low constraint items in the Music task (High: M (SD) = 6.04 (1.31) vs. Low: M (SD) = 4.88 (1.74),  $z = 13.98, p < 0.001$ ). There were no significant effects or interactions involving Group.

In the full model, similar to the simple model, there were significant effects of Task ( $\chi^2(1) = 120.25, p < 0.001, R^2 = 0.233$ ), Constraint ( $\chi^2(1) = 54.47, p < 0.001, R^2 = 0.059$ ), and a significant Task  $\times$  Constraint interaction ( $\chi^2(1) = 326.20, p < 0.001, R^2 = 0.057$ ), such that differences in ratings between High constraint items and Low constraint items were only significant in the Music task but not in the Language Task (Music High: M (SD) = 6.04 (1.31) vs. Music Low: M (SD) = 4.88 (1.74),  $z = 13.42, p < 0.001$ ; Language High: M (SD) = 6.71 (0.76) vs. Language Low: M (SD) = 6.62 (0.89),  $z = 0.33, p = 0.739$ ). There were also significant effects of pitch

matching deviation ( $\chi^2(1) = 5.00, p = 0.025, R^2 = 0.009$ ) and sound duration ( $\chi^2(1) = 8.88, p = 0.003, R^2 = 0.003$ ), such that ratings were higher among those who could better pitch match ( $B = -0.0008, SE = 0.0004$ ) and for longer stimuli ( $B = -0.05, SE = 0.02$ ). Additionally, there was a significant effect of production category ( $\chi^2(1) = 81.39, p < 0.001, R^2 = 0.018$ ), suggesting that participants' perception and production performance was correlated (Frequent Response: M (SD) = 6.49 (1.09) vs. Infrequent Response: M (SD) = 5.82 (1.55),  $z = 9.02, p < 0.001$ ). Similar to the simple model, there were no significant effects or interactions involving group.

## Discussion

Table 6 shows a summary of the significant effects and interactions on the three measures in the simple and full models from both experiments. Across both experiments, participants' perception and production performance was significantly correlated, with age and pitch matching ability predicting the production of the most frequent responses across music and language tasks. However, the significant effect of receptive vocabulary on production response was only observed in Experiment 1. Furthermore, Experiment 2 revealed no significant effects or interactions involving Group across all measures. Thus,

**TABLE 6** Summary of the results from the simple and full models on production category, reaction time (RT), and perceptual ratings in both experiments.

|                           | Experiment 1 (Mandarin speakers) |      |        |      |        |      | Experiment 2 (English speakers) |      |        |      |        |      |
|---------------------------|----------------------------------|------|--------|------|--------|------|---------------------------------|------|--------|------|--------|------|
|                           | Production                       |      | RT     |      | Rating |      | Production                      |      | RT     |      | Rating |      |
|                           | Simple                           | Full | Simple | Full | Simple | Full | Simple                          | Full | Simple | Full | Simple | Full |
| Group                     |                                  |      |        |      |        |      |                                 |      |        |      |        |      |
| Task                      | ✓                                | ✓    | ✓      | ✓    | ✓      |      | ✓                               |      | ✓      | ✓    | ✓      | ✓    |
| Constraint                | ✓                                | ✓    | ✓      | ✓    | ✓      | ✓    | ✓                               | ✓    | ✓      | ✓    | ✓      | ✓    |
| Group × Task              | ✓                                | ✓    | ✓      | ✓    |        |      |                                 |      |        |      |        |      |
| Group × Constraint        | ✓                                |      | ✓      |      |        |      |                                 |      |        |      |        |      |
| Task × Constraint         | ✓                                | ✓    |        |      |        |      | ✓                               | ✓    | ✓      | ✓    | ✓      | ✓    |
| Group × Task × Constraint | ✓                                | ✓    |        |      |        |      |                                 |      |        |      |        |      |
| Age                       |                                  |      | ✓      |      |        |      |                                 |      | ✓      |      |        |      |
| Musical training          |                                  |      |        |      |        |      |                                 |      |        |      |        |      |
| Non-verbal IQ             |                                  |      |        |      |        |      |                                 |      |        |      |        |      |
| Receptive vocabulary      |                                  |      | ✓      |      |        |      |                                 |      |        |      |        |      |
| Digit span                |                                  |      |        |      |        |      |                                 |      |        |      |        |      |
| Pitch matching deviation  |                                  |      | ✓      |      |        |      |                                 |      | ✓      |      |        | ✓    |
| Sound duration            |                                  |      |        |      |        |      |                                 |      |        | ✓    |        | ✓    |
| Perceptual rating         |                                  |      | ✓      |      |        |      |                                 |      | ✓      |      |        |      |
| Production category       |                                  |      |        |      |        | ✓    |                                 |      |        |      |        | ✓    |

matching musical training experience and receptive vocabulary between the two groups in Experiment 2 eliminated the Group × Task × Constraint interaction on production response and the Group × Task interaction on reaction time as observed in Experiment 1. While pitch matching ability and sound duration were significant predictors of expectedness ratings of the music and language stimuli in Experiment 2, these effects were not observed in Experiment 1. We discuss the implications of the findings from both experiments below.

## GENERAL DISCUSSION

Examining two samples of autistic and non-autistic individuals with different language, music, and cognitive abilities, we compared predictive processing of language and music in autism using closely matched melodic and sentence cloze tasks as well as perceptual rating tasks. Participants in both experiments were generally sensitive to the different degrees of expectation based on the linguistic or melodic stem, as indexed by the significant effect of Constraint in all the models examined. This suggests that the prediction tasks used in this study are indeed measuring predictive processing (i.e., otherwise, there would not be a difference in performance between high constraint and low constraint stems). Importantly, the results from both experiments suggest that the initial group differences qualified by a Group × Task × Constraint interaction as observed in Experiment 1 were likely driven by variations in musical training experience

and receptive vocabulary skills in the Mandarin sample. That is, individual differences in musical and linguistic abilities among the participants in Experiment 1, rather than autism diagnosis, might have mediated the relationship between group (autistic vs. non-autistic) and predictive processing of language and music. In keeping with the extensive evidence indicating the significant roles cognitive abilities play in pitch and melodic processing in autism (Chowdhury et al., 2017; Jamey et al., 2019; Kargas et al., 2015; Ong et al., 2023), our findings suggest the importance of accounting for confounding factors in studies of music and language processing in autism.

Regarding musical prediction, we observed generally similar performance between autistic and non-autistic groups in both production (in number of most frequent notes produced and in reaction time) and perceptual rating tasks, in both experiments. It should be noted that whereas the Chinese autistic participants produced more frequent notes than the Chinese non-autistic participants for the high constraint melodic stems in Experiment 1, the same pattern of results was not seen among the English participants in Experiment 2. At first glance, it might seem like the Chinese autistic participants showed *enhanced* musical skills (at least with respect to producing the most frequent notes) whereas the English autistic participants showed *intact* musical skills relative to their respective non-autistic counterparts. However, we caution against such an interpretation due to how the task was scored. Specifically, performance on the music task in both experiments was scored against norms

established by Western musicians. That is, in Experiment 1 on Chinese speakers, although the sentence cloze task was re-normed in a Mandarin-speaking sample, we did not re-norm the melodic cloze task considering that performance on that task should be judged against expert performance. However, one may question whether performance on Western melodies among Chinese participants relative to the performance of Western musicians is itself a cognitively informative measure. Indeed, while Chinese non-musicians have been shown to be familiar with the tonality, syntactic structure, and phrase boundary of Western music (Jiang et al., 2016, 2017; Nan et al., 2009), we cannot rule out the possibility that the Chinese autistic participants in Experiment 1 may be more aware (or alternatively, the Chinese non-autistic participants being less aware) of the “rules” in Western tonal music than the Chinese non-autistic participants because of their musical training experience. Future studies should replicate the melodic cloze task with norms from a Mandarin-speaking sample to see if group differences among Chinese participants persist. What is clear, however, is that the autistic participants in both experiments showed intact musical processing skills at the very least. This is consistent with the literature suggesting intact musical processing in autism (Chen et al., 2022; Janzen & Thaut, 2018; O'Connor, 2012; Ouimet et al., 2012; Quintin, 2019; Ong et al., 2023), including memory and labeling of musical pitch and segmentation of chords (Heaton, 2003), perception of pitch intervals and melodic contours (Heaton, 2005; Jiang et al., 2015), local and global processing of music (Germain et al., 2019; Mottron et al., 2000), musical phrase boundary processing (DePriest et al., 2017), perception of musical melodies (Jamey et al., 2019), processing of temporal sequences and musical structures (Quintin et al., 2013), and recognition of musical emotions (Gebauer et al., 2014; Heaton et al. 1999; Quintin et al., 2011). Note, however, that there are counter-examples in which autistic individuals had lower performance than non-autistic individuals in musical tasks (e.g., Bhatara et al., 2010; Ong et al., 2023; Sota et al., 2018). Some autistic individuals also possess exceptional musical skills such as perfect pitch (Heaton, Pring, et al., 1999; Mottron et al., 2009; Rimland & Fein, 1988; Young & Nettelbeck, 1995). Thus, although impaired pitch, emotion, melodic (Bhatara et al., 2010; Ong et al., 2023; Sota et al., 2018), beat (Kasten et al., 2023; Morimoto et al., 2018; Vishne et al., 2021), rhythmic (Steinberg et al., 2021), and metrical processing (DePape et al., 2012) has also been observed in autism, our finding of intact musical prediction in autistic Mandarin and English speakers is in line with meta-analyses suggesting intact implicit learning in autism as evidenced in contextual cueing and other statistical learning tasks (Foti et al., 2015; Obeid et al., 2016). Most importantly, our findings are consistent with research suggesting intact enculturation to the pitch structure of Western music

(DePape et al., 2012) and utilizing prediction during music listening in autism (Venter et al., 2023).

A few computational models have tried to explain how melodic expectations are formed, including the domain-specific model of Gestalt-like principles derived from music theory (Probabilistic Model of Melody Perception; Temperley, 2008, 2014), the domain-general Markov model of statistical learning based on transition probabilities and *n*-grams (Information Dynamics of Music; IDyOM model; Hansen & Pearce, 2014; Pearce & Wiggins, 2006), and the expectation networks of statistical learning inspired by generalized scale degree associations (Verosky & Morgan, 2021). Using the data from the melodic cloze task (Fogel et al., 2015), Morgan et al. (2019) compared the Temperley and IDyOM models and suggested that both Gestalt principles and statistical learning of *n*-gram probabilities contribute to melodic expectations, with the latter playing a stronger role than the former (Morgan et al., 2019). However, neither model can fully explain the variance in the data, especially related to how the tonic is expected in authentic cadence (Morgan et al., 2019). Taking into account another type of statistical learning, expectation networks (Verosky, 2019), Verosky and Morgan (2021) compared all three models and concluded that expectation networks can better distinguish authentic cadence from non-cadential melodies and that the three models each explain distinct aspects of melodic expectations. Thus, melodic expectations are driven by a combination of Gestalt principles (Temperley, 2008, 2014), statistical learning of *n*-gram probabilities (Hansen & Pearce, 2014; Pearce & Wiggins, 2006), expectation networks (Verosky, 2019), as well as other non-local melodic relationships (Krumhansl, 1997; Kuhn & Dienes, 2005; Morgan et al., 2019; Verosky & Morgan, 2021). Our finding of intact musical prediction in autism thus suggests that, like non-autistic individuals, autistic individuals are able to make predictions of musical events following the principles of melodic expectations, which may be related to their intact implicit/statistical learning skills across a variety of statistical learning tasks (Foti et al., 2015; Obeid et al., 2016).

Regarding linguistic prediction, our results suggest a significant group difference in both response pattern and reaction time in the sentence cloze task among Mandarin speakers in Experiment 1, although no group difference was observed in the English-speaking sample in Experiment 2. Thus, it is important to acknowledge that it may not be autism diagnosis per se, but individual differences in linguistic abilities as reflected by receptive vocabulary, that have driven the reduced predictive processing of language among autistic participants in Experiment 1. In previous studies, autistic individuals also show difficulties in global sentence processing and semantic integration in context (Booth & Happé, 2010; Frith & Snowling, 1983; Happé, 1997; Henderson et al., 2011; Jolliffe & Baron-Cohen, 1999) and demonstrate delayed N400 latency

during semantic processing (DiStefano et al., 2019). Extensive research has also shown reduced reading comprehension in autism, especially in text comprehension at sentence and passage levels (Lucas & Norbury, 2014; McIntyre et al., 2017; Nation et al., 2006; Ricketts et al., 2013). The difficulties in oral language and reading comprehension in autism have been attributed to “weak central coherence” (Happé & Frith, 2006), decreased top-down modulation (Henderson et al., 2011) and monitoring (Koolen et al., 2013), social impairments (Ricketts et al., 2013), limited vocabulary or semantic knowledge (Brown et al., 2013; Lucas & Norbury, 2014), reduced processing speed of world knowledge (Howard et al., 2017), and impaired structural language ability (Eberhardt & Nadig, 2018). Comparing results from Experiments 1 and 2, it is likely that Mandarin-speaking autistic individuals’ atypical linguistic prediction as revealed in Experiment 1 may be mediated by their reduced linguistic skills, rather than be due to autism *per se*.

Similar to melodic expectations, linguistic expectations are also based on rule-like principles such as syntactic structure (Gibson, 1998) and statistical learning of word *n*-gram probabilities (Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Saffran, 2003). Language processing also involves regularity extraction and integration via statistical learning (Erickson & Thiessen, 2015). Generally speaking, prediction during language comprehension involves probabilistic computation of upcoming events in a context at multiple levels of representation, including perceptual, phonological, syntactic, semantic, and orthographic (Kuperberg & Jaeger, 2016). These predictions are made by drawing on information from the speaker, prior context, world knowledge, as well as extra-linguistic cues such as gestures or other visual stimuli (Hagoort & van Berkum, 2007). Studies and theoretical accounts examining autistic differences in perceptual adaptation stipulate increased sensory overload at subcortical levels (Font-Alamino et al., 2020), atypical top-down influence on sensory and higher-level information processing (Gomot & Wicker, 2012), attenuated use of prior higher-level social stimuli (Pellicano & Burr, 2012; Sinha et al., 2014; van Boxtel & Lu, 2013; Van de Cruys et al., 2014), and reduced speed of integrating new information to guide behavior—“slow-updating” (Lieder et al., 2019; Vishne et al., 2021). In our sentence production and rating tasks, participants need to process higher-level complex stimuli by a novel talker, which requires vigorous integration of bottom-up and top-down processes (Heald & Nusbaum, 2014; Kleinschmidt & Jaeger, 2015). It has been reported that autistic adults show delayed auditory feedback during speech production (Lin et al., 2015), and they also lack perceptual adaptation to novel talkers (Alispahic et al., 2022). Other studies have also shown atypical semantic processing (Ahtam et al., 2020; Grisoni et al., 2019; Kamio et al., 2007; O’Rourke & Coderre, 2021) and

reduced top-down modulation during online semantic processing in autism (Henderson et al., 2011). Nevertheless, our finding of intact linguistic prediction among autistic English speakers in Experiment 2 suggests that, if equipped with intact linguistic skills, autistic individuals are equally capable of performing predictive processing during language production and comprehension.

While we propose that the differences in findings between experiments may be attributed to how well the groups are matched within each experiment, there are two other possibilities that we should address. Across both experiments, participants also differed in their language experience and age (Mandarin speakers with a mean age between 10 and 11 in Experiment 1 and English speakers with a mean age between 27 and 28 in Experiment 2). Concerning language experience, given that the language task was conducted in their own native language, we think it is unlikely that this is the reason for the group differences observed among Mandarin speakers but not among English speakers in the language task. Concerning age, one might argue that perhaps autistic individuals’ predictive processing ability impacts the language domain more so than the music domain when they are younger. Even though predictive processing is said to be domain-general (e.g., Cannon et al., 2021; Sinha et al., 2014), it may operate on domain-specific representations, which may be weaker in the language domain among autistic individuals given that autism is associated with delayed language development (Hart & Curtin, 2023). Thus, group differences in the language task may be more likely to be observed among younger participants and disappear when they are older or when their language ability “catches up” with their non-autistic peers (Brignell et al., 2018). This should be confirmed in future research with younger English speakers. Moreover, participants (particularly autistic participants) with varying levels of linguistic ability should be examined in future research as the current study has focused on matching the autistic participants with their non-autistic peers, resulting in limited generalizability to the broader autistic population.

Apart from music and language processing, atypical predictive processing in autism has been frequently reported in other domains, including social functioning, visual processing, sensory processing, theory of mind, and motor anticipation (Cannon et al., 2021; Lawson et al., 2014; Pellicano & Burr, 2012; Sinha et al., 2014; Van de Cruys et al., 2014). However, combining the results from Experiments 1 and 2, we conclude that predictive processing of music and language in autism is likely influenced by individual differences in musical, linguistic, and cognitive abilities. Future studies should investigate the role of other factors (e.g., motor skills, comorbid conditions, and other potentially confounding variables) that may influence predictive processing in autism. Finally, given that our stimuli are relatively

simple, i.e., involving isolated melodies and sentences without larger contexts, future studies should explore how autistic individuals make predictions from more complex musical and linguistic inputs in their everyday environments.

In conclusion, our study is the first attempt to compare prediction in music and language in autistic and non-autistic individuals using matched cloze probability tasks across the two domains. Based on results from two experiments on two different samples, our findings suggested that performance on musical and linguistic prediction in autism is largely dependent on individual differences in musical, linguistic, and cognitive abilities. Future studies should employ different tasks (e.g., with more complex stimuli in naturally occurring settings) and different samples to further explore the complex relationship between autism, music, and language, while considering potential confounding variables impacting task performance.

## AUTHOR CONTRIBUTIONS

Fang Liu, Aniruddh D. Patel, Allison R. Fogel, Chen Zhao, Anamarija Veic, and Bhismadev Chakrabarti designed the study. Allison R. Fogel, Chen Zhao, Anamarija Veic, Li Wang, and Jia Hoong Ong created the stimuli. Chen Zhao, Li Wang, Qingqi Hou, and Anamarija Veic collected the data. Chen Zhao, Anamarija Veic, Dipsikha Das, Cara Crasto, and Fang Liu processed the data. Chen Zhao, Fang Liu, and Jia Hoong Ong analyzed the data. Fang Liu, Aniruddh D. Patel, Jia Hoong Ong, and Chen Zhao wrote the manuscript. Cunmei Jiang, Tim I. Williams, and Ariadne Loutrari commented on the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This study was supported by a European Research Council (ERC) Starting Grant, ERC-StG-2015, CAASD, 678733, to Fang Liu and Cunmei Jiang. We thank Sanrong Xiao for assistance with participant recruitment in China, Georgia Matthews for help with literature review of the sentence task, Ellie Packham, Roshni Vagadia, and Zivile Bernotaite for help with participant recruitment and data collection in the United Kingdom, Leyan Zheng for help with pre-processing of the Mandarin data, and Hiba Ahmed, Leah Jackson, and Maleeha Sujawal for help with pre-processing of the English data. We would also like to thank the associate editor and two anonymous reviewers for their insightful suggestions and comments which have helped us significantly improve our work.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Jia Hoong Ong  <https://orcid.org/0000-0003-1503-8311>  
 Cunmei Jiang  <https://orcid.org/0000-0002-0264-5924>  
 Li Wang  <https://orcid.org/0000-0001-5318-2408>  
 Bhismadev Chakrabarti  <https://orcid.org/0000-0002-6649-7895>  
 Fang Liu  <https://orcid.org/0000-0002-7776-0222>

## REFERENCES

Ahtam, B., Braeutigam, S., & Bailey, A. (2020). Semantic processing in autism spectrum disorders is associated with the timing of language acquisition: A magnetoencephalographic study. *Frontiers in Human Neuroscience*, 14, 267. <https://doi.org/10.3389/fnhum.2020.00267>

Alispahic, S., Pellicano, E., Cutler, A., & Antoniou, M. (2022). Auditory perceptual learning in autistic adults. *Autism Research*, 15(8), 1495–1507. <https://doi.org/10.1002/aur.2778>

American Psychiatric Association. (2013). *DSM-V: The American Journal of Psychiatry*. <https://doi.org/10.1176/appi.books.9780890425596.744053>

Amoruso, L., Narzisi, A., Pinzino, M., Finisguerra, A., Billeci, L., Calderoni, S., Fabbro, F., Muratori, F., Volzone, A., & Urgesi, C. (2019). Contextual priors do not modulate action prediction in children with autism. *Proceedings of the Royal Society B: Biological Sciences*, 286(1908), 20191319. <https://doi.org/10.1098/rspb.2019.1319>

Arcuri, S. M., Rabe-Hesketh, S., Morris, R. G., & McGuire, P. K. (2001). Regional variation of cloze probabilities for sentence contexts. *Behavior Research Methods, Instruments, & Computers*, 33(1), 80–90. <https://doi.org/10.3758/BF03195350>

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3), 349–371. <https://doi.org/10.1177/0023830913484891>

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82. <https://doi.org/10.1016/j.jml.2009.09.005>

Balsters, J. H., Apps, M. A. J., Bolis, D., Lehner, R., Gallagher, L., & Wenderoth, N. (2017). Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain*, 140(1), 235–246. <https://doi.org/10.1093/brain/aww287>

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289. <https://doi.org/10.1016/j.tics.2007.05.005>

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 1–43. <https://doi.org/10.1016/j.jml.2012.11.001>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. *Frontiers in Neuroscience*, 8, 60. <https://doi.org/10.3389/fnins.2014.00060>

Bhatara, A., Quintin, E.-M., Levy, B., Bellugi, U., Fombarone, E., & Levitin, D. J. (2010). Perception of emotion in musical performance in adolescents with autism spectrum disorders. *Autism Research: Official Journal of the International Society for Autism Research*, 3(5), 214–225. <https://doi.org/10.1002/aur.147>

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, 42(3), 665–670. <https://doi.org/10.3758/BRM.42.3.665>

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

Booth, R., & Happé, F. (2010). 'Hunting with a knife and Fork': Examining central coherence in autism, attention deficit/hyperactivity disorder, and typical development with a linguistic task. *Journal of Experimental Child Psychology*, 107(4), 377–393. <https://doi.org/10.1016/j.jecp.2010.06.003>

Brignell, A., Morgan, A. T., Woolfenden, S., Klopper, F., May, T., Sarkozy, V., & Williams, K. (2018). A systematic review and meta-analysis of the prognosis of language outcomes for individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, 3, 2396941518767610. <https://doi.org/10.1177/2396941518767610>

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Trends in Cognitive Sciences*, 16(12), 573–574. <https://doi.org/10.1016/j.tics.2012.10.005>

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, 108(3), 896–904. <https://doi.org/10.1016/j.cognition.2008.06.007>

Brown, H. M., Oram-Cardy, J., & Johnson, A. (2013). A meta-analysis of the Reading comprehension skills of individuals on the autism spectrum. *Journal of Autism and Developmental Disorders*, 43(4), 932–955. <https://doi.org/10.1007/s10803-012-1638-1>

Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in autism spectrum disorder: A systematic review of empirical evidence. *Autism Research*, 14(4), 604–630. <https://doi.org/10.1002/aur.2482>

Caria, A., Venuti, P., & de Falco, S. (2011). Functional and dysfunctional brain circuits underlying emotional processing of music in autism spectrum disorders. *Cerebral Cortex*, 21(12), 2838–2849. <https://doi.org/10.1093/cercor/bhr084>

Catarino, A., Luke, L., Waldman, S., Andrade, A., Fletcher, P. C., & Ring, H. (2011). An fMRI investigation of detection of semantic incongruities in autistic spectrum conditions. *European Journal of Neuroscience*, 33(3), 558–567. <https://doi.org/10.1111/j.1460-9568.2010.07503.x>

Chen, Y., Tang, E., Ding, H., & Zhang, Y. (2022). Auditory pitch perception in autism spectrum disorder: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 65, 4866–4886.

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology*, 29(23), 4084–4092.e4. <https://doi.org/10.1016/j.cub.2019.09.067>

Chik, P. P., Ho, C. S., Yeung, P., Chan, D. W., Chung, K. K., Luan, H., Lo, L., & Lau, W. S. (2012). Syntactic skills in sentence reading comprehension among Chinese elementary school children. *Reading and Writing*, 25(3), 679–699. <https://doi.org/10.1007/s11145-010-9293-4>

Chowdhury, R., Sharda, M., Foster, N. E. V., Germain, E., Tryfon, A., Doyle-Thomas, K., Anagnostou, E., & Hyde, K. L. (2017). Auditory pitch perception in autism spectrum disorder is associated with nonverbal abilities. *Perception*, 46(11), 1298–1320. <https://doi.org/10.1177/0301006617718715>

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371. <https://doi.org/10.1016/j.cognition.2009.10.009>

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and 'phenotypes' of poor singing. *Annals of the New York Academy of Sciences*, 1169, 99–107. <https://doi.org/10.1111/j.1749-6632.2009.04558.x>

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. <https://doi.org/10.1038/nn1504>

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

DePriest, J., Glushko, A., Steinhauer, K., & Koelsch, S. (2017). Language and music phrase boundary processing in autism spectrum disorder: An ERP study. *Scientific Reports*, 7(1), 14465. <https://doi.org/10.1038/s41598-017-14538-y>

Di Vesta, F. J., Hayward, K. G., & Orlando, V. P. (1979). Developmental trends in monitoring text for comprehension. *Child Development*, 50(1), 97–105. <https://doi.org/10.2307/1129046>

DiStefano, C., Senturk, D., & Jeste, S. S. (2019). ERP evidence of semantic processing in children with ASD. *Developmental Cognitive Neuroscience*, 36, 100640. <https://doi.org/10.1016/j.dcn.2019.100640>

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised*. American Guidance Service.

Eberhardt, M., & Nadig, A. (2018). Reduced sensitivity to context in language comprehension: A characteristic of autism spectrum disorders or of poor structural language ability? *Research in Developmental Disabilities*, 72, 284–296. <https://doi.org/10.1016/j.ridd.2016.01.017>

Eccles, J. S. (1999). The development of children ages 6 to 14. *The Future of Children*, 9(2), 30–44. <https://doi.org/10.2307/11602703>

Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3), 533–553. <https://doi.org/10.3758/s13415-013-0161-y>

Eigsti, I.-M., & Fein, D. A. (2013). More is less: Pitch discrimination and language delays in children with optimal outcomes from autism: Pitch perception, language delay, and outcomes. *Autism Research*, 6(6), 605–613. <https://doi.org/10.1002/aur.1324>

Ellis Weismer, S., & Saffran, J. R. (2022). Differences in prediction may underlie language disorder in autism. *Frontiers in Psychology*, 13, 897187. <https://doi.org/10.3389/fpsyg.2022.897187>

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition*, 128(1), 82–102. <https://doi.org/10.1016/j.cognition.2012.12.006>

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37, 66–108. <https://doi.org/10.1016/j.dr.2015.05.002>

Fang, L., Shang, J., & Chen, N. (2017). Perception of Western musical modes: A Chinese study. *Frontiers in Psychology*, 8, 1905. <https://doi.org/10.3389/fpsyg.2017.01905>

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, 27(6), 443–448. <https://doi.org/10.1177/0963721418794491>

Fogel, A. R., Rosenberg, J. C., Lehman, F. M., Kuperberg, G. R., & Patel, A. D. (2015). Studying musical and linguistic prediction in comparable ways: The melodic cloze probability method. *Frontiers in Psychology*, 6, 1718. <https://doi.org/10.3389/fpsyg.2015.01718>

Font-Alamillos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., & Escera, C. (2020). Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biological Psychology*, 149, 107807. <https://doi.org/10.1016/j.biopsych.2019.107807>

Foti, F., Crescenzo, F. D., Vivanti, G., Menghini, D., & Vicari, S. (2015). Implicit learning in individuals with autism spectrum disorders: A meta-analysis. *Psychological Medicine*, 45(5), 897–910. <https://doi.org/10.1017/S0033291714001950>

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage Publications.

Frith, U., & Happé, F. G. E. (1994). Autism: Beyond "theory of mind". *Cognition*, 50(1), 115–132. [https://doi.org/10.1016/0010-0277\(94\)90024-8](https://doi.org/10.1016/0010-0277(94)90024-8)

Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, 1(4), 329–342. <https://doi.org/10.1111/j.2044-835X.1983.tb00906.x>

Fröhholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, 68, 96–110. <https://doi.org/10.1016/j.neubiorev.2016.05.002>

Gebauer, L., Skewes, J., Westphael, G., Heaton, P., & Vuust, P. (2014). Intact brain processing of musical emotions in autism spectrum disorder, but more cognitive load and arousal in happy vs. sad music. *Frontiers in Neuroscience*, 8, 192. <https://doi.org/10.3389/fnins.2014.00192>

Germain, E., Foster, N. E. V., Sharda, M., Chowdhury, R., Tryfon, A., Doyle-Thomas, K. A. R., Anagnostou, E., & Hyde, K. L. (2019). Pitch direction ability predicts melodic perception in autism. *Child Neuropsychology*, 25(4), 445–465. <https://doi.org/10.1080/09297049.2018.1488954>

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76. [https://doi.org/10.1016/S0010-0277\(98\)00034-1](https://doi.org/10.1016/S0010-0277(98)00034-1)

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research*, 8(2), 153–163. <https://doi.org/10.1002/aur.1432>

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *Journal of Neuroscience*, 39(47), 9397–9409. <https://doi.org/10.1523/JNEUROSCI.0428-19.2019>

Gomot, M., & Wicker, B. (2012). A challenging, unpredictable world for people with autism spectrum disorder. *International Journal of Psychophysiology*, 83(2), 240–247. <https://doi.org/10.1016/j.ijpsycho.2011.09.017>

Goris, J., Braem, S., Nijhof, A. D., Rigoni, D., Deschrijver, E., Van de Cruys, S., Wiersema, J. R., & Brass, M. (2018). Sensory prediction errors are less modulated by global context in autism spectrum disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(8), 667–674. <https://doi.org/10.1016/j.bpsc.2018.02.003>

Greene, R. K., Zheng, S., Kinard, J. L., Mosner, M. G., Wiesen, C. A., Kennedy, D. P., & Dichter, G. S. (2019). Social and nonsocial visual prediction errors in autism spectrum disorder. *Autism Research*, 12(6), 878–883. <https://doi.org/10.1002/aur.2090>

Grisoni, L., Moseley, R. L., Motlagh, S., Kandia, D., Sener, N., Pulvermüller, F., Roepke, S., & Mohr, B. (2019). Prediction and mismatch negativity responses reflect impairments in action semantic processing in adults with autism spectrum disorders. *Frontiers in Human Neuroscience*, 13, 395. <https://doi.org/10.3389/fnhum.2019.00395>

Háden, G. P., Németh, R., Török, M., & Winkler, I. (2015). Predictive processing of pitch trends in newborn infants. *Brain Research*, 1626, 14–20. <https://doi.org/10.1016/j.brainres.2015.02.048>

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1481), 801–811. <https://doi.org/10.1098/rstb.2007.2089>

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, 5, 1052. <https://doi.org/10.3389/fpsyg.2014.01052>

Happé, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology*, 15(1), 1–12. <https://doi.org/10.1111/j.2044-835X.1997.tb00721.x>

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

Hart, C. M., & Curtin, S. (2023). Trajectories of vocabulary development in children with autism spectrum disorder across multiple measures. *Journal of Autism and Developmental Disorders*, 53(4), 1347–1362. <https://doi.org/10.1007/s10803-021-05379-w>

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8, 35. <https://doi.org/10.3389/fnsys.2014.00035>

Heaton, P. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 44(4), 543–551.

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

Heaton, P., Hermelin, B., & Pring, L. (1999). Can children with autistic spectrum disorders perceive affect in music? An experimental investigation. *Psychological Medicine*, 29(6), 1405–1410.

Heaton, P., Pring, L., & Hermelin, B. (1999). A pseudo-savant: A case of exceptional musical splinter skills. *Neurocase*, 5(6), 503–509. <https://doi.org/10.1080/1355479908402745>

Heaton, P., Tsang, W., Jakubowski, K., Mullensiefen, D., & Allen, R. (2018). Discriminating autism and language impairment and specific language impairment through acuity of musical imagery. *Research in Developmental Disabilities*, 80, 52–63. <https://doi.org/10.1016/j.ridd.2018.06.001>

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

Howard, P. L., Liversedge, S. P., & Benson, V. (2017). Investigating the use of world knowledge during on-line comprehension in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47(7), 2039–2053. <https://doi.org/10.1007/s10803-017-3129-x>

Huang, H. (2012). Why Chinese people play Western classical music: Transcultural roots of music philosophy. *International Journal of Music Education*, 30(2), 161–176. <https://doi.org/10.1177/0255761411420955>

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation* (p. 462). The MIT Press.

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*, 44(6), 1086–1105. <https://doi.org/10.1080/02664763.2016.1193725>

Jamey, K., Foster, N. E. V., Sharda, M., Tuerk, C., Nadig, A., & Hyde, K. L. (2019). Evidence for intact melodic and rhythmic perception in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 64, 1–12. <https://doi.org/10.1016/j.rasd.2018.11.013>

Janzen, T. B., & Thaut, M. H. (2018). Rethinking the role of music in the neurodevelopment of autism spectrum disorder. *Music & Science*, 1, 205920431876963. <https://doi.org/10.1177/2059204318769639>

Jiang, C., Liu, F., & Thompson, W. F. (2016). Impaired explicit processing of musical syntax and tonality in a group of Mandarin-speaking congenital amusics. *Music Perception: An Interdisciplinary Journal*, 33(4), 401–413.

Jiang, C., Liu, F., & Wong, P. C. M. (2017). Sensitivity to musical emotion is influenced by tonal structure in congenital amusia. *Scientific Reports*, 7(1), 7624. <https://doi.org/10.1038/s41598-017-08005-x>

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of melodic contour and intonation in autism spectrum disorder: Evidence from mandarin speakers. *Journal of Autism and Developmental Disorders*, 1–9, 2067–2075. <https://doi.org/10.1007/s10803-015-2370-4>

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger syndrome: Is local coherence impaired? *Cognition*, 71(2), 149–185. [https://doi.org/10.1016/S0010-0277\(99\)00022-0](https://doi.org/10.1016/S0010-0277(99)00022-0)

Jones, C. R. G., Happé, F., Baird, G., Simonoff, E., Marsden, A. J. S., Tregay, J., Phillips, R. J., Goswami, U., Thomson, J. M., & Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, 47(13), 2850–2858. <https://doi.org/10.1016/j.neuropsychologia.2009.06.015>

Kamio, Y., Robins, D., Kelley, E., Swainson, B., & Fein, D. (2007). Atypical lexical/semantic processing in high-functioning autism spectrum disorders without early language delay. *Journal of Autism and Developmental Disorders*, 37(6), 1116–1122. <https://doi.org/10.1007/s10803-006-0254-3>

Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., & Pellecchio, E. (2016). Central tendency effects in time interval reproduction in autism. *Scientific Reports*, 6, 28570. <https://doi.org/10.1038/srep28570>

Kargas, N., López, 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. <https://doi.org/10.1007/s10803-014-2219-2>

Kasten, K., Jacoby, N., & Ahissar, M. (2023). Poor synchronization yet adequate tempo-keeping in adults with autism. *Autism Research*, 16(6), 1161–1173. <https://doi.org/10.1002/aur.2926>

Kinard, J. L., Mosner, M. G., Greene, R. K., Addicott, M., Bizzell, J., Petty, C., Cernasov, P., Walsh, E., Eisenlohr-Moul, T., Carter, R. M., McLamb, M., Hopper, A., Sukhu, R., & Dichter, G. S. (2020). Neural mechanisms of social and nonsocial reward prediction errors in adolescents with autism spectrum disorder. *Autism Research*, 13(5), 715–728. <https://doi.org/10.1002/aur.2273>

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>

Knight, E. J., Oakes, L., Hyman, S. L., Freedman, E. G., & Foxe, J. J. (2020). Individuals with autism have no detectable deficit in neural markers of prediction error when presented with auditory rhythms of varied temporal complexity. *Autism Research*, 13(12), 2058–2072. <https://doi.org/10.1002/aur.2362>

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, 23(1), 63–77. <https://doi.org/10.1016/j.tics.2018.10.006>

Koolen, S., Vissers, C. T. W. M., Egger, J. I. M., & Verhoeven, L. (2013). Can monitoring in language comprehension in autism spectrum disorder be modulated? Evidence from event-related potentials. *Biological Psychology*, 94(2), 354–368. <https://doi.org/10.1016/j.biopsych.2013.07.008>

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making sense of the world: Infant learning from a predictive processing perspective. *Perspectives on Psychological Science*, 15(3), 562–571. <https://doi.org/10.1177/1745691619895071>

Krumhansl, C. L. (1997). Effects of perceptual organization and musical form on melodic expectancies. In M. Leman (Ed.), *Music, gestalt, and computing* (pp. 294–320). Springer. <https://doi.org/10.1007/BFb0034122>

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334–368. <https://doi.org/10.1037/0033-295X.89.4.334>

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5(4), 579–594.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1417–1432. <https://doi.org/10.1037/0278-7393.31.6.1417>

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–163. <https://doi.org/10.1038/307161a0>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, 8, 302. <https://doi.org/10.3389/fnhum.2014.00302>

Lenroot, R., & Yeung, P. K. (2013). Heterogeneity within autism spectrum disorders: What have we learned from neuroimaging studies? *Frontiers in Human Neuroscience*, 7, 7.

Lenth, R., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means (1.5.0). <https://CRAN.R-project.org/package=emmeans>

Leung, F. Y. N., Sin, J., Dawson, C., Ong, J. H., Zhao, C., Veic, A., & Liu, F. (2022). Emotion recognition across visual and auditory modalities in autism spectrum disorder: A systematic review and meta-analysis. *Developmental Review*, 63, 1–47.

Leung, F. Y. N., Stojanovik, V., Micai, M., Jiang, C., & Liu, F. (2023). Emotion recognition in autism spectrum disorder across age groups: A cross-sectional investigation of various visual and auditory communicative domains. *Autism Research: Official Journal of the International Society for Autism Research*, 16, 783–801. <https://doi.org/10.1002/aur.2896>

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., & Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience*, 22(2), 256–264. <https://doi.org/10.1038/s41593-018-0308-9>

Lin, I.-F., Mochida, T., Asada, K., Ayaya, S., Kumagaya, S., & Kato, M. (2015). Atypical delayed auditory feedback effect and Lombard effect on speech production in high-functioning adults with autism spectrum disorder. *Frontiers in Human Neuroscience*, 9, 510. <https://doi.org/10.3389/fnhum.2015.00510>

Lord, C., Luyster, R. J., Gotham, K., & Guthrie, W. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2) manual (part II): Toddler module*. Western Psychological Services.

Lucas, R., & Norbury, C. F. (2014). Levels of text comprehension in children with autism spectrum disorders (ASD): The influence of language phenotype. *Journal of Autism and Developmental Disorders*, 44(11), 2756–2768. <https://doi.org/10.1007/s10803-014-2133-7>

Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics*, 70(5), 841–852. <https://doi.org/10.3758/PP.70.5.841>

Martin, N. A., & Brownell, R. (2011). *ROWPVT-4: Receptive one-word picture vocabulary test-fourth edition*. Pearson Assessment.

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. <https://doi.org/10.1007/s10803-014-2207-6>

McIntyre, N. S., Solari, E. J., Grimm, R. P., Lerro, E., Gonzales, E., & Mundy, P. C. (2017). A comprehensive examination of reading heterogeneity in students with high functioning autism: Distinct reading profiles and their relation to autism symptom severity. *Journal of Autism and Developmental Disorders*, 47(4), 1086–1101. <https://doi.org/10.1007/s10803-017-3029-0>

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *The American Journal of Psychology*, 63(2), 176–185. <https://doi.org/10.2307/1418920>

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and gestalt-like principles predict melodic expectations. *Cognition*, 189, 23–34. <https://doi.org/10.1016/j.cognition.2018.12.015>

Morimoto, C., Hida, E., Shima, K., & Okamura, H. (2018). Temporal processing instability with millisecond accuracy is a cardinal feature of sensorimotor impairments in autism spectrum disorder: Analysis using the synchronized finger-tapping task. *Journal of Autism and Developmental Disorders*, 48(2), 351–360. <https://doi.org/10.1007/s10803-017-3334-7>

Mottron, L., Dawson, M., & Soulières, I. (2009). Enhanced perception in savant syndrome: Patterns, structure and creativity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1522), 1385–1391. <https://doi.org/10.1098/rstb.2008.0333>

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

Nan, Y., Knösche, T. R., & Friederici, A. D. (2009). Non-musicians' perception of phrase boundaries in music: A cross-cultural ERP study. *Biological Psychology*, 82(1), 70–81. <https://doi.org/10.1016/j.biopsych.2009.06.002>

Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 36(7), 911–919. <https://doi.org/10.1007/s10803-006-0130-1>

Neville, M. H., & Pugh, A. K. (1976). Context in reading and listening: Variations in approach to cloze tasks. *Reading Research Quarterly*, 12(1), 13–31. <https://doi.org/10.2307/747495>

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., ... Wolfsturn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791), 20180522. <https://doi.org/10.1098/rstb.2018.0522>

Nippold, M. (2000). Language development during the adolescent years: Aspects of pragmatics, syntax, and semantics. *Topics in Language Disorders*, 20, 15–28. <https://doi.org/10.1097/00011363-200020020-00004>

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, 90(2), 142–171. <https://doi.org/10.1016/j.jecp.2004.11.003>

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. G. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, 7, 1245. <https://doi.org/10.3389/fpsyg.2016.01245>

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>

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

Ong, J. H., Zhao, C., Bacon, A., Leung, F. Y. N., Veic, A., Wang, L., Jiang, C., & Liu, F. (2023). The relationship between autism and pitch perception is modulated by cognitive abilities. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-023-06075-7>

O'Rourke, E., & Coderre, E. L. (2021). Implicit semantic processing of linguistic and non-linguistic stimuli in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 51(8), 2611–2630. <https://doi.org/10.1007/s10803-020-04736-5>

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

Palumbo, L., Burnett, H. G., & Jellema, T. (2015). Atypical emotional anticipation in high-functioning autism. *Molecular Autism*, 6(1), 47. <https://doi.org/10.1186/s13229-015-0039-7>

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674–681.

Patel, A. D. (2013). Sharing and nonsharing of brain resources for language and music. In M. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship* (pp. 329–355). MIT Press.

Patel, A. D., & Morgan, E. (2017). Exploring cognitive relations between prediction in language and music. *Cognitive Science*, 41(S2), 303–320. <https://doi.org/10.1111/cogs.12411>

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, 9(2), 60–68. <https://doi.org/10.1016/j.tics.2004.12.008>

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, 23(5), 377–405. <https://doi.org/10.1525/mp.2006.23.5.377>

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504–510. <https://doi.org/10.1016/j.tics.2012.08.009>

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5), 233–238. <https://doi.org/10.1016/j.tics.2006.03.006>

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America*, 128(4), 2182–2190. <https://doi.org/10.1121/1.3478782>

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, 20(4), 747–753. <https://doi.org/10.3758/s13423-013-0401-8>

Piper, B., Mueller, S. T., Talebzadeh, S., & Ki, M. J. (2016). Evaluation of the validity of the psychology experiment building language tests of vigilance, auditory memory, and decision making. *PeerJ*, 4, e1772. <https://doi.org/10.7717/peerj.1772>

Prescott, K. E., Mathéée-Scott, J., Reuter, T., Edwards, J., Saffran, J., & Ellis Weismer, S. (2022). Predictive language processing in young autistic children. *Autism Research*, 15(5), 892–903. <https://doi.org/10.1002/aur.2684>

Quintin, E.-M. (2019). Music-evoked reward and emotion: Relative strengths and response to intervention of people with ASD. *Frontiers in Neural Circuits*, 13, 49. <https://doi.org/10.3389/fncir.2019.00049>

Quintin, E.-M., Bhatara, A., Poissant, H., Fombonne, E., & Levitin, D. J. (2011). Emotion perception in music in high-functioning adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 41(9), 1240–1255. <https://doi.org/10.1007/s10803-010-1146-0>

Quintin, E.-M., Bhatara, A., Poissant, H., Fombonne, E., & Levitin, D. J. (2013). Processing of musical structure by high-functioning adolescents with autism spectrum disorders. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 19(3), 250–275. <https://doi.org/10.1080/09297049.2011.653540>

R: A Language and Environment for Statistical Computing. (2019). R Foundation for Statistical Computing. <https://www.R-project.org/>

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

Ricketts, J., Jones, C. R. G., Happé, F. G. E., & Charman, T. (2013). Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of Autism and Developmental Disorders*, 43(4), 807–816. <https://doi.org/10.1007/s10803-012-1619-4>

Rimland, B., & Fein, D. (1988). Special talents of autistic savants. In *The exceptional brain: Neuropsychology of talent and special abilities* (pp. 474–492). Guilford Press.

Ring, H., Sharma, S., Wheelwright, S., & Barrett, G. (2007). An electrophysiological investigation of semantic incongruity processing by people with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 37(2), 281–290. <https://doi.org/10.1007/s10803-006-0167-1>

Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, 83(2), 164–175. <https://doi.org/10.1016/j.ijpsycho.2011.12.010>

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, 12(4), 110–114. <https://doi.org/10.1111/1467-8721.01243>

Sang, B., & Miao, X. (1990). The revision of trail norm of Peabody picture vocabulary test revised (PPVT-R) in Shanghai proper. *Psychological Science*, 5, 20–25.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, 58(1), 75–125. [https://doi.org/10.1016/0010-0277\(95\)00665-6](https://doi.org/10.1016/0010-0277(95)00665-6)

Schuwerk, T., Sodian, B., & Paulus, M. (2016). Cognitive mechanisms underlying action prediction in children and adults with autism spectrum condition. *Journal of Autism and Developmental Disorders*, 46(12), 3623–3639. <https://doi.org/10.1007/s10803-016-2899-x>

Sheppard, E., van Loon, E., Underwood, G., & Ropar, D. (2016). Difficulties predicting time-to-arrival in individuals with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 28, 17–23. <https://doi.org/10.1016/j.rasd.2016.05.001>

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., & Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, 111(42), 15220–15225. <https://doi.org/10.1073/pnas.1416797111>

Slevc, L. R. (2012). Language and music: Sound, structure, and meaning. *WIREs Cognitive Science*, 3(4), 483–492. <https://doi.org/10.1002/wcs.1186>

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319. <https://doi.org/10.1016/j.cognition.2013.02.013>

Sota, S., Hatada, S., Honjyo, K., Takatsuka, T., Honer, W. G., Morinobu, S., & Sawada, K. (2018). Musical disability in children with autism spectrum disorder. *Psychiatry Research*, 267, 354–359. <https://doi.org/10.1016/j.psychres.2018.05.078>

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, 82, 1–17. <https://doi.org/10.1016/j.jml.2015.02.004>

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, 18(8), 1380–1393. <https://doi.org/10.1162/jocn.2006.18.8.1380>

Steinberg, S., Shivers, C. M., Liu, T., Cirelli, L. K., & Lense, M. D. (2021). Survey of the home music environment of children with various developmental profiles. *Journal of Applied Developmental Psychology*, 75, 101296. <https://doi.org/10.1016/j.appdev.2021.101296>

Stringer, L., & Iverson, P. (2020). Non-native speech recognition sentences: A new materials set for non-native speech perception research. *Behavior Research Methods*, 52(2), 561–571. <https://doi.org/10.3758/s13428-019-01251-z>

Sun, L., Thompson, W. F., Liu, F., Zhou, L., & Jiang, C. (2020). The human brain processes hierarchical structures of meter and harmony differently: Evidence from musicians and nonmusicians. *Psychophysiology*, 57(9), e13598. <https://doi.org/10.1111/psyp.13598>

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

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, 30(4), 415–433. <https://doi.org/10.1177/107769905303000401>

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, 32(2), 418–444. <https://doi.org/10.1080/03640210701864089>

Temperley, D. (2014). Probabilistic models of melodic interval. *Music Perception*, 32(1), 85–99. <https://doi.org/10.1525/mp.2014.32.1.85>

The MathWorks Inc. (2022). *MATLAB version: 9.12.0 (R2022a) [Computer software]*. The MathWorks Inc. <https://www.mathworks.com>

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, 112(32), 10062–10067. <https://doi.org/10.1073/pnas.1505114112>

Trainor, L. J. (2012). Predictive information processing is a fundamental learning mechanism present in early development: Evidence from infants. *International Journal of Psychophysiology*, 83(2), 256–258. <https://doi.org/10.1016/j.ijpsycho.2011.12.008>

van Boxtel, J. J. A., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, 4, 19. <https://doi.org/10.3389/fpsyg.2013.00019>

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, 121(4), 649–675. <https://doi.org/10.1037/a0037665>

Venter, F., Morelli, J., & Erasmus, E. (2023). Understanding the lived music listening experiences of adults on the autism spectrum. *Psychology of Music*, 51(3), 971–985. <https://doi.org/10.1177/03057356221126201>

Verosky, N. J. (2019). Corpus-based learning of tonal expectations with expectation networks. *Journal of New Music Research*, 48(2), 145–158. <https://doi.org/10.1080/09298215.2018.1542447>

Verosky, N. J., & Morgan, E. (2021). Pitches that wire together fire together: Scale degree associations across time predict melodic expectations. *Cognitive Science*, 45(10), e13037. <https://doi.org/10.1111/cogs.13037>

Vishne, G., Jacoby, N., Malinovitch, T., Epstein, T., Frenkel, O., & Ahissar, M. (2021). Slow update of internal representations impedes synchronization in autism. *Nature Communications*, 12(1), 5439. <https://doi.org/10.1038/s41467-021-25740-y>

Vuust, P., Heggli, O. A., Friston, K. J., & Kringelbach, M. L. (2022). Music in the brain. *Nature Reviews Neuroscience*, 23(5), 287–305. <https://doi.org/10.1038/s41583-022-00578-5>

Wang, L., Beaman, C. P., Jiang, C., & Liu, F. (2022). Perception and production of statement-question intonation in autism spectrum disorder: A developmental investigation. *Journal of Autism and Developmental Disorders*, 52(8), 3456–3472. <https://doi.org/10.1007/s10803-021-05220-4>

Wang, L., Ong, J. H., Ponsot, E., Hou, Q., Jiang, C., & Liu, F. (2023). Mental representations of speech and musical pitch contours reveal a diversity of profiles in autism spectrum disorder. *Autism*, 27(3), 629. <https://doi.org/10.1177/1362361322111207>

Wang, L., Pfordresher, P. Q., Jiang, C., & Liu, F. (2021). Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation. *Autism Research*, 14(11), 2355–2372. <https://doi.org/10.1002/aur.2569>

Wang, L., Xiao, S., Jiang, C., Hou, Q., Chan, A. H. D., Wong, P. C. M., & Liu, F. (2023). The form and function processing of lexical tone and intonation in tone-language-speaking children with autism spectrum disorder. *The Journal of the Acoustical Society of America*, 143(4), 2601–2612. <https://doi.org/10.1121/1540819231177001>

*Society of America, 154*(1), 467–481. <https://doi.org/10.1121/10.0020271>

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition Administration and Scoring Manual*. San Antonio, TX: Pearson.

Xu, J., Zhou, L., Liu, F., Xue, C., Jiang, J., & Jiang, C. (2022). The autistic brain can process local but not global emotion regularities in facial and musical sequences. *Autism Research, 15*(2), 222–240. <https://doi.org/10.1002/aur.2635>

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody, 01-1*, 7–10.

Yang, Y., & Welch, G. (2023). A systematic literature review of Chinese music education studies during 2007 to 2019. *International Journal of Music Education, 41*(2), 175–198. <https://doi.org/10.1177/02557614221096150>

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

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

Zhang, H.-c., & Wang, X.-p. (1989). Standardization research on Raven's Standard Progressive Matrices in China. *Acta Psychologica Sinica, 21*(2), 113–121.

Zhang, M., Xu, S., Chen, Y., Lin, Y., Ding, H., & Zhang, Y. (2022). Recognition of affective prosody in autism spectrum conditions: A systematic review and meta-analysis. *Autism, 26*(4), 798–813. <https://doi.org/10.1177/1362361321995725>

Zhou, L., Liu, F., Jiang, J., Jiang, H., & Jiang, C. (2019). Abnormal neural responses to harmonic syntactic structures in congenital amusia. *Psychophysiology, 56*(9), e13394. <https://doi.org/10.1111/psyp.13394>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Zhao, C., Ong, J. H., Veic, A., Patel, A. D., Jiang, C., Fogel, A. R., Wang, L., Hou, Q., Das, D., Crasto, C., Chakrabarti, B., Williams, T. I., Loutrari, A., & Liu, F. (2024). Predictive processing of music and language in autism: Evidence from Mandarin and English speakers. *Autism Research, 17*(6), 1230–1257. <https://doi.org/10.1002/aur.3133>