Pitch Decoding and Encoding in Individuals with and without Autism Spectrum Disorder

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

School of Psychology and Clinical Language Sciences

University of Reading

Li Wang
July 2021

Abstract

How pitch is processed has been investigated across different auditory domains (speech vs. music), processing levels (low-level vs. high-level), and modalities (perception vs. production) in individuals with autism spectrum disorder (ASD). However, mixed results have been reported with no substantial evidence to inform the ongoing theoretical debates in ASD, such as (a) whether speech and music processing share underlying mechanisms; (b) whether perception correlates with production; and (c) whether high-level information processing is intact in individuals with ASD. The present thesis reports three studies that examine pitch processing in individuals with and without ASD to inform the three aforementioned theoretical debates and to reconcile inconsistencies across studies. In relation to pitch processing in speech versus music, the results show that pitch perception in both domains is intact in individuals with ASD. However, when imitating speech and song, individuals with ASD demonstrate impaired absolute but not relative pitch production. These findings from perception and production suggest that speech and music pitch processing, whether intact or impaired, likely share underlying mechanisms. With respect to the relationship between perception and production, the findings indicate an association between these two modalities, since the ability to identify statement-question intonation is associated with the ability to imitate the intonation in both groups. Concerning low- and high-level processing, the findings reveal that high-level processing is not impaired in ASD, and that sensitivity to low-level pitch predicts performance on higher-level processing in both groups. Furthermore, perception ability increases with age in individuals with and without ASD whereas imitation ability does not. The findings from this thesis heighten our understanding of how pitch is decoded and encoded in ASD, and provide theoretical implications for pitch processing in this population and in typical development.

Acknowledgments

I would first like to express my deepest gratitude to my fantastic supervisors, Fang Liu and Beaman Philip. I am very fortunate to have Fang as my first supervisor, who has seen my potential and has been a source of encouragement and inspiration to me. Without her constant support and guidance, I would surely not have been able to complete my Ph.D. in the field of Psychology. My second supervisor, Phil, has enriched my understanding of Bayesian analysis. His patient guidance and useful critiques of this research work, as well as his willingness to give his time so generously, are very much appreciated. Many thanks also go to my other co-authors, Cunmei Jiang, Peter Pfordresher, Jia Hoong Ong, Emmanuel Ponsot, and Jean-Julien Aucouturier for providing insightful suggestions on my projects and manuscripts. I would also like to extend my appreciation to my fellow CAASD lab members - Florence Leung, Jia Hoong Ong, Chen Zhao, Anamarija Veic, and Ariadne Loutrari for their unwavering support and willingness to help me. They have colored my Ph.D. life. I thankfully acknowledge the funding received towards my PhD from the European Research Council (678733, CAASD), which has relieved my financial burden and allowed my research to go smoothly. Furthermore, I owe special thanks to my Ph.D. fellow Yesi Cheng, who successfully became Dr. Cheng in June of this year. He is not only a considerable friend to me in everyday life, but also an excellent writing tutor on this research work. Without his help and encouragement, I would not have been able to survive this bittersweet journey. Last but not least, I would really like to thank all the participants of my experiments and their families for their time and participation.

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Li Wang

Declaration of Authorship

Chapters 2, 3 and 4 are based on the following papers, which report on the original work that I conducted in collaboration with my supervisors and other co-authors during my Ph.D. period.

Study 1

Wang, L., Beaman, C. P., Jiang, C., & Liu, F. (2021). Perception and Production of Statement-Question Intonation in Autism Spectrum Disorder: A Developmental Investigation. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-021-05220-4

Study 2

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*. https://doi.org/10.1002/aur.2569

Study 3

Wang, L., Ong, J., Ponsot, E., Aucouturier, J., Jiang, C., & Liu, F. (To be submitted). Mental representations of speech and music pitch contours in autism spectrum disorder

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1. General Introduction

ASD is a complex neurodevelopmental disorder, well known for impairments in social communication and interaction, as well as restricted and repetitive behaviours and interests (American Psychiatric Association, 1994). More recently, an increasing number of studies have shown that ASD is also associated with atypical auditory processing (Germain et al., 2019; Haesen et al., 2011; O'Connor, 2012). Indeed, subsequent to the publication of the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders—5th edition), sensory issues, including hearing (e.g., adverse response to specific sounds), were added to the symptoms that diagnose ASD (American Psychiatric Association, 2013). As a key auditory attribute of sounds, pitch is ubiquitous in our everyday listening experience, including language communication and music appreciation (Plack et al., 2006, 2014; Plack & Oxenham, 2005). Given its necessity in auditory processing, numerous studies have investigated pitch decoding and encoding in ASD. Mixed findings have been reported in studies examining either pitch perception or production, and the results can vary depending on the auditory domains (speech vs. music) and the processing levels (low-level vs. high-level) (Haesen et al., 2011; O'Connor, 2012; Ouimet et al., 2012). To reconcile the inconsistencies, this thesis reports three studies that used comparative and developmental designs to investigate pitch perception and production across domains and processing levels in individuals with and without ASD. In particular, this thesis has four general aims. The aims and how these aims are met in a specific study will be described below.

One aim of this thesis is to inform a longstanding theoretical debate in psychology on whether speech and music share underlying processing mechanisms. Within research with typically developing individuals and individuals with auditory disorders (e.g., congenital amusia), some researchers propose that speech and music processing may involve distinct mechanisms, due to dissociation between performance on speech and music observed in auditory disorders (Fodor, 1983, 2001; Peretz, 2009; Peretz et al., 2015; Peretz & Coltheart, 2003; Peretz & Zatorre, 2005), whilst others argue that the mechanisms used to process information are shared between speech and music, as these domains influence one another (Koelsch, 2011; Koelsch & Siebel, 2005; A. D. Patel, 2010; Sammler et al., 2009). Meanwhile, within research in ASD, it has been suggested that individuals with ASD exhibit superior musical processing but inferior linguistic processing (Heaton, 2009; Lai et al., 2012; Sharda et al., 2015). This intriguing difference between linguistic and musical processing abilities makes ASD a particularly interesting testing case to help probe the modularity mystery, explicitly in terms of whether deficits can be observed only in one domain but not the other.

As previously mentioned, pitch is important for both music and speech (Plack et al., 2006). Specifically, in music, pitch is thought to be one of the most relevant perceptual factors in most forms of music and is crucial to music appreciation (Krumhansl & Shepard, 1979; Trehub et al., 1986). Individual pitches form melodies when sequentially presented as well as chords when simultaneously presented (Krumhansl, 2004). In term of speech, all languages, including tone (e.g., Mandarin, Cantonese) and intonation languages (e.g., English), use pitch to convey prosodic meanings, such as emotions and intonations (Crystal, 1969; Xu, 2005). In tone languages, pitch is also used to convey semantic information, such as lexical tones (Yip, 2002). Thus, atypical pitch processing can affect not only music appreciation, but also language development and social communication. Prior research that investigated pitch processing in either speech or music in ASD has presented an ambiguous picture where either enhanced or intact or impaired performance was observed in either domain (Bonnel et al., 2003, 2010; Haesen et al., 2011; Heaton, Hudry, et al., 2008; Jones et al., 2009; Lau et al., 2020; O'Connor, 2012). Meanwhile, this pattern of mixed findings has also been attested in a small number of studies that tapped into both domains (Cheng et al., 2017; Heaton, Hudry, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007; J. Jiang et al., 2015). To inform the theoretical debate and to better understand how ASD impacts pitch processing across domains, Studies 1, 2 and 3 use different tasks to examine pitch perception (Study 1 and 3) and production (Study 1 and 2) in speech and music in ASD.

Study 1 investigates how individuals with ASD process the pitch contours embedded in statement-question intonation in speech and music, and whether the pitch contour is processed differently across domains. While previous studies investigated pitch perception in individuals with ASD using matched speech and musical stimuli and found enhanced pitch processing ability in individuals with ASD across domains (Heaton, Hudry, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007), it is unknown that how matched speech and musical stimuli are discriminated when stimuli entail prosodic cues. Prosody is related to the suprasegmental elements of speech, such as intonation, stress and rhythm, which reflect elements beyond the literal meaning of the speech itself, such as differentiating questions and statements (Boutsen, 2003). Since ASD is often associated with prosody processing difficulties, even amongst those who have high verbal abilities and can communicate without problems (McCann et al., 2007; Peppé et al., 2007, 2011), individuals with ASD may perform worse on speech than music when stimuli require processing prosodic cues. Thus, Study 1 tests this possibility by using speech and musical stimuli that are matched with global pitch contours derived from statement-question intonation, which primarily differ in the direction of the pitch contour on the final word (i.e., statements have a downward glide and questions have an upward glide).

Study 2 examines speech and song imitation (i.e., imitating sentences either spoken or sung) in ASD using acoustic methods to gauge pitch production. While previous studies have reported atypical speech production (J. J. Diehl & Paul, 2012; Fosnot & Jun, 1999; Hubbard & Trauner, 2007; McCann et al., 2007; McCann & Peppé, 2003; Paul et al., 2008; Peppé et al., 2007, 2011) but normal to superior musical production (Applebaum et al., 1979) in individuals with ASD, the conclusion was simply drawn based on studies that either examined speech or music but not both. No existing studies have systematically investigated pitch production when speech is compared to music within the same participants. Therefore, Study 2 conducts comparative designs using sentences and melodies that are matched in linguistic content and pitch contour to address this issue.

Similar to Study 1, Study 3 also investigates pitch perception in matched speech and music tasks, but using a novel data-driven method called the "reverse-correlation paradigm". While the majority of research has used standard behavioural measures (e.g., discrimination or identification tasks), like Study 1, to examine pitch perception in either speech or music, or both, these standard behavioural measures are normally considered as hypothesis-driven studies with some inherent shortcomings. For example, the features (e.g., pitch contours with final rises for questions and falls for statements) have been controlled by experimenters (e.g., intentionally using stimuli that represent questions with a final rising contour) before these features are examined experimentally, which may cause confirmation biases and diminish individual differences (e.g., different people may ask questions in different ways) (Burred et al., 2019). In particular, while a question intonation is generally assumed to be characterized by a final rise in pitch (Kügler, 2003), in reality, the rise may occur as briefly as in a single syllable or an entire word or in an even longer sequence, and the fundamental frequency (F0) of the rise can either be lower than 50 Hz or higher than 100 Hz (Warren, 2005). To overcome these issues, Study 3 adopts the reverse-correlation paradigm. This paradigm allows us to create a large number of variants of the same utterance, with each variant having randomly manipulated signal features (e.g., pitch contour). By presenting pairs of these variants to participants, the mental representation of an "optimal stimulus" of the feature that drives participants' responses (e.g., which one of the pairs sounds more like a question) can be systematically determined.

The second aim of this thesis is to explore the relationship between pitch perception and production of speech in ASD. It is known that successful communication relies on voices to be physically produced and perceived (Hutchins & Moreno, 2013). While few studies have reported that perception and production abilities of speech prosody were associated in individuals with ASD, these findings were based on subjective ratings of production performance (Järvinen-Pasley, Peppé, et al., 2008; Peppé et al., 2007). For example, in Peppé et al. (2007), results from subjective ratings by speech and language therapists suggested that production performance in individuals with ASD was worse than controls. However, due to the fact that participants' diagnostic status was not blind to the raters, this may have resulted in rating biases. Indeed, some studies observed inconsistencies between subjective ratings and objective instrumental judgements, which likely resulted from bias in clinicians' perceptual ratings (Van Santen et al., 2010). Also, the subjective approach has made it difficult to reveal which aspects (e.g., pitch, or rhythm) contribute to the impaired production in individuals with ASD. Therefore, to re-evaluate the relationship between perception and production in ASD, I adopt acoustic analysis in Study 1, a relatively objective measure, to examine individuals' capabilities to identify (as perception) and imitate (as production) statement-question intonation that differs mainly in pitch direction of the final word.

The third aim of this thesis is to investigate how individuals with ASD process pitch when it involves low-level or high-level information, since there is an ongoing debate on whether individuals with ASD have intact low-level processing but impaired high-level processing (Frith, 1989; Happé & Booth, 2008; Happé & Frith, 2006; Mottron et al., 2006; Mottron & Burack, 2001). Low-level processing refers to the processing of information in the early stages after entering into the brain's perceptual system (Germain et al., 2019), e.g., individual tones or pure tones. Subsequent higher-level processing refers to the integration of low-level information and higher functioning cognitive processes (Germain et al., 2019; Haesen et al., 2011), e.g., utterance-level prosodic cues, global pattern recognition and manipulation in melodies (Chowdhury et al., 2017; Nahum et al., 2008; D. L. Williams et al., 2006). Some studies report that individuals with ASD have intact or enhanced low-level processing ability but impaired ability to process high-level information (Frith, 1989; Happé & Booth, 2008; Happé & Frith, 2006). In contrast, other studies suggest that ASD is associated with enhanced or intact low-level information processing abilities in the absence of high-level processing deficits (Mottron et al., 2000, 2006; Mottron & Burack, 2001).

Given that the processing of pitch is involved in both low-level perceptual and high-level cognitive processes, pitch offers us an insight into the connection between early lower-level perception and subsequent higher-level functions in ASD. One recent study has examined how individuals with ASD process different levels of pitch information using a low-level pitch direction task and a high-level melodic task (Germain et al., 2019). Participants were presented with either pairs of tones in the low-level task or melodies in the high-level task and asked to judge if the pitch pattern was going down or up. The results indicated that the ability to process pitch is intact in both tasks in ASD and that low-level pitch perception predicts higher-level musical processing (Germain et al., 2019). However, the findings reported by Germain et al. (2019) were mainly focused on music domain. Therefore, it remains to be determined how low-level and high-level pitch information are processed and correlated in speech, particularly when speech entails prosodic cues, in ASD. This is addressed in Study 1 using a low-level pitch direction task and high-level statement-question intonation tasks.

The fourth aim of this thesis is to explore the developmental changes in pitch processing in individuals with ASD. By doing so, in addition to obtaining a better understanding of the phenotypes in ASD across the lifespan, the findings may also help clarify how age interacts with pitch processing across domains, levels and modalities in individuals with ASD. It has been suggested that sensitivity to pitch increases with age in typically developing (TD) controls (i.e., children, adolescents and adults), but not in individuals with ASD (Mayer et al., 2016). However, this has only been examined using pure tones and monosyllables, which did not contain prosody information, and has not been tested in vocal production in both speech and music within the same study. Hence, this thesis fills these gaps in the literature by testing participants with and without ASD from children to adults in terms of pitch perception using stimuli that entail prosody, such as statement-question intonation in Study 1, and in terms of vocal imitation using stimuli that are either spoken or sung in Study 2.

In the remainder of this chapter, I first introduce the relevant theoretical frameworks to account for the existing debates, namely the relationship between speech versus music, perception versus production, and low-level versus high-level, and then summarize the state-of-the-art of the research in individuals with ASD with regard to the four general research questions of this thesis I have mentioned previously.

1.1 The processing of speech versus music

Speech and music are two elementary means of human communication, with extensive research comparing and contrasting the differences and similarities between these two domains (Bidelman et al., 2011; Gold et al., 2011). Consequently, various findings in psychology have raised a longstanding debate regarding whether speech and music are processed by distinct mechanisms, or, alternatively, whether the two domains share underlying processing systems. The debate primarily surrounds two frameworks: domain-specific and domain-general.

1.1.1 Domain-specific framework

The domain-specific framework proposes that speech and music may involve distinct modules or mechanisms that operate a specific aspect of the input and its output, either exclusively or more effectively than any other mechanisms (Fodor, 1983, 2001; Peretz, 2009; Peretz et al., 2015; Peretz & Coltheart, 2003; Peretz & Zatorre, 2005). The most vigorous evidence for the existence of distinct mechanisms between speech and music comes from those cases with auditory impairments that show dissociation between speech and music, such as congenital amusia (CA), and patients with brain damage (e.g., acquired amusia, verbal agnosia) (Ayotte et al., 2002; Dalla Bella et al., 2009; Marin & Perry, 1999; Mendez, 2001; Steinke et al., 2001; Tremblay-Champoux et al., 2010). Along with such evidence, the domain-specific framework postulates that it is possible to show impairments in one domain without influencing the other, and that the processing of speech and music appears to be operating via separate mechanisms. For instance, Mendez (2001) reported a case of auditory agnosia following a right temporoparietal stroke. This patient lost the ability to comprehend speech and had difficulties in recognising environmental sounds, but was able to perceive and sing melodies. The domain-specific framework has also been supported by a number of neuroimaging studies (e.g., Peretz, 2009; Peretz & Zatorre, 2005; Saito et al., 2006), suggesting a hemispheric asymmetry in the brain with the right hemisphere mainly recruited by musical processing and the left by speech.

1.1.2 Domain-general framework

Contrary to the domain-specific framework, the domain-general framework claims that the underlying mechanisms between speech and music may not be entirely separate. Instead, they are shared between these two domains (Asaridou & McQueen, 2013; Koelsch, 2011; Koelsch & Siebel, 2005; A. D. Patel, 2010; Sammler et al., 2009). The evidence to support this framework stems mainly from studies testing the bidirectional interactions between speech and music, including how musical experience influences speech, and how tone language background affects music processing (Asaridou & McQueen, 2013; Bidelman et al., 2011, 2013; Elmer et al., 2012; Wong & Perrachione, 2007). Behavioural studies have found that musical training benefits speech processing of lexical tone (Wong & Perrachione, 2007), consonant (Marie et al., 2011) and phonetic categorization (Elmer et al., 2012). For example, Marie et al. (2011) examined French musicians' and non-musicians' ability to discriminate tonal, vowel and consonantal variations in Mandarin, and found that musicians outperformed non-musicians across all tasks, suggesting that speech processing involves mechanisms, at least partially, shared with musical processing. Recent findings in individuals with congenital amusia (CA) have also found that, contrary to the original thought, the pitch deficit in CA is not domain-specific to music, because such individuals also show impairments in speech pitch perception (Liu et al., 2010; S. Nguyen et al., 2009; Vuvan et al., 2015).

In addition to these behavioural findings, the bidirectional transfer effects between speech and music have also been supported by neural evidence (Magne et al., 2006; Marques et al., 2007; Schön et al., 2004). It has been shown that subcortical and cortical auditory processing can be shaped by long-term experience occurring in either domain (Asaridou & McQueen, 2013; Bidelman et al., 2011; Oechslin et al., 2010; Wong et al., 2007). In particular, Wong et al. (2007) investigated brainstem encoding of lexical tones by gauging the frequency following response (FFR), and found that musicians showed enhanced subcortical encoding relative to non-musicians.

In summary, the aforementioned behavioural and neuroscientific findings indicate that there are bidirectional effects between music and speech. Strikingly, a domain-specific experience with language or music may shape brainstem neurons, which have domain-general consequences for the processing of speech and music (A. D. Patel, 2011). Taken together, the domain-general framework argues that the two domains share underlying processing mechanisms.

1.1.3 Domain-specificity or domain-generality in pitch processing

One central topic of the debate between the domain-specific and domain-general frameworks concerns pitch processing (Wisniewski et al., 2013), since pitch is crucial in delivering prosody and semantic information for speech, as well as constituting melodies and chords for music. The domain-specific framework proposes that tonal encoding of pitch is a music-specific module that processes the input and output of musical pitch exclusively and that it is not shared with speech pitch (Peretz & Coltheart, 2003). In contrast to this view, the domain-general framework suggests that pitch is processed in a similar manner between speech and music (Bidelman et al., 2013; Bradley, 2012; Pfordresher & Brown, 2009). For example, Bidelman et al. (2013) comprehensively examined pitch and music perception in English-speaking musicians and non-musicians, as well as Cantonese-speaking non-musicians using a set of measures, including pitch memory, pitch differences limens, and discrimination of melodies. They found that musicians showed superior performance across all tasks compared with the other two non-musician groups, and that for the non-musician groups, Cantonese speakers consistently outperformed English speakers, suggesting that long-term experiences in either music or in tone languages benefited pitch processing ability.

While the two frameworks regarding whether pitch processing uses shared mechanisms in speech and music are still a source of debate, research on individuals with ASD, who seem to show dissociation between speech and music, with superior musical processing abilities (Heaton, 2009; Molnar-Szakacs & Heaton, 2012) but largely impaired speech processing (Kwok et al., 2015; Peppé et al., 2007), might be able to inform the ongoing debate. The findings may also have implications for speech and music therapy in ASD, e.g., whether extensive musical training enhances speech perception and production. In the following sections, I provide an overview of the studies that have investigated pitch processing in speech and music in individuals with ASD. Research findings are reviewed based on different modalities, ranging from perceptive to productive findings.

1.1.4 The processing of pitch in speech versus music in ASD

1.1.4.1 Pitch perception in speech and music

In perception studies, increased or intact sensitivity to musical pitch is frequently observed in individuals with ASD (Heaton, 2009), including exceptional sensitivity to a local pitch change of one note (Heaton, Pring, et al., 1999; Heaton, 2005; Mottron et al., 2000; Stanutz et al., 2014), enhanced identification of melodic pitch contour (J. Jiang et al., 2015) and individual pitches in a chord (Heaton, 2003), intact melodic pitch direction discrimination (Germain et al., 2019), spared contour change discrimination (Altgassen et al., 2005; J. Jiang et al., 2015), as well as music perception in terms of melodic pitch, rhythm, and memory (Jamey et al., 2019). In addition, a number of studies suggested that absolute pitch (AP), even though as a rare musical gift, has been commonly reported amongst individuals with ASD (Brown et al., 2003; Heaton et al., 1998; Mottron et al., 1999; Stanutz et al., 2014). AP (also called perfect pitch) refers to an ability to name or sing isolated musical tones without taking advantage of a reference tone (Deutsch, 2002; Takeuchi & Hulse, 1993). The occurrence of AP among the general population and musicians was approximately 0.01% (Takeuchi & Hulse, 1993) and 0.64% (Brown et al., 2003) respectively, whereas the prevalence of this skill reached 5% among individuals with ASD (Miller, 1999; Rimland & Fein, 1988). Thus, pitch processing ability in the music domain appears to be an exceptional skill, or at least an unimpaired skill, in individuals with ASD.

In contrast, other neurophysiological studies have commonly revealed impairments when processing pitch-mediated linguistic information in speech (e.g., vowels and syllables) in individuals with ASD (Čeponienė et al., 2003; Kuhl et al., 2005; Lau et al., 2020; Lepistö et al., 2005, 2006; Wang et al., 2017; Yu et al., 2015; Zhang et al., 2019). For instance, by adopting event-related potential (ERP) measurements, less robust neural responses to speech stimuli were commonly observed in individuals with ASD, including poor encoding in early stages of sensory processing, which are thought to be reflected by P1 and N2 components (Čeponienė et al., 2003; Lepistö et al., 2005); diminished responses in N4 component, which reflect impaired speech classification during early cognitive processing (Lepistö et al., 2005, 2006); as well as impaired speech discrimination (reflected by MMN, a component associated with stimulus discrimination and integration) (Kuhl et al., 2005; Lepistö et al., 2006; Wang et al., 2017; Yu et al., 2015; Zhang et al., 2019) and reduced perceptual salience to speech changes (reflected by P3a, a component related to attention to important environmental stimuli) (Čeponienė et al., 2003; Lepistö et al., 2005, 2006).

However, most behavioural studies found no group difference or even enhanced sensitivity to linguistic pitch in ASD, as compared with controls (Cheng et al., 2017; Heaton, Hudry, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007; Mayer et al., 2016). For instance, using pitch sequence matching tasks in which participants were required to match four visual representations of pitch contours (i.e., rising, falling, rising-falling, falling-rising) of speech and musical analogues (five syllables and five-tone sequences), Järvinen-Pasley et al. (2008) found that participants with ASD performed significantly better than controls in both speech and music tasks. Interestingly, no domain effect was observed in either of the groups, since the two groups showed similar pitch sensitivity across speech and music domains.

The inconsistencies across studies in terms of whether speech pitch processing is impaired or not in ASD may be explained by the processing mode that is required when measuring speech pitch sensitivity. Specifically, in neurophysiological experiments, participants are often tested in a passive mode, e.g., watching a video silently while stimuli are being presented, whereas in behavioural studies, participants are generally in an active mode where they have to listen to the stimuli carefully, and are required to make forced judgments after each trial. To further investigate whether the deficit in speech processing observed in some studies was caused by a lack of attention to speech sounds, or difficulties in processing linguistic information from speech sounds, Whitehouse and Bishop (2008) set up two processing modes: a passive mode where participants were told to ignore the sounds, and an active mode where participants responded by clicking on the mouse while hearing nonstandard sounds. They found that children with ASD showed diminished ERP responses (indicated by P1-N2-P3-N4-P3a) to speech stimuli but not to nonspeech stimuli in the passive mode, whereas the speech encoding deficits disappeared in the active mode where participants were asked to pay attention to the sound stream. Thus, it appears that while individuals with ASD can process linguistic information in speech when in an active mode as suggested by previous behavioural studies, they do not do so spontaneously when in a passive mode as shown in some of the neurophysiological studies.

Thus, previous behavioural studies that examined pitch processing in both speech and matched music tasks seem to suggest that individuals with ASD process speech and music pitch similarly (Järvinen-Pasley & Heaton, 2007; Mayer et al., 2016), which support the domain-general framework. Nevertheless, Jiang et al. (2015) examined pitch perception using unmatched speech intonation (i.e., disyllabic words in Mandarin) and melodic contour (i.e., five-tone sequences) perception tasks. They found enhanced/intact melodic contour perception but impaired statement-question intonation perception in individuals with ASD. These findings indicate that pitch deficits in ASD only exist in speech but not in music, which is compatible with the domain-specific framework. Therefore, the current existing evidence reported from pitch perceptive studies is inconclusive as to whether speech and music pitch processing are processed similarly, or whether they are processed independently.

1.1.4.2 Pitch production in speech and music

In production studies, ASD is usually linked to atypical speech production (J. J. Diehl & Paul, 2012; Fosnot & Jun, 1999; Hubbard & Trauner, 2007; McCann et al., 2007; McCann & Peppé, 2003; Paul et al., 2008; Peppé et al., 2007, 2011) but normal to superior musical production (Applebaum et al., 1979). Spoken language in individuals with ASD has been described as monotonous, sing-song, pedantic, exaggerated, parrotlike, and robotic (Chan & To, 2016; J. J. Diehl et al., 2009, 2015; J. J. Diehl & Paul, 2013), with pitch being suggested to be the primary contributor to these atypical patterns (DePape et al., 2012). It has been reported that vocal characteristics are the primary contributors of perceived social oddness when interacting with individuals with ASD, even among those without language impairments (Paul et al., 2005; Van Bourgondien & Woods, 1992). For instance, speech with a high pitch can result in an unintended, but negative, impression of domineering force (Shriberg et al., 2001). Thus, misuse of pitch in speech may also play a negative role in the impression others give to individuals with ASD.

Indeed, acoustic studies have commonly identified increased pitch variability in the speech generated by individuals with ASD, whether produced while reading aloud (Fosnot & Jun, 1999), imitating utterances (Fosnot & Jun, 1999), freely describing pictures (Filipe et al., 2014), making up stories (J. J. Diehl et al., 2009), or during a social interaction (e.g., Autism Diagnostic Observation Schedule (ADOS) interview) (Parish-Morris et al., 2016; Sharda et al., 2010). For example, during both spontaneous and structured communication, participants with ASD employed greater pitch range compared with controls (Nadig & Shaw, 2012). Additionally, the speech pronounced by both groups was rated by speech and language pathology students, and the ASD participants received higher scores than controls for abnormal prosody (Nadig & Shaw, 2012). When reading sentences themselves and imitating sentences from others, individuals with ASD also produced a wider pitch range and greater pitch variation than controls (Fosnot & Jun, 1999). Furthermore, in order to determine whether acoustic features of vocal production (including pitch, intensity, and duration) can be a marker of ASD, Fusaroli et al. (2017) conducted a meta-analysis. The results showed that the mean pitch and the pitch range differed significantly between the ASD and control groups, but the effect sizes were small with an approximately 61-64% discriminative accuracy. No acoustic parameters other than pitch were found to be statistically significant. Thus, the misuse of pitch during vocal production appears to be the main contributor to atypical speech identified in ASD.

In contrast, individuals with ASD show enhanced performance in musical production. For example, Applebaum et al. (1979) tested three autistic children with no musical experience, and three age matched controls who had considerable musical experience. Participants were asked to imitate either individual tones or a series of tones, generated by the piano, voice, or synthesizer. The accuracy of imitations of the pitch, rhythm, and duration of the examples was rated by two independent observers and the results revealed that children with ASD performed as well as, or even better, than the controls. However, this is the only study examining musical production in ASD, and more studies with larger sample sizes are needed to substantiate the findings.

While there are only a few studies that have investigated pitch production in speech and music in ASD, the existing findings seem to favour the domain-specific framework, as those with ASD showed atypical speech production (Bonneh et al., 2011; J. J. Diehl & Paul, 2012) but intact/enhanced music production (Applebaum et al., 1979). However, with only one study investigating musical production, one cannot draw well-founded conclusions to inform the theoretical debate regarding music and language production in the ASD population. In addition, these studies have only focused on either the speech or music domain, and no studies have yet directly compared productive abilities in speech versus music in ASD using matched linguistic and musical tasks. Thus, it remains unknown whether impaired speech production but spared/enhanced musical production would be present in the same sample of participants.

1.2 The relationship between perception and production

Both speech and vocal music rely on voices being physically produced and perceived, which are necessary for communication to occur (Hutchins & Moreno, 2013).

Therefore, revealing the relationship between vocal perception and production is crucial for understanding the nature of communication, and for guiding clinical treatment to help those with communication difficulties, as seen in some individuals with ASD (Eigsti et al., 2011), which leads us to the second aim of this thesis. In the literature, three different theoretical models have been put forward to account for the relationship between vocal perception and production, with the motor model and perception-based model predicting an associated relationship between perception and production, and the dual-route model claiming that the two can be independent of each other (Hutchins & Moreno, 2013).

1.2.1 Motor model and Perception-based model

One of the widely known models in the vocal perception literature is the motor model (Liberman, 1996; Liberman & Mattingly, 1985). It proposes that vocal perception involves access to the vocal motor system. It also claims that sounds are initially processed for motor-related features, and then relayed into perception for symbolic representations, suggesting that impaired vocal production may have a negative effect on vocal perception, although this effect is one-way and not vice versa (Hutchins & Moreno, 2013; Liberman, 1996; Liberman & Mattingly, 1985). Neurophysiological studies have provided strong evidence to support the motor model (Galantucci et al., 2006; Möttönen et al., 2005; Pekkola et al., 2006; Schwartz et al., 2012). For example, brain areas that are associated with motor-related features of speech (e.g., primary motor cortices, the left inferior frontal gyrus, etc.) have been activated during speech perception (Möttönen et al., 2005; Pekkola et al., 2006).

Similarly, the perception-based model also posits a close relationship between perception and production, although it proposes a different processing stream from the motor model (R. L. Diehl et al., 2004; Schwartz et al., 2012). Specifically, this model predicts that the symbolic representations of sound are first processed, and then used to produce the intended sounds, which means that vocal perception abilities can influence vocal production, whereas vocal production abilities may not affect perception ability (R. L. Diehl et al., 2004; Hutchins & Moreno, 2013; Schwartz et al., 2012). The model has been typically supported by studies on individuals losing hearing ability, whose speech production has been deleteriously impacted by their impaired perception ability (Busby et al., 1991; C. R. Smith, 1975).

In short, regardless of the proposed differential processing streams, the perception-based model and motor model assume that there exists a very close relationship between vocal production and perception.

1.2.2 Dual-route model

In contrast to the motor model and perception-based model, the dual-route model claims that vocal perception and production should be uncorrelated (Hickok & Poeppel, 2007; Loui, 2015). Instead, it proposes that impairments in either vocal perception or vocal production do not influence one another, as motor-relevant features and symbolic representations are processed via two different, independent pathways. For example, it has been suggested that some individuals with congenital amusia showed intact ability to imitate pairs of pitches but impaired ability to identify the pitch direction (Loui et al., 2008). Also, the dissociation between perception and production can be presented in a reverse direction, with some amusics exhibiting poor pitch production and intact pitch perception (Williamson et al., 2012). In accordance with the dual-route model, these findings suggest that it is possible that perception and production involve distinct pathways, without affecting each other.

1.2.3 The relationship between perception and production in ASD

In ASD, the relationship between vocal perception and production is also controversial. Using standardized language measures to investigate expressive language and receptive language in ASD, McCann et al. (2007) claimed that children with ASD had more problems with expressive than receptive language. However, this is contrary to the earlier finding by Kjelgaard and Tager-Flusberg (2001) who also used standardized language measures that children with ASD showed either the opposite pattern (more difficulty with receptive than expressive language), or equal expressive and receptive skills.

Similarly, contradicting findings have also been reported in the studies using the same speech prosody stimuli applied to both perception and production tasks. Specifically, some research suggests that individuals with ASD are able to perceive prosody in general, but the way they produce prosody is unnatural and atypical (Chevallier et al., 2009; DePape et al., 2012; Filipe et al., 2014; Grossman et al., 2010). For example, Filipe et al. (2014) examined both perception and production of statement-question intonation in children with ASD, and production performance was judged using both perceptual ratings by adult listeners and acoustic measures. They reported that while children with ASD perceived statement-question intonation as accurate as controls, their production of the intonation was judged as unnatural or odd by adult listeners. Also, acoustic measures suggested that children with ASD had greater variability in pitch, including pitch range, mean and median pitch, compared with controls. Thus, these results provide evidence of dissociation between perception and production in ASD, with intact perception but impaired production. Such findings lend support to the dual-route model.

However, other research shows that individuals with ASD have difficulty with both perception and production of prosody (J. J. Diehl & Paul, 2012; Paul et al., 2005, 2008; Shriberg et al., 2001), and these abilities are closely associated (Peppé et al., 2007). Peppé et al. (2007) investigated prosodic perception and production, including statement-question intonation. They found that children with ASD performed worse than controls not only in the production tasks, but also in the perception tasks, and that the perception and production performance were correlated in both groups. These findings clearly do not support the dual-route model, but instead favour the claim of the motor model and perception-based model.

1.3 The processing of low-level versus high-level information

In addition to the controversies relating to domains (speech vs. music) and modalities (perception vs. production), there is also an ongoing theoretical debate that is specific to the ASD population. In particular, when processing hierarchical information, including low-level (e.g., pure tones and single tones) and subsequent high-level information (e.g., prosody in speech and melodies), whether individuals with ASD have intact low-level processing but impaired high-level processing is unclear. Thus, the third aim of this thesis is to inform two theoretical models that have been put forward to address the question, namely Weak Central Coherence (WCC) and Enhanced Perceptual Functioning (EPF).

1.3.1 WCC theory and EPF theory

Originally proposed by Frith (1989), the WCC proposes that individuals with ASD tend to pay more attention to details or low-level information, and as a result, that they show reduced tendency to integrate details into a whole or higher-level information (Happé & Booth, 2008; Happé & Frith, 2006). Thus, the difficulties in integrating multiple low-level information and processing holistically, which require higher-level processing abilities, may be caused by low-level bias in hierarchical processing and superior performance in low-level perceptual operations in ASD.

While agreeing with the WCC that ASD is associated with enhanced or intact low-level information perception, proponents of the EPF argue against the claim that a preference for low-level information leads to a deficit in the perception of high-level information in ASD (Mottron et al., 2006; Mottron & Burack, 2001). Rather, the EPF proposes a more moderate prediction of the ability to process global or higher-level information such that the processing is more of an option for ASD (i.e., they can perceive when they want to or are instructed to) as opposed to typically developing populations where it is mandatory (Haesen et al., 2011; Mottron et al., 2006; Mottron & Burack, 2001).

There is extensive evidence for enhanced or intact low-level auditory processing in ASD as proposed by both the WCC and EPF (Jones et al., 2009; Khalfa et al., 2004; O'Riordan & Passetti, 2006). For example, Khalfa et al. (2004) examined loudness thresholds in individuals with ASD using pure tones, and found comparable perception and reduced tolerance to uncomfortable loudness in the ASD group compared to the control group. However, research findings with regard to higher-level information processing ability in ASD are mixed. Some studies reported impaired high-level information processing, such as processing prosody of utterances (Järvinen-Pasley, Peppé, et al., 2008; J. Jiang et al., 2015; Kujala et al., 2005; McCann et al., 2007; Peppé et al., 2007). For instance, Peppé et al. (2007) assessed a range of prosodic abilities, including affect (liking versus disliking), intonation (statements versus questions), chunking (e.g., "chocolate-biscuits and jam" versus "chocolate, biscuits and jam"), and focus (e.g., "I wanted **red** and blue socks" versus "I wanted red and **blue** socks") in children with ASD using a battery called Profiling Elements of Prosodic Systems in Children (PEPS-C). They found that children with ASD performed worse than controls in 11 out of the 12 prosody tasks across receptive and expressive tasks, revealing a prevalent prosodic impairment in ASD. Also, it has been reported that the ability to perceive speech in noise is impaired in individuals with ASD (Alcántara et al., 2004; Schelinski & von Kriegstein, 2020).

In contrast, other studies found intact abilities to perceive high-level information in individuals with ASD (Grossman et al., 2010; Heikkinen et al., 2010; Jones et al., 2011), including processing musical emotions (e.g., anger, fear, triumph, tenderness and contemplation) (Heaton, Allen, et al., 2008; Heaton, Hermelin, et al., 1999) and structures (Heaton et al., 2007), as well as affective prosody and lexical stress (Grossman et al., 2010). For instance, Grossman et al. (2010) investigated perception of affective prosody (i.e., sad, happy and neural) and lexical stress (e.g., HOTdog vs. hotDOG) in individuals with ASD. The results showed comparable accuracy between the ASD and control groups in both tasks, indicating that individuals with ASD are able to perceive affective prosody and lexical stress. It has been suggested that these inconsistencies in high-level information processing in ASD are likely triggered by different tasks, different stimuli used, as well as intra ASD population differences (O'Connor, 2012). Thus, while both theoretical models, along with existing auditory findings, agree that individuals with ASD tend to focus on low-level information, whether their higher-level information processing capacity is impaired remains open to debate.

1.3.2 Low-level versus high-level processing of pitch in ASD

Pitch conveys both low-level and high-level information, which provides an ideal opportunity to test the competing theories about the abilities to process low-level and high-level information in ASD. In the studies reviewed below, some used the terms "local" and "global" to describe hierarchical information in music, which is similar to the terms of low-level and high-level information (Germain et al., 2019; Haesen et al., 2011).

Consistent with the above reviewed auditory findings, as well as the WCC and EPF theories, individuals with ASD have been described to manifest enhanced or intact low-level pitch processing (Bonnel et al., 2003, 2010; O'Riordan & Passetti, 2006). For example, a study conducted by Bonnel et al. (2003) examined pitch processing ability of individuals with ASD for pure tones using a discrimination task (i.e., same or different) and a categorization task (i.e., low or high). They found that individuals with ASD performed better than controls in both tasks. Similar to the enhanced pitch processing in low-level pure tones, when discriminating pairs of melodies, the tasks that taped local pitch processing (e.g., detecting pitch changes in contour-preserved melodies) also reported superior local pitch processing in individuals with ASD (Mottron et al., 2000).

While enhanced or intact low-level pitch processing has been found in ASD, a deficit in high-level or global pitch processing, as postulated by the WCC theory, is less widely confirmed (Bouvet et al., 2014; Heaton, 2005; Mottron et al., 2000; also see Haesen et al. (2011) and Heaton (2009) for reviews). In particular, Bouvet et al. (2014) investigated both local (i.e., a group of three notes) and global (i.e., 9-note melody/ three groups of local elements) congruency, which can be falling or rising. They reported that while individuals with ASD showed superior local pitch processing, their global pitch processing was also intact.

In a recent study that tested pitch perception in both low and high levels using pitch direction discrimination and melodic global-local tasks, respectively (Germain et al., 2019), the results found that the ASD and control groups performed comparably in both tasks. Thus, the findings suggested that both low-level and higher-level pitch processing were intact in ASD, which is in favour of the EPF theory. Also, in contrast to what has been proposed by the WCC, the study also suggested that low-level processing can predict higher-level perception in ASD and that the increased coupling between low- and high-level pitch processing may reflect a bottom-up cascade, which means that the extent of impairment or improvement in low-level processing impacts the ability to process higher-level signals at later stages (Germain et al., 2019; Stevenson et al., 2014). Thus, the currently existing findings regarding pitch processing across levels, at least from the music domain, appear to favor the EPF theory over the WCC.

1.4 Developmental changes of pitch processing in ASD

Development is a crucial factor that can affect pitch processing. Whilst pitch processing has been widely investigated across domains (speech vs. music), modalities (perception vs. production) and processing levels (low-level vs. high-level) in individuals with ASD, little information is available about the developmental changes of pitch processing in these aspects in ASD. Mayer et al. (2016) examined pitch perception in adults with and without ASD, and compared them with groups of children and adolescents in previous studies that used the same paradigm. The results suggested that pitch sensitivity increased with age in controls, but not in individuals with ASD. In particular, pitch processing ability was enhanced in childhood in ASD compared with controls, whereas no increase was observed in ASD throughout development. Interestingly, the developmental trajectories of pitch processing in speech and music were similar across both ASD and control groups, indicating shared mechanisms between speech and musical pitch processing in both groups.

However, whether the different developmental trajectories between individuals with ASD and controls observed by Mayer et al. (2016) apply to higher-level pitch-related prosody remains unresolved. If the answer is yes, individuals with ASD would be expected to show no obvious or slight developmental increase in pitch-related prosodic perception since their pitch processing ability is stable over time. By contrast, controls would show improvement with age. However, this prediction contradicts the finding of no developmental improvement in prosody performance, including statement-question intonation, in controls and in those ASD participants with normal language ability, but an improvement with age in ASD participants with low language ability (Lyons et al., 2014). If the answer is no, the dissociation between pitch processing and the perception of pitch-related prosody would challenge previous findings claiming low-level pitch perception can predict higher-levels of pitch processing (Germain et al., 2019).

Hence, it is essential to map the developmental changes of pitch processing and those prosody that rely on pitch cues separately, as well as to examine the relation between abilities of pitch and prosody perception across the lifespan. Notably, although Lyons et al. (2014) reported developmental changes in prosodic performance of preadolescents (9–12 years old) and adolescents (13–17 years old) with ASD, age related changes in prosody across the lifespan from children, adolescents to adults are yet to be examined. Previous studies suggested that while typically developing children continue to develop understanding and using of prosody during school years (Cruttenden, 1985; Wells et al., 2004), they did not reach adult-like levels (Arciuli & Ballard, 2017; Ballard et al., 2012).

Furthermore, based on the findings reported by previous observational studies on vocal production, the early presentation of atypical vocal characteristics in individuals with ASD, such as deficits in the use of pitch, tends to be persistent and has no qualitative improvement over time, even when other aspects of language have improved (DeMyer et al., 1973; Kanner, 1971; McCann & Peppé, 2003; Simmons & Baltaxe, 1975). However, since no vocal studies have used acoustic measures to map developmental changes in ASD, it remains unclear whether the observations over time can be supported by acoustic measures adopting the same standards and methods across the lifespan.

1.5 This thesis

The literature discussed above has shown that although pitch processing in individuals with ASD has been extensively investigated, in order to inform many ongoing theoretical debates, e.g., whether speech and music share underlying mechanisms in this population (the domain-specific vs. domain-general debate), whether high-level pitch processing is intact in ASD (the WCC vs. EPF debate), and whether perception and production are correlated (the perception-based model and motor model vs. dual-route model debate), more studies using well-matched comparative designs are required.

Comparative analysis allows us to examine theories across multiple tasks to evaluate their applicability, and these findings can contribute to the generalization of theories regardless of whether the theoretical debates are specific to the ASD population or beyond. It is known that ASD is a heterogeneous group, characterized by high intra variability (Hall et al., 2012; Vivanti et al., 2018). Using comparative analysis is, therefore, particularly important in ASD because it might be helpful in teasing apart the mixed findings across studies that solely focus on one aspect (e.g., studies focused on either speech or music). For the benefit of individuals with ASD, comparative analysis can offer implications for using a range of alternative options to identify or overcome possible difficulties existing in a similar situation. These include using musical therapy to improve their language comprehension, conducting training on low-level information processing to advance their high-level processing abilities, and practicing vocal imitation to improve their understanding of language.

While a scarce number of studies have investigated how individuals with ASD perceive pitch under both speech and music conditions (Cheng et al., 2017; Heaton, Hudry, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007; Mayer et al., 2016), no studies have examined pitch production comparatively across these two domains. In addition, among those comparative studies on speech and music pitch perception in ASD, the speech stimuli used (e.g., vowels or syllables) were relatively simple and did not involve prosodic cues. Since prosody is a landmark deficit domain of ASD, whether individuals with ASD show impairments in speech prosody but not in musical analogues is little known. Exploring pitch perception using prosody, such as statement-question intonation, in speech and music, would not only inform the ongoing domain-specific versus domain-general debate, but also illustrate the possibility of using music therapy to help with speech—an area where individuals with ASD are commonly reported to have problems.

Several comparative studies have examined prosody perception and production in ASD. However, the production performance in these studies relied on subjective perceptual ratings, rather than acoustic analyses, which have made it difficult to identify what aspects of production are atypical in ASD (e.g., pitch, duration, or intensity) (Grossman et al., 2010). Also, given the bias of subjective ratings observed in previous studies (Van Santen et al., 2010), further comparative research is warranted to examine the relationship between vocal perception and production using objective acoustic analyses to quantitatively inform the debate over whether there is a correlation between perception and production. Clarifying the relationship between perception and production is particularly important for ASD, since it has implications for conducting training on perception, which could potentially have a positive influence on language comprehension as well as language production.

Concerning pitch processing in low and high levels, a recent study has comprehensively examined how individuals with ASD perceive pitch across levels, suggesting pitch processing was intact in both levels (Germain et al., 2019). However, whether these findings of unimpaired pitch processing in high-level melodies can be replicated in the speech domain, is not yet clear. Thus, comparative studies are warranted to investigate how low-level and high-level speech pitch information is processed, as well as to test the relationship between the two in ASD. The findings of intact pitch processing ability in ASD across levels suggested by Germain et al. (2019), especially with regard to the association in pitch processing between low- and higher-level processes, is encouraging. If the association between low-level and high-level perception also holds true in the speech domain, it would imply that training on low-level pitch would positively affect higher-level speech pitch processing, which in turn would be helpful for social interactions and communication (Xu, 2005).

In comparative studies, in addition to cross-section designs (e.g., across-domains, across-levels, etc.), a developmental dimension should also be considered to address the question of whether the processing abilities in ASD grow over time. Adding a developmental dimension helps to fully understand these developmental changes and helps clarify the role age plays in pitch processing across domains, modalities and processing levels in individuals with ASD.

This thesis reports three studies that aim to test the theoretical models and address those aforementioned issues through comparative and developmental designs. Study 1 investigates pitch perception in individuals with and without ASD across domains (statement-question intonation vs. musical analogue), modalities (identification vs. imitation) and processing levels (low-level pitch direction discrimination versus high-level processing of statement-question intonation), as well as the age-related changes across different age cohorts (children, adolescents and adults), using standard methods (discrimination, identification and imitation tasks). Study 2 examines pitch production in individuals with and without ASD when imitating stimuli across domains (statement-question intonation in spoken versus sung versions) and the developmental changes across age cohorts (children versus adults). Study 3 also investigates pitch perception in children with and without ASD across domains (question intonation and musical analogue). Unlike Study 1, Study 3 adopts a novel data-driven method called reverse-correlation paradigm to systematically reveal the mental representation of pitch contours in speech and music.

To briefly summarise the results, Studies 1 and 3, while using different methods, consistently revealed that pitch perception ability is intact in individuals with ASD in both speech and music conditions. The acoustic measures used in Study 2 suggested that individuals with ASD are impaired in absolute but not relative pitch imitation in both speech and song. Regarding the high-level pitch processing tested in Study 1, individuals with ASD performed comparably as controls, revealing that the ability to perceive high-level pitch is not impaired in individuals with ASD. In addition, performance on high-level pitch processing can be predicted by the sensitivity to low-level pitch, and such predictions exist in both the ASD and control groups. Study 1 also suggested that the ability to identify statement-question intonation is associated with the ability to imitate the intonation in both the ASD and control groups. Furthermore, Study 1 found that, similar to the controls, performance on pitch perception of statement-question intonation and its musical analogue, as well as low-level pitch direction discrimination increased with age. However, Study 2 failed to observe an increase in pitch production over time. Instead, children imitated spoken pitch more accurately than did adults.

In summary, I argue that the findings from the three studies of associated pitch perception and production abilities in speech and music, whether intact or impaired, are in support of the domain-general framework, in which the mechanisms of speech and music pitch information processing are shared. Results of correlated pitch perception and production are compatible with the perception-based model and motor model, indicating that the more accurate the perception and understanding, the better the performance on production, or vice versa. In terms of high-level pitch processing, the findings favour the Enhanced Perceptual Functioning theory, which proposes that the ability to process high-level information may be unimpaired in ASD. Therefore, these findings provide evidence for closely linked relationships in pitch processing across different auditory domains (speech vs. music), modalities (perception vs. production), and processing levels (low-level vs. high-level) in individuals with ASD. The following three chapters report these three studies in more detail.

Chapter 2. Study 1: Perception and production of statement-question intonation in autism spectrum disorder: A developmental investigation

Abstract

Prosody or "melody in speech" in autism spectrum disorder (ASD) is often perceived as atypical. This study examined perception and production of statements and questions in 84 children, adolescents and adults with and without ASD, as well as participants' pitch direction discrimination thresholds. The results suggested that the abilities to discriminate (in both speech and music conditions), identify, and imitate statement-question intonation are intact in individuals with ASD across age cohorts. Sensitivity to pitch direction predicted performance on intonation processing in both groups, who also exhibited similar developmental changes. These findings provide evidence for shared mechanisms in pitch processing between speech and music, as well as associations between low- and high-level pitch processing and between perception and production of pitch.

Keywords: autism spectrum disorder, speech, music, intonation, pitch

2.1 Introduction

Prosody is a suprasegmental feature of speech that adds additional pragmatic, affective, or grammatical information via changes in frequency, intensity, and duration of spoken utterances (McAlpine et al., 2014; McCann & Peppé, 2003; Paul et al., 2005). It plays an important role in speech communication and social interaction (Xu, 2005). While the acquisition of prosody starts from infancy (Levitt, 1993) and lays the foundations for children's sociopragmatic development (Hübscher & Prieto, 2019), atypical prosody can become a barrier to everyday linguistic and social functioning, as seen in autism spectrum disorder (ASD) (McCann & Peppé, 2003; Paul et al., 2005).

ASD is a neurodevelopmental disorder associated with deficits in social communication and interaction as well as restricted and repetitive behaviours and interests (American Psychiatric Association, 2013). Prosodic deficits have been frequently observed in ASD across a variety of perception and production tasks (J. J. Diehl & Paul, 2012; Nakai et al., 2014; Peppé et al., 2007; Shriberg et al., 2011; Tager-Flusberg et al., 2005). They can occur even in highly verbal individuals with ASD and tend to be lifelong even when other areas of language, such as semantics and syntax, improve (McCann & Peppé, 2003). Among the different areas of prosody, recognising and differentiating the rising from falling intonation in questions and statements represents an important aspect of conversational and linguistic competence (Dahan, 2015; Xie et al., 2021), and the literature in ASD has produced mixed findings (Chevallier et al., 2009; Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; Jiang et al., 2015; McCann et al., 2007; McCann & Peppé, 2003; Paul et al., 2005; Peppé et al., 2007). The current study investigated this issue by examining the roles of response bias, stimulus type, age, and pitch discrimination thresholds in the perception and production of statement-question intonation in ASD.

Perception of statement-question intonation and response bias in ASD

Several studies used the same test battery, PEPS-C (Profiling Elements of Prosodic Systems - Children) (Peppé & McCann, 2003), to examine discrimination (e.g., same vs. different) and identification (e.g., question vs. statement) of statements and questions in ASD (Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; McCann et al., 2007; Peppé et al., 2007). Within this battery, statement-question identification is assessed using a turn-end task with single words, e.g., “Carrots.” vs. “Carrots?”. Statement-question discrimination is assessed within a short-item discrimination task, which contains the laryngographic sounds (devoid of meaning) of the statement-question pairs, as well as those of the liking-disliking pairs, e.g., “tea” pronounced as though the speakers like it or dislike it, from the affect subtask in PEPS-C. Thus, the identification and discrimination tasks are unmatched in stimulus type (speech vs. laryngographic sounds) and in the number of relevant stimuli (only statements or questions are used in the identification task, whereas both statement-question and liking-disliking pairs are included in the discrimination task) in these studies (Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; McCann et al., 2007; Peppé et al., 2007). Results from these studies suggest that individuals with ASD are unimpaired in statement-question identification compared to typically developing (TD) peers (Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; McCann et al., 2007; Peppé et al., 2007). However, impaired discrimination between statements and questions was observed in one sample of participants (31 ASD vs. 72 TD participants) (McCann et al., 2007; Peppé et al., 2007), while a different sample showed intact discrimination (21

ASD vs. 21 TD participants) (Järvinen-Pasley, Peppé, et al., 2008). In summary, studies using PEPS-C suggest intact statement-question identification but the results on statement-question discrimination in ASD are unclear, in part due to limitations of the design, but also mixed results from different studies (Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; McCann et al., 2007; Peppé et al., 2007).

Using prosodic tasks other than PEPS-C, (e.g., sentence stimuli from Patel et al. (1998)), previous studies also reported intact statement-question identification beyond single-word stimuli in ASD (Chevallier et al., 2009; Järvinen-Pasley, Peppé, et al., 2008; Paul et al., 2005). However, using disyllabic phrases from Jiang et al., (2010), Jiang et al. (2015) revealed impaired identification and discrimination of statement-question intonation in Mandarin speakers with ASD. While the different results between Jiang et al. (2015) and other studies (Chevallier et al., 2009; Järvinen-Pasley, Peppé, et al., 2008; Paul et al., 2005) may be attributed to the different language or cultural background of the participants: Mandarin Chinese versus English, it may also be the case that the discrepancy was due to differences in task difficulty across these studies. Indeed, participants from both the ASD and TD groups performed at ceiling in Paul et al. (2005), which used the stimuli from Patel et al. (1998). In that stimulus set large pitch contrasts exist between the statements and questions (Patel et al., 1998), and research has shown that even individuals with congenital amusia, a neurodevelopmental disorder of pitch processing, can perform as well as TD controls on both identification and discrimination of these statements/questions (Ayotte et al., 2002; A. D. Patel et al., 2008). Addressing the issue with ceiling performance in the literature, Liu et al. (2010) designed and created a new set of ecologically valid stimuli with relatively subtle pitch contrasts between statements and questions, and revealed prosodic deficits in congenital amusia. Thus, using stimuli from Liu et al. (2010), the current study aimed to examine whether English-speaking individuals with ASD would also show impaired statement-question identification and discrimination when task difficulty is increased.

In addition to identification/discrimination accuracy rates, it has been suggested that participants' response patterns should also be scrutinised in order to detect possible response biases in ASD (Järvinen-Pasley, Peppé, et al., 2008). Specifically, Peppé et al. (2007) observed that while children with ASD performed as well as controls in terms of judgement accuracy in statement-question identification, they were biased towards judging questions as statements. In this study, 12.9% of the ASD participants and 2.7% of the control participants judged all the questions as statements, showing a declarative bias, although this percentage difference did not reach statistical significance (Peppé et al., 2007). For discrimination, impaired performance in ASD was mainly driven by false alarms, i.e., judging the same items as different (Peppé et al., 2007). To investigate the declarative bias in ASD further, Järvinen-Pasley, Peppé, et al. (2008) examined another sample of participants and included the identification task in Patel et al. (1998) in addition to the turn-end task in PEPS-C. While no significant group difference in response patterns was observed for the turn-end task in PEPS-C, a significant declarative bias was observed among 50% of participants with ASD (in comparison to 10% of controls) for the identification task from Patel et al. (1998) (Järvinen-Pasley, Peppé, et al., 2008). However, no response bias emerged among Mandarin speakers with ASD for either identification or discrimination in Jiang et al. (2015), although significantly lower accuracy rates were observed in ASD compared to TD. Thus, among the studies that examined response biases in ASD, mixed findings have been presented, with some studies indicating a response bias based on either statistics or simply percentage comparison (Järvinen-Pasley, Peppé, et al., 2008; McCann et al., 2007; Peppé et al., 2007), whereas others reporting no response bias (Jiang et al., 2015), depending on the tasks and samples.

In summary, despite much research (Chevallier et al., 2009; Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; Jiang et al., 2015; McCann et al., 2007; McCann & Peppé, 2003; Paul et al., 2005; Peppé et al., 2007), it remains unclear whether individuals with ASD are associated with deficits in identification and/or discrimination of statements and questions, and whether there are response biases driving the observed accuracy rates. These questions need to be addressed, as the answers have implications for the prosody phenotypes of ASD. As mentioned earlier, due to the limitations of the design in PEPS-C (Peppé & McCann, 2003), the short-item discrimination task contains not only statement-question pairs but also liking-disliking pairs, and in laryngographic sounds rather than in natural speech. Thus, one cannot make inferences about the ability to discriminate statements from questions in everyday language from this task by individuals with ASD, or any other neurodevelopmental disorder. However, if it is indeed the case that ASD is associated with intact identification but impaired discrimination as reported in Peppé et al. (2007), this dissociation between identification and discrimination may be interpreted as a special feature related to ASD phenotypes (Peppé et al., 2007). An association between identification and discrimination has been observed in other studies: both are intact (Järvinen-Pasley, Peppé, et al., 2008); or both impaired (Jiang et al., 2015). To further clarify this issue and to help understand the phenotypes of ASD, the current study employed both identification and discrimination tasks from Liu et al. (2010) to investigate response patterns and the relationship between statement-question identification and discrimination in ASD.

Production of statement-question intonation in ASD

In contrast to the mixed findings reported in perception studies, evidence from production studies has consistently suggested atypical intonation production in ASD (Filipe et al., 2014; Fusaroli et al., 2017; McCann et al., 2007; McCann & Peppé, 2003). Specifically, statement responses of individuals with ASD were more likely to be judged as questions or ambiguous than those of controls (McCann et al., 2007; Peppé et al., 2007). In addition, utterances by individuals with ASD were much less likely to be judged as normal or natural than those of controls (Filipe et al., 2014). These ratings were either given by the experimenter (“tester”) (McCann et al., 2007; Peppé et al., 2007) or by independent adult participants (Filipe et al., 2014). Although informative, subjective ratings do not reveal what aspects of intonation production were atypical in ASD (e.g., pitch, duration, and intensity). In studies using objective acoustic measures, individuals with ASD showed significantly greater pitch range, mean pitch, and maximum pitch than controls for both statements and questions (Filipe et al., 2014), and increased and inappropriate use of pitch accents as well as difficulty in producing high frequency boundary tones (Fosnot & Jun, 1999). These findings were supported by Fusaroli et al. (2017), who systematically reviewed the literature quantifying acoustic patterns in ASD and identified significant differences in pitch production (e.g., pitch range and mean pitch) between individuals with ASD and controls, while finding no significant differences in other acoustic features (e.g., intensity, duration).

In sum, the atypical production of intonation in ASD seems to be related to one parameter—pitch (DePape et al., 2012; Fusaroli et al., 2017). It is known that question and statement intonation are heavily dependent upon pitch direction, with rising tones representing questions and falling tones representing statements (Cruttenden, 1997; Lieberman, 1960). Therefore, misuse of pitch itself can cause not only atypical intonation production but also misperception of statements and questions. Studies on congenital amusia suggest that impaired identification or discrimination of pitch direction can be coupled with intact imitation (Hutchins & Peretz, 2012; Loui et al., 2008). Like controls, individuals with congenital amusia also performed better on imitation than identification of statements and questions (Liu et al., 2010). Previous studies on intonation production in ASD (Filipe et al., 2014; Fosnot & Jun, 1999; McCann et al., 2007; Peppé et al., 2007) have not conducted acoustic analysis to verify the acoustic realisation of pitch direction in statements and questions in ASD. Thus, it remains to be determined whether intonation production and perception abilities are related or dissociated among individuals with ASD. The current study addressed this issue by including an intonation imitation task and using acoustic measures to assess pitch direction of the final words in the produced statements and questions (Liu et al., 2010).

Perception of pitch in speech versus music in ASD

As in speech, pitch is also used extensively in music to convey meaning and emotion (A. D. Patel, 2010). It has been intensely debated whether pitch processing is domain-specific or domain-general between speech and music domains (Mantell & Pfordresher, 2013; A. D. Patel, 2010; Peretz & Coltheart, 2003). In particular, Peretz and Coltheart (2003) proposed that pitch information within a musical context is processed by a tonal encoding module which is absent in spoken pitch processing. Other researchers, however, argued for shared systems underlying the processing of information across both domains (Koelsch, 2011; Koelsch & Siebel, 2005; A. D. Patel, 2010; Sammler et al., 2009). Comparing intonation perception with melodic contour perception, Jiang et al. (2015) observed enhanced/intact melodic contour identification/discrimination but impaired statement-question identification and discrimination in Mandarin speakers with ASD. This finding suggested pitch processing deficits specific to the speech domain in ASD (Jiang et al., 2015). However, other studies indicated enhanced identification of pitch contours (e.g., rising, falling, falling-rising, rising-falling) across speech and musical stimuli (Järvinen-Pasley, Wallace, et al., 2008), as well as superior discrimination of pitch patterns across speech-speech and speech-music stimulus pairs in ASD versus TD (Järvinen-Pasley & Heaton, 2007). Therefore, further research is warranted to clarify the domain specificity or generality of pitch processing in ASD. To our knowledge, no studies have yet compared pitch perception in ASD using speech and musical stimuli that are matched in global pitch contours derived from statement-question intonation. The present study aimed to fill this gap by investigating whether individuals with ASD would process intonation embedded in speech and musical stimuli differently, using the musical analogues of the statement-question discrimination task in Liu et al. (2010).

The development of prosodic abilities and its relationship with pitch sensitivity

Studies of prosodic development in TD children suggest that there are significant improvements in the perception and production of statement-question intonation between ages 5 and 11 (Wells et al., 2004). As children grow older, pitch becomes the primary cue for the statement-question contrast compared to intensity and duration in production (R. Patel & Grigos, 2006). While 4-year-olds used lengthened duration of the final syllable rather than a rising pitch contour to signify questions, 7-year-olds used multiple acoustic cues (including pitch, intensity and duration) and 11-year-olds used pitch cues predominantly to differentiate statements from questions (R. Patel & Grigos, 2006). Given that language delay and impairment are prevalent among children and youth with ASD (Kwok et al., 2015), it may be the case that the development of prosodic skills is also delayed in ASD. Lyons et al. (2014) investigated the developmental changes of four prosodic functions, including the perception and production of statement-question intonation, stress, phrasing, and affect, in “language-normal” and “language-impaired” preadolescents (9-12 years old) and adolescents (13-17 years old) with and without ASD. The results suggest that TD preadolescents performed as well as TD adolescents on statement-question identification and production, and thus no developmental improvement was observed among TD participants due to ceiling performance. The same pattern of results was also seen in “language-normal” ASD preadolescents and adolescents, who performed similarly to the TD groups on both identification and production of statements and questions. For the “language-impaired” ASD groups, however, significant age-related improvement was observed for identification, but not for production, of statements and questions. That is, while impaired statement-question identification was only observed among “language-impaired” ASD preadolescents, but not among adolescents, impaired statement-question production persisted among “language-impaired” ASD preadolescents and adolescents. Thus, there are developmental delays in the perception and production of statements and questions among “language-impaired” individuals with ASD (Lyons et al., 2014).

In addition to the close relationship with language abilities (Lyons et al., 2014), prosodic skills also correlate significantly with pitch processing abilities (Liu et al., 2010, 2012; Vuvan et al., 2015). In typical development, there are age-related improvements in the ability to discriminate the direction of pitch changes between ages 6-11 (Fancourt et al., 2013). However, it has been reported that individuals with ASD show enhanced pitch discrimination early in development, and this ability maintains across children, adolescents and adults and does not correlate with receptive vocabulary (Mayer et al., 2016). By contrast, controls show significant gains in pitch discrimination performance across development, which also correlates significantly with receptive vocabulary scores (Mayer et al., 2016). This raises the questions as to whether and how pitch processing abilities influence intonation perception and production in individuals with ASD, and if age plays a role in these abilities across the lifespan. The current study addressed these questions by examining the development of statement-question perception and production across children, adolescents and adults with and without ASD, as well as its relationship with pitch direction discrimination thresholds. Implications of the results of our studies for pitch processing (in language and in music) generally and for the correct interpretation of prosodic perception and production in ASD specifically, and possible directions for intervention where prosody is observed to be impaired, are considered in the Discussion.

Present study

In the current study, we matched ASD and TD children, adolescents and adults for age, gender and cognitive ability and incorporated both perception and production as well as pitch threshold tasks from Liu et al. (2010). Focusing on the prosodic feature of statement-question intonation and the acoustic parameter of pitch, we examined intonation processing in ASD and TD from the perspectives of task condition (discrimination, identification, imitation), response bias, stimulus type (speech, music), developmental changes, and its association with pitch thresholds. We asked whether individuals with ASD differed from controls in their ability to discriminate, identify, and imitate statement-question intonation, whether individuals with ASD showed response bias in discrimination and identification tasks, and whether performance on intonation perception and production related to pitch direction discrimination thresholds. We also examined whether individuals with ASD would perform better on musical pitch processing than on linguistic pitch processing, comparing discrimination of natural speech and their musical analogues. Finally, we examined the effect of age on pitch and intonation perception and production for both ASD and control groups. Based on previous findings, we predicted that: (a) participants with ASD would show impaired performance compared to controls in intonation discrimination and identification tasks, and they would show response biases towards judging the same pairs as different and identifying questions as statements; (b) participants with ASD would show poorer performance on the imitation task compared with controls; (c) participants with ASD would perform better on the musical condition than the speech condition in the discrimination task; (d) across both groups, performance on intonation processing would be associated with pitch direction discrimination thresholds; and (e) participants with ASD would show different developmental trajectories for pitch and intonation processing compared with controls.

2.2 Methods

Participants

A priori power analysis was conducted using G*Power (Faul et al., 2009). To detect the interaction of Group (ASD vs. control) by Condition (speech vs. music or identification vs. imitation) by Age (child, adolescent vs. adult) in the present design, 64 participants (with 32 in each group) were required to reach a power of 0.80, with a large effect size (f = 0.40) and an alpha of 0.05. Given the mixed findings in the ASD literature and to further increase the power of our study, we recruited a total of 84 participants, 42 with ASD (12 female, 30 male) and 42 controls (12 female, 30 male), resulting in a power of 0.91.

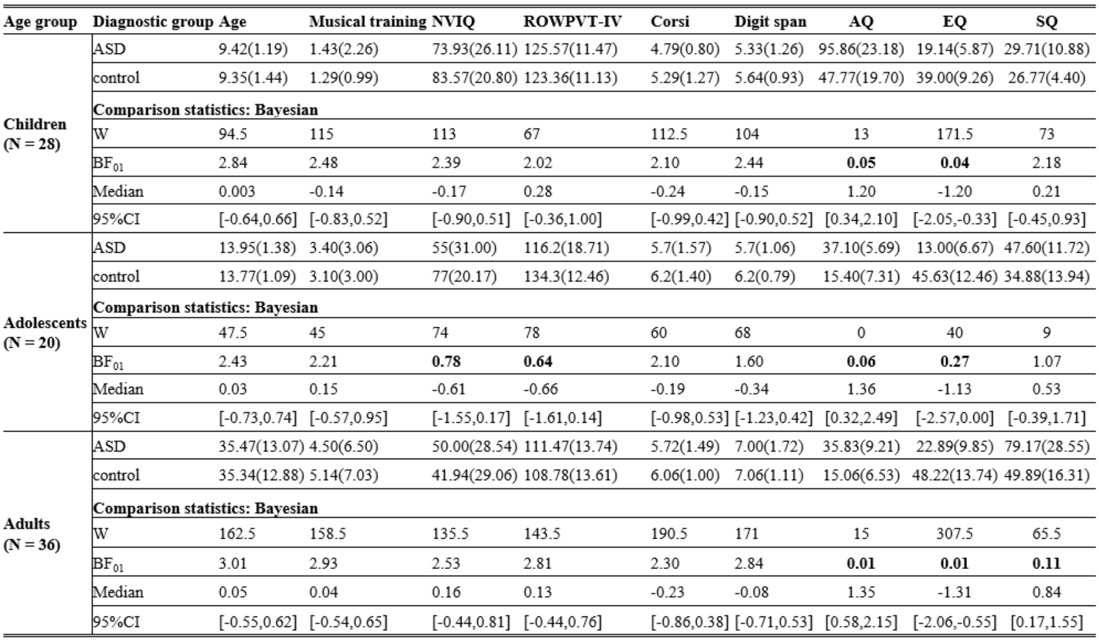
All participants were native speakers of British English, recruited through email lists, word of mouth, online and social media advertisements, local schools, charities and organisations, and departmental participant databases. Participants in the ASD group all received a formal diagnosis of ASD by professional clinicians, and their high autistic traits were also confirmed using the cut-off scores of 32 (adults), 30 (adolescents) and 76 (children) on the Autism Spectrum Quotient (AQ) (Auyeung et al., 2008; Baron-Cohen et al., 2001, 2006). All control participants scored below these cut-offs. In addition to the AQ, Empathy Quotient (EQ) and Systemizing Quotient (SQ) were also collected through questionnaires. All participants had normal hearing in both ears, with pure-tone air conduction thresholds of 25 dB HL or better at frequencies of 0.5, 1, 2, and 4 kHz. The study was approved by the University of Reading Research Ethics Committee. Written informed consent/assent was obtained from the participants and/or their parents prior to the experiment.

Given the significant effects of IQ, receptive vocabulary, short-term memory, and musical training on pitch and prosodic processing (Acton & Schroeder, 2001;

Bidelman et al., 2013; Chowdhury et al., 2017; Heaton, Hudry, et al., 2008; Mayer et al., 2016; McCann et al., 2007; Peppé et al., 2007; Tillmann et al., 2016), we gathered related background measures from all participants (Table 1). Specifically, participants completed a nonverbal IQ test using the Raven’s Standard Progressive Matrices Test (Raven et al., 1998) and a receptive vocabulary test using the Receptive One Word Picture Vocabulary Test IV (ROWPVT-IV) (Martin & Brownell, 2011). The Corsi block-tapping task was used to assess participants’ nonverbal short-term memory span (Kessels et al., 2000), and the digit span task was used to assess verbal short-term memory (Wechsler, 2003). Participants’ musical training background was collected using a questionnaire, and their years of formal musical training were summed across all instruments including voice (Pfordresher & Halpern, 2013).

Following the age cut-offs for the Autism-Spectrum Quotient (Auyeung et al., 2008; Baron-Cohen et al., 2001, 2006), participants were divided into three age cohorts: children (7-11 years), adolescents (12-15 years), and adults (>= 16 years). The age range of the child cohort was between 7.39 to 11.92 years, that of the adolescent cohort was between 12.08 to 15.75, and that of the adult cohort was between 18 to 55.72 years. Table 1 shows the characteristics of the participants.

Table 1. Characteristics of the ASD (N = 42) and control groups (N = 42).



Age group	Diagnostic group	Age	Musical training	NVIQ	ROWPVT-IV	Corsi	Digit span	AQ	EQ	SQ
Children (N = 28)	ASD	9.42(1.19)	1.43(2.26)	73.93(26.11)	125.57(11.47)	4.79(0.80)	5.33(1.26)	95.86(23.18)	19.14(5.87)	29.71(10.88)
	control	9.35(1.44)	1.29(0.99)	83.57(20.80)	123.36(11.13)	5.29(1.27)	5.64(0.93)	47.77(19.70)	39.00(9.26)	26.77(4.40)
	Comparison statistics: Bayesian									
	W	94.5	115	113	67	112.5	104	13	171.5	73
	BF_{01}	2.84	2.48	2.39	2.02	2.10	2.44	**0.05**	**0.04**	2.18
	Median	0.003	-0.14	-0.17	0.28	-0.24	-0.15	1.20	-1.20	0.21
	95%CI	[-0.64,0.66]	[-0.83,0.52]	[-0.90,0.51]	[-0.36,1.00]	[-0.99,0.42]	[-0.90,0.52]	[0.34,2.10]	[-2.05,-0.33]	[-0.45,0.93]
Adolescents (N = 20)	ASD	13.95(1.38)	3.40(3.06)	55(31.00)	116.2(18.71)	5.7(1.57)	5.7(1.06)	37.10(5.69)	13.00(6.67)	47.60(11.72)
	control	13.77(1.09)	3.10(3.00)	77(20.17)	134.3(12.46)	6.2(1.40)	6.2(0.79)	15.40(7.31)	45.63(12.46)	34.88(13.94)
	Comparison statistics: Bayesian									
	W	47.5	45	74	78	60	68	0	40	9
	BF_{01}	2.43	2.21	**0.78**	**0.64**	2.10	1.60	**0.06**	**0.27**	1.07
	Median	0.03	0.15	-0.61	-0.66	-0.19	-0.34	1.36	-1.13	0.53
	95%CI	[-0.73,0.74]	[-0.57,0.95]	[-1.55,0.17]	[-1.61,0.14]	[-0.98,0.53]	[-1.23,0.42]	[0.32,2.49]	[-2.57,0.00]	[-0.39,1.71]
Adults (N = 36)	ASD	35.47(13.07)	4.50(6.50)	50.00(28.54)	111.47(13.74)	5.72(1.49)	7.00(1.72)	35.83(9.21)	22.89(9.85)	79.17(28.55)
	control	35.34(12.88)	5.14(7.03)	41.94(29.06)	108.78(13.61)	6.06(1.00)	7.06(1.11)	15.06(6.53)	48.22(13.74)	49.89(16.31)
	Comparison statistics: Bayesian									
	W	162.5	158.5	135.5	143.5	190.5	171	15	307.5	65.5
	BF_{01}	3.01	2.93	2.53	2.81	2.30	2.84	**0.01**	**0.01**	**0.11**
	Median	0.05	0.04	0.16	0.13	-0.23	-0.08	1.35	-1.31	0.84
	95%CI	[-0.55,0.62]	[-0.54,0.65]	[-0.44,0.81]	[-0.44,0.76]	[-0.86,0.38]	[-0.71,0.53]	[0.58,2.15]	[-2.06,-0.55]	[0.17,1.55]

Note: Age and Musical training are in years; NVIQ and ROWPVT-IV are percentile points of nonverbal IQ and standard scores of receptive verbal ability respectively; Corsi and Digit span are the raw scores of nonverbal and verbal short-term memory respectively; AQ, EQ and SQ are the scores of Autism Spectrum, Empathy and Systemizing Quotient respectively. Bayes factors from a default prior 2-tailed Bayesian Mann-Whitney-Wilcoxon Test are expressed in terms of the Bayes factor in favour of the null hypothesis of no difference (BF01). The delta effect size in these Bayesian comparisons is given by the median of a posterior distribution and 95% credible intervals.

The groups in each of the three age cohorts were largely matched on the background measures, with the exception that the ASD adolescents showed lower receptive vocabulary and nonverbal IQ scores than the control adolescents. To control for the possible contribution of receptive verbal ability and nonverbal IQ to the current results, these scores were entered as covariates in the analysis of Bayesian ANCOVA in the Results section.

Tasks

The present study consisted of one pitch direction discrimination thresholds task and four intonation perception/production tasks from Liu et al. (2010). The pure tones in the pitch direction discrimination thresholds task were generated in Matlab (MATLAB, 2010). Stimuli from the intonation tasks were recorded or generated using Praat (Boersma & Weenink, 2001), with 44.1 kHz sampling rate and 16-bit amplitude resolution.

Pitch direction discrimination thresholds task

Thresholds were measured using a three-interval two-alternative forced-choice procedure with Matlab. In each trial, participants were presented with three pure tones of 600ms each. All three tones contained pitch glides centring on 500 Hz, with two moving in the same direction and the other moving in the opposite direction. Participants were required to identify the “odd-one-out” among the three tones, which always appeared in the first or last position. An adaptive-tracking procedure with a “two down, one up” staircase method and a variable change in step size was used. Starting with a default excursion of six semitones, the initial step size was one semitone which reduced to 0.1 semitones after four reversals and 0.02 semitones after eight reversals. The task was terminated after 14 reversals. The threshold was calculated as the mean excursion size of the target glide of the last six reversals.

The practice session consisted of four trials, with excursion sizes greater than those in the testing session. To ensure that all participants understood the task, they were required to achieve 100% correct on the four practice trials (with feedback) before proceeding to the testing session. Given that inattention may impact performance on adaptive-tracking pitch thresholds tasks, especially in children (Fancourt et al., 2013; Horváth et al., 2009; McDermott & Oxenham, 2008), participants were required to make their responses orally for the experimenters to input into the computer, in order to maintain their attention.

Intonation tasks

The intonation tasks consisted of four subtests assessing discrimination of statements and questions in natural speech and in their musical analogues (composed of gliding tones), and identification and imitation of these statements and questions. The tasks were presented in counterbalanced order across participants using Praat. Taken from Liu et al. (2010), the speech stimuli were 18 statement-question pairs recorded by a 28-year-old female native British English speaker with a slight London accent. The pairs were cross-spliced so that each pair began with the same stem and differed only in the final word. As shown in the examples in Figure 1, the statement had a downward glide and the question had an upward glide, to signify the statement vs. question intonation (For full details of the speech stimuli, see Table 2).

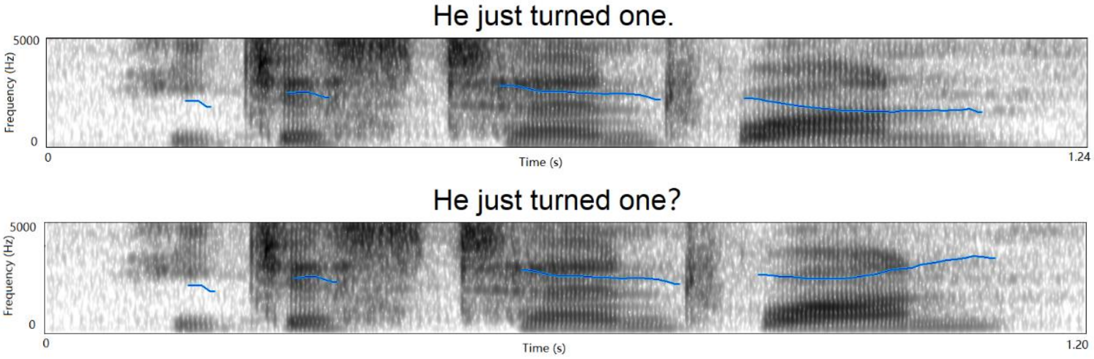


Figure 1. Spectrograms of an example sentence pair “He just turned one./?” with different final pitch contours in the statement (falling) and the question (rising).

Table 2: Sentences used in intonation tasks.

Sentence	Sentence Rate (syl/s)		Size of Final Pitch Glide (st)		Rate of Final Pitch Glide (st/s)		Duration of Final Pitch Glide (s)	
	S	Q	S	Q	S	Q	S	Q
It’s a lie./?	4.7	4.4	-2.5	7.1	-11.3	40.4	0.22	0.18
This is love./?	3.9	3.5	-3.6	2.6	-17.4	20.7	0.21	0.12
He hurt his knee./?	4.3	4.2	-3.0	3.7	-16.7	21.6	0.18	0.17
The answer is no./?	5.6	5.6	-2.7	4.7	-24.3	28.3	0.11	0.17
The deal is still on./?	5.5	5.3	-3.7	4.4	-24.3	31.4	0.15	0.14
He just turned one./?	3.9	4.0	-5.1	4.0	-29.5	16.6	0.17	0.24
She looks like Anne./?	3.9	4.0	-2.4	3.4	-17.2	26.3	0.14	0.13
She changed her name./?	4.3	4.3	-4.5	5.5	-16.7	35.3	0.27	0.16
It's a menu./?	5.4	5.5	-5.4	2.9	-18.8	17.2	0.29	0.17
She looks manly./?	4.3	3.9	-6.2	4.1	-14.8	12.4	0.41	0.33
He lives in Ealing./?	5.7	5.6	-6.4	8.1	-27.1	32.6	0.24	0.25
She grew up in Ely./?	5.8	5.8	-4.0	9.6	-18.3	29.2	0.22	0.33
They were in a limo./?	6.7	6.4	-3.5	6.9	-17.5	40.7	0.20	0.17
They named her Lilly./?	5.3	5.4	-4.2	4.9	-25.5	26.5	0.17	0.18
It’s from Emily./?	6.1	6.1	-4.1	5.0	-26.3	12.7	0.16	0.39
He speaks Romany./?	4.6	4.4	-3.3	2.9	-7.5	10.3	0.44	0.28
He was born in Illinois./?	6.1	5.9	-3.0	3.2	-8.0	18.9	0.37	0.17
He considers her his enemy./?	6.7	6.5	-3.7	4.8	-10.4	19.3	0.35	0.25
Mean	5.1	5.0	-4.0	4.9	-18.4	24.5	0.24	0.21
SD	0.9	1.0	1.2	1.9	6.6	9.3	0.10	0.08

S = statement; Q = question; syl=syllable; st=semitone.

Musical analogues of the sentences were created in Praat (Boersma & Weenink, 2001), matching the original sentences in pitch and temporal patterns, following the procedure in Patel et al. (1998). These musical tones were made of the fundamental frequency and its seven odd harmonics of the individual syllables in the original sentences, with peak amplitudes normalized to match those of the sentences (see Liu et al. (2010) for full details).

The speech and musical discrimination tasks were conducted in two separate blocks (order counterbalanced), where participants were presented with 36 pairs of stimuli (either speech or musical analogues) in either the same or different condition: nine statement-statement pairs, nine question-question pairs, nine statement-question pairs, and nine question-statement pairs. Participants were asked whether the pairs in each trial were the same or different, with their answers recorded by an experimenter by clicking a button on the computer. The interstimulus interval was 750ms and the intertrial interval was 2 seconds. The sounds were presented at a comfortable hearing level, which was determined and adjusted by participants themselves, through Sennheiser HD280 pro headphones and a Roland RUBIX22 USB Audio Interface. Two additional pairs were included as practice trials to familiarise participants with the procedure.

The identification and imitation tasks were conducted in the same block, where participants were presented with the 36 speech sentences one at a time. They were instructed to first imitate the sentence just played as exactly as possible (while their voices were recorded), and then to indicate whether the original sentence (not their imitation) was a statement or a question. The experimenter recorded the identification responses in Praat. Prior to the experiment, participants were familiarised with the procedure using two additional sentences (one statement and one question) in a practice session.

Data analysis

In the pitch direction discrimination task, thresholds were transformed using log transformation for parametric statistical analysis (Howell, 2009). Nine ASD participants and two controls did not complete this task, and their missing data were labelled “Not Applicable” (NA). Additionally, to screen for inattentive performers, following Moore et al. (2008), the visual tracks were closely inspected on an individual basis. Two children were identified as “non-compliant” (a term used in previous literature; Fancourt et al. (2013)) performers due to fluctuations of attention, and their data were excluded from further analysis (see Figure S1 for their visual tracks and Table S1 for the remaining participants' demographic characteristics in the pitch direction discrimination task).

Signal detection analysis (D-prime) was carried out for the intonation data. Specifically, in the discrimination tasks, correct responses to "different" trials were coded as hits; and in the identification and imitation tasks, correct responses to "question" trials were coded as hits, and d' was calculated using the *psycho* package in Rstudio (Makowski, 2018; RStudio Team, 2020). In the *Psycho* package, to estimate d' from extreme values (hits = 100% or 0%), corrections are made following the log-linear rule, for which the frequency of each category (e.g., Hit and False alarm) increases by 0.5 (Hautus, 1995). For the recordings from the imitation task, we employed acoustic analysis as a quantitative measure of performance accuracy using an earlier version of ProsodyPro (Xu, 2013) in Praat. Signed glide sizes (in Hz) were extracted from the final words of the sentences, with negative values indicating downward glides and recorded as statements, and positive values indicating upward glides and recorded as questions (Liu et al., 2010).

Bayesian analyses were run using JASP software (JASP Team, 2020). Bayes Factors indicate the strength of the evidence obtained and are particularly helpful in determining when the evidence supports the null hypothesis over an alternative. Unlike frequentist statistics, Bayes Factors test the relative probability of the two hypotheses given the data, rather than the probability of the data given the null hypothesis and so can be used to support both alternative and null hypotheses (Dienes, 2014). The Bayes Factor (BF) in favour of the null is the reciprocal of the BF in favour of the alternative. Also unlike frequentist significance testing, BFs give continuous measures of the likelihood of one hypothesis over another, which means cut-off values (e.g., p = 0.05) are inappropriate. For the interpretation of BFs as evidence for hypotheses, Raftery (1995, p.139) suggested ranges of values equivalent to different “strengths” of evidence, where a BF value above 1 and less than 3 is “weak” evidence and a BF between 3-20 represents “positive” evidence for an hypothesis.

2.3 Results

Pitch direction discrimination task

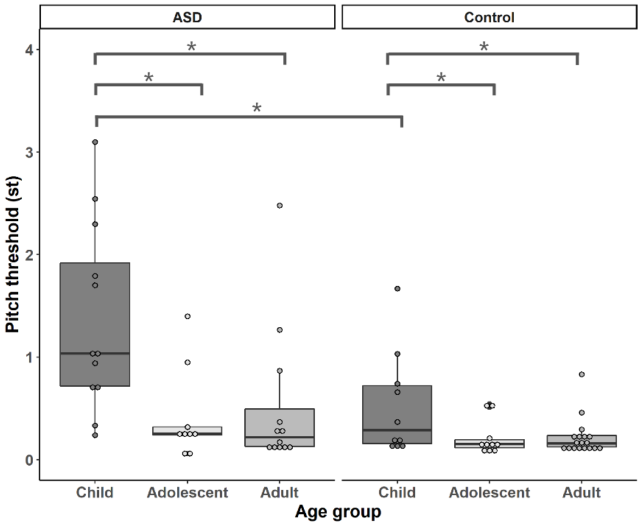


Figure 2. Pitch threshold in semitone (st) of each age cohort by group from the pitch direction discrimination task.

Figure 2 shows boxplots of the pitch thresholds for the ASD and control groups. Bayesian ANCOVA was fit to the data with group (ASD vs. control) and age cohorts (child, adolescent and adult) as the between-subjects variables and receptive vocabulary and nonverbal IQ scores as the covariates. The model revealed positive main effects of age (BF_{10} = 400381.43) and group (BF_{10} = 23.85) on pitch thresholds. The interaction between age and group also received positive support from Bayesian factor (BF_{10} = 4.70). The Bayesian post-hoc analysis suggested that both groups showed similar developmental trajectories, with the adult and adolescent cohorts performing better than the child cohort, though this trend was more pronounced in the ASD group than in the control group (see Table 3). The main effect of group was mainly driven by the difference across the child cohorts, as ASD children showed worse pitch thresholds than control children (BF_{10} = 7.07). The difference between the adult cohorts was weak (BF_{10} = 1.79) and no difference was found between the adolescent cohorts (BF_{10} = 0.46, weak evidence in favour of H0).

Table 3: Differences of pitch thresholds between age cohorts within each group.

		ASD	Control
Adolescent	Child	$BF_{10} = 15.00$	$BF_{10} = 1.52$
Adult	Child	$BF_{10} = 14.97$	$BF_{10} = 2.54$
Adult	Adolescent	$BF_{10} = 0.40$	$BF_{10} = 0.38$

There was a main effect of nonverbal IQ (BF_{10} = 251), whereas the main effect of receptive verbal ability was only weakly supported by Bayes factor (BF_{10} = 1.82). A Bayesian Kendall correlation analysis (1-tailed) showed that nonverbal IQ and receptive verbal ability were weakly associated with the performance of the control group (NVIQ: tau = -0.25, BF_{-0} = 1.14; ROWPVT-IV: tau = -0.18, BF_{-0} = 1.30) but not those of the ASD group (NVIQ: tau = -0.14, BF_{-0} = 0.74; ROWPVT-IV: tau = -0.04, BF_{-0} = 0.30).

Intonation discrimination tasks

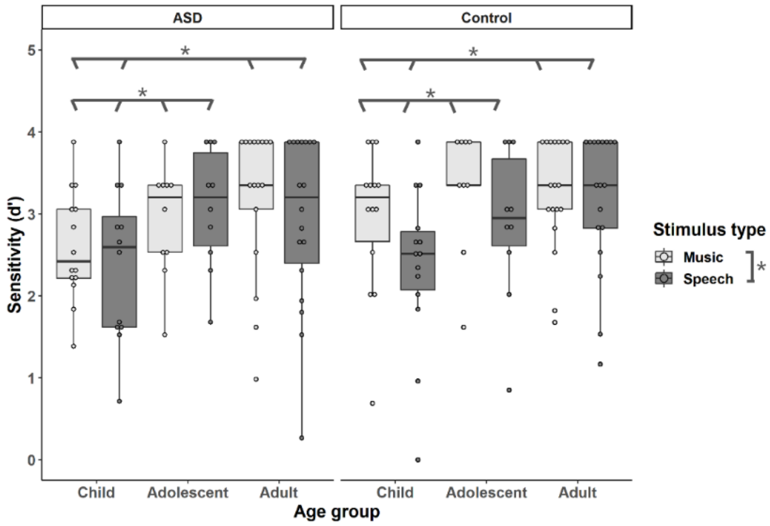


Figure 3. d' of each age cohort by stimulus type and by group from the discrimination task.

Figure 3 shows boxplots of the sensitivity d’ on the discrimination tasks for the ASD and control groups. Bayesian repeated measures ANCOVA was fit to the data with group (ASD vs. control) and age cohorts (child, adolescent and adult) as the between-subjects variables, stimulus type (speech vs. music) as the within subject-variable, and receptive vocabulary and nonverbal IQ scores as the covariates. The model revealed a positive main effect of stimulus type (BF_{10} = 7.09), with all groups performing better on the music condition than on the speech condition, and a main effect of age (BF_{10} = 3.57). The Bayesian post-hoc analysis showed that the adult cohort performed better than the child cohort (BF_{10} = 47.19), and the adolescent cohort also performed better than the child cohort (BF_{10} = 2.95), whereas the adult and adolescent cohorts performed comparably (BF_{10} = 0.28, positive evidence in favour of H0). No evidence for other main effects or interactions was observed, though the evidence for the null hypotheses was also weak: group (BF_{10} = 0.35), stimulus type by age cohort (BF_{10} = 0.59), stimulus type by group (BF_{10} = 0.58), age cohort by group (BF_{10} = 0.35) and stimulus type by age cohort by group (BF_{10} = 0.34). The evidence for main effects of receptive verbal ability (BF_{10} = 0.995), and nonverbal IQ (BF_{10} = 0.91), was equivocal.

To assess the relationship between pitch thresholds and the performance on intonation discrimination, 1-tailed Bayesian Kendall's correlation analysis was carried out separately for the groups and tasks, results are reported in Table 4. Pitch thresholds were negatively correlated with performance on both tasks for both groups: the lower (better) the pitch thresholds, the better performance on the speech and musical tasks.

Table 4. Kendall's correlations between performance on pitch thresholds and intonation discrimination tasks by group.

	ASD group	natural speech	musical analogues	**Control group**	natural speech	musical analogues
Pitch threshold	tau	-0.33	-0.34	tau	-0.38	-0.45
	BF_{-0}	11.88	15.60	BF_{-0}	91.05	765.98

Intonation identification and imitation tasks

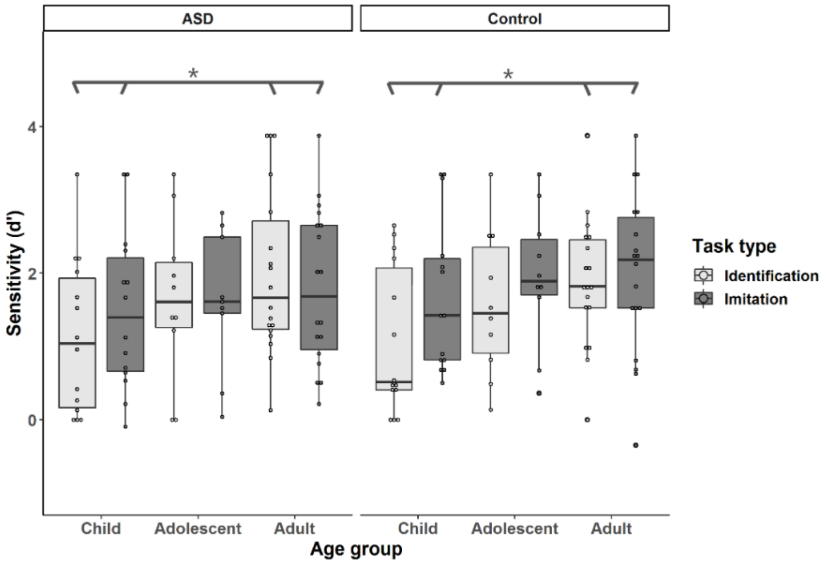


Figure 4. d' of each age cohort by group from the identification and imitation task.

Figure 4 shows boxplots of the sensitivity d’ on the identification and imitation tasks for the ASD and control groups. A Bayesian repeated measures ANCOVA was conducted with group (ASD vs. control) and age (child, adolescent and adult) as the between-subjects variables, task type (identification vs. imitation) as the within-subject variable, and receptive vocabulary and nonverbal IQ scores as the covariates. The model revealed a main effect of age (BF_{10} = 4.06). The post-hoc analysis showed that the adult cohort performed better than the child cohort (BF_{10} = 18.01) but similarly to the adolescent cohort (BF_{10} = 0.57), and the adolescent and the child cohorts performed comparably (BF_{10} = 0.47). For all other main effects and interactions, there was positive evidence in favour of the null hypotheses: task type (BF_{10} = 0.23); group (BF_{10} = 0.14); task type by age cohort (BF_{10} = 0.33); task type by group (BF_{10} = 0.13); age cohort by group (BF_{10} = 0.11) and task type by age cohort by group (BF_{10} = 0.01).

In addition, there was a weak main effect of receptive verbal ability, BF_{10} = 2.93, whereas no main effect of nonverbal IQ was observed, BF_{10} = 0.73. However, a Bayesian Kendall correlation analysis (1-tailed) showed that receptive verbal ability was not associated with performance on the imitation task (ASD: tau = 0.08, BF_{+0} = 0.40; control: tau = 0.10, BF_{+0} = 0.50) or the identification task (ASD: tau = 0.02, BF_{+0} = 0.23; control: tau = 0.07, BF_{+0} = 0.38) in either group, with Bayes factors supporting the null hypotheses positively.

Performance on identification and imitation was positively correlated for both groups (ASD: tau = 0.40, BF_{+0} = 329.97; control: tau = 0.28, BF_{+0} = 11.50). Using 1-tailed Bayesian Kendall’s correlation analysis, we analysed the relationship between pitch thresholds and identification as well as imitation tasks (see Table 5). The results indicated that pitch thresholds were negatively correlated with the performance on both tasks in the control group, but were only weakly associated with the identification performance in the ASD group.

Table 5. Kendall’s correlations between performance on pitch thresholds and intonation identification/imitation tasks by group.

	ASD group	intonation identification	intonation imitation	**Control group**	intonation identification	intonation imitation
Pitch threshold	tau	-0.22	-0.15	tau	-0.43	-0.26
	BF_{-0}	2.06	0.80	BF_{-0}	509.47	5.09

Response bias

To measure whether individuals with ASD showed response biases between the same versus different pairs (discrimination task) or questions versus statements (identification task), i.e., judging the same pairs as different or questions as statements, we calculated the percentage of correct responses to same/different pairs in the discrimination task and questions/statements in the identification task for the two groups.

A 2x2 mixed ANOVA with Bayesian analysis was conducted. Group (ASD vs. control) was the between-subjects factor, and response type (same vs. different in the discrimination task and question vs. statement in the identification task) was the within-subjects factor. In the discrimination task, there was a main effect of response type ($BF_{10} = 137.42$) with participants showing poorer performance on different pairs (M = 0.90, SD = 0.12) than on same pairs (M = 0.94, SD = 0.07). No main effect of group, or group by response type interaction was found, with Bayes factors tending to support the null hypotheses in both cases ($BF_{10} = 0.32$ and $BF_{10} = 0.52$, respectively). In the identification task, there was a main effect of response type, $BF_{10} = \infty$, with participants showing poorer performance on questions (M = 0.62, SD = 0.22) than statements (M = 0.90, SD = 0.15). No main effect of group or group by stimulus type interaction was observed, with $BF_{10} = 0.20$ and $BF_{10} = 0.24$, respectively.

To inspect individual response patterns, following Steffens et al. (2020), we calculated the probability that each individual accuracy rate was due to random guessing based on the binomial distribution. Accuracy rates with probabilities > 0.05 were interpreted as being likely due to random chance alone, whereas accuracy rates with probabilities ≤ 0.05 were interpreted as being unlikely due to chance alone (Steffens et al., 2020). We found that all participants performed above chance level in the discrimination task, while 24 participants showed chance level performance in the identification task (12 ASD vs. 12 control). A 2x2 mixed ANOVA with Bayesian analysis with Group as the between-subjects factor and response type as the within-subjects factor on the responses of the 24 participants revealed no main effect of group, $BF_{10} = 0.28$, or group by stimulus type interaction, $BF_{10} = 0.37$. There was a main effect of stimulus type, as participants were less able to identify questions (M = 0.27, SD = 0.24) than statements (M = 0.82, SD = 0.20).

2.4 Discussion

Using pitch thresholds and intonation perception and production tasks, the present study examined the abilities of individuals with and without ASD to use pitch to differentiate, identify, and imitate intonation (statements vs. questions) and whether these abilities would be affected by response bias, age (child, adolescent vs. adult), stimulus type (speech vs. music), and pitch direction discrimination thresholds. The main results showed that the performance of intonation discrimination (in both speech and music conditions), identification, and imitation was comparable between the ASD and TD groups within each age cohort, and that performances across tasks were largely independent of participants’ receptive verbal ability and nonverbal IQ, especially for participants with ASD. In addition, no response bias was observed in the discrimination and identification of statements and questions among participants with ASD. Participants’ abilities to discriminate, identify and imitate intonation were associated with their pitch direction discrimination thresholds for both groups. There were also age-related improvements across all tasks for both groups. These findings suggest that some individuals with ASD may have genuinely intact abilities to differentiate, identify, and imitate statement-question intonation, and they may also show similar developmental trajectories as typically developing individuals, with performance on both intonation and pitch thresholds increasing with age.

Perception of statement-question intonation and response bias in ASD

Regarding discrimination and identification of statements and questions, we found no group differences in response accuracy across all three age cohorts. Bayesian analyses supported our null results weakly for the discrimination task but positively for the identification task. Thus, no strong conclusions can be drawn about intonation discrimination abilities between the ASD and control groups. These findings are consistent with the majority of the literature that suggests intact statement-question identification in ASD (Chevallier et al., 2009; Filipe et al., 2014; Järvinen-Pasley, Peppé, et al., 2008; Paul et al., 2005). However, they contradict the findings indicating impaired discrimination (McCann et al., 2007; Peppé et al., 2007), impaired identification among “language-impaired” preadolescents (Lyons et al., 2014), and impaired discrimination and identification among Mandarin speakers (Jiang et al., 2015). Notably, as previously mentioned, the impaired discrimination suggested by McCann et al. (2007) and Peppé et al. (2007) was evaluated using the short-item discrimination task within PEPS-C, which contains the laryngographic sounds of the statement-question pairs, as well as those of the liking-disliking pairs from the affect subtask. Thus, the discrimination performance in ASD reported by these two studies may be confounded by the unnaturalness of the stimuli as well as by participants’ ability to discriminate affective pairs.

In addition, the inconsistency between the present study and Jiang et al. (2015) may be explained by language differences. Jiang et al. (2015) used stimuli in Mandarin which is a tone language, while the present study used stimuli in English which is a non-tone language. It has been suggested that the perception of statement-question intonation in tone languages is complicated by the changes in tones, which convey lexical meaning (Jiang et al., 2015; Liu & Xu, 2005; Xu, 2005), resulting in the tasks in Jiang et al. (2015) being more difficult than the present study. Finally, impaired statement-question identification was only observed among “language-impaired” preadolescents in Lyons et al. (2014). Indeed, prosodic skills correlate significantly with language ability in ASD (McCann et al., 2007; Peppé et al., 2007), In our current study, there was a weak main effect of receptive verbal ability on intonation identification/imitation (although correlations were nonsignificant), but not on intonation discrimination, which may be because our ASD and TD participants were largely matched on a range of cognitive abilities (Table 1).

We also examined the response biases that were reported in some studies

(Järvinen-Pasley, Peppé, et al., 2008; Peppé et al., 2007). Inconsistent with those results, but consistent with the findings of Jiang et al. (2015), individuals with ASD in our study did not show a tendency to judge the same pairs as different or identify questions as statements and this null result receives substantial support from the Bayes factors. Similar to controls, our participants with ASD displayed poorer performance when discriminating different pairs than same pairs, and when identifying questions than statements. While the response bias in the discrimination task reported by Peppé et al. (2007) reached a significant level, the declarative bias in identification from the PEPS-C turn-end task lacked statistical support (Järvinen-Pasley, Peppé, et al., 2008; Peppé et al., 2007). Using the speech stimuli from Patel et al. (1998), Järvinen-Pasley, Peppé, et al. (2008) observed a significant declarative bias among 50% of participants with ASD (in comparison to 10% of controls) for the identification task. Following Jiang et al. (2015) and Steffens et al. (2020), the present study used ANOVA models to inspect participants' response patterns, and found no response bias in either discrimination or identification tasks in ASD. It is worth noting that the difference in results between our study and previous studies is not due to our sample size being smaller. In fact, our ASD sample size is the largest among all these studies: 42 (our study), 31 (McCann et al., 2007; Peppé et al., 2007), 21 (Järvinen-Pasley, Peppé, et al., 2008), and 17 (Jiang et al., 2015). Given the reproducibility problems in science (Begley & Ioannidis, 2015), further studies are needed to determine whether there are genuine response biases in intonation discrimination and identification in ASD.

Production of statement-question intonation in ASD

In the imitation task, participants with ASD showed comparable performance to controls. Although previous research has found a deficit in intonation production in ASD, either based on subjective perceptual judgements or objective acoustic measures (Filipe et al., 2014; Fosnot & Jun, 1999; McCann et al., 2007; Peppé et al., 2007, 2011), the current study did not observe this deficit and the balance of evidence provided by our Bayesian analyses is sufficient for us to be confident in our null results. This discrepancy mainly results from the different methods used in reporting/analysing production data among these studies. Unlike previous studies using subjective judgements of the sentences produced (Filipe et al., 2014; McCann et al., 2007; Peppé et al., 2007, 2011), we explored objective measures by calculating glide sizes to verify the acoustic realisation of pitch direction in statements and questions in ASD. Thus, imitations were scored as correct only if participants shared the same sign in glide size as the models (i.e., statements imitated as statements with final falls would have negative glide sizes, and questions imitated as questions with final rises would have positive glide sizes). Additionally, while the present study was inspired by previous acoustic studies suggesting the key role pitch plays in atypical intonation production in ASD (Filipe et al., 2014; Fosnot & Jun, 1999), we used a different acoustic analysis method than those studies, in order to capture the production of pitch direction specifically. That is, when calculating imitation accuracy, we did not consider mean pitch, pitch range, or other variables, which measured the characteristics of speech production rather than imitation accuracy per se (Filipe et al., 2014; Fosnot & Jun, 1999). Rather, we focused on using objective measures to gauge the relationship between identification and imitation. The results suggested that identification abilities positively correlated with imitation abilities in both groups and that the association was stronger in the ASD group than in the control group, which were substantially supported by Bayes factors. These findings are consistent with the finding of intact identification and production of turn-end sentences in previous studies (Järvinen-Pasley, Peppé, et al., 2008; Peppé et al., 2007). Our correlation analysis further indicates that the more accurate the participants were on prosody perception and understanding, the better their performance on prosody production. Thus, an increase in receptive prosodic skills might result in amelioration of expressive prosodic disorder in ASD, and vice versa.

Perception of pitch in speech versus music in ASD

The third aim of this study was to investigate intonation processing in speech versus music in ASD. The results showed that, like the controls, participants with ASD performed better on discriminating between musical glides than on speech utterances. These findings received positive support from our Bayesian analysis. While the better performance on music than on speech in the intonation task is consistent with our hypothesis for ASD, the same perceptual pattern was also noted in the control group. It has been suggested that semantic information might hamper controls' performance due to their overly selective attention towards the content (Järvinen-Pasley & Heaton, 2007). Similarly, Bijou and Ghezzi (1999) have proposed a Behavior Interference Theory which states that typically developing children tend to focus on social stimuli (i.e., the human voice), whereas these stimuli do not easily obtain attention from children with ASD. For example, when two typically developing children in a playpen observed their mother walking back and forth from one side of the room to the other side of the room, their eyes firmly fixed on the mother while they reliably followed their mother back-and-forth. Conversely, a child with ASD sat on the floor of the room, overly focused on the shiny glare of a phone case and manipulated the plastic phone case in a non-functional manner (Bijou & Ghezzi, 1999). In the present study, both groups showed semantic interference with perceptual processing and performed better on discrimination of musical analogues than natural speech. This perceptual pattern is consistent with the findings of Liu et al. (2010) for amusic participants, Cheng et al. (2017) for ASD and control participants, as well as Francis & Ciocca (2003) for typically developing English and Cantonese listeners. However, other studies reported mixed findings regarding the effect of stimulus type on intonation processing among amusic and control participants (C. Jiang et al., 2010; Liu et al., 2012; A. D. Patel et al., 2005, 2008). Further studies are required to tease apart the effects of stimulus type, perceptual acuity, and sensory preference on intonation processing among different participant groups.

Since no group difference was observed in discrimination of speech and musical stimuli, the current findings provide evidence for shared mechanisms of pitch processing between music and speech in both individuals with ASD and controls (Liu et al., 2010). Similar findings were also reported in Cheng et al. (2017). However, numerous previous studies have suggested enhanced musical processing in ASD compared to controls, including local music processing (Mottron et al., 2000), melodic contour identification (Jiang et al., 2015), as well as memory and labelling of musical tones and segmentation of chords (Heaton, 2003). Nevertheless, more recent studies reported comparable or even impaired musical processing in ASD versus TD (Jamey et al., 2019; Schelinski et al., 2017; Sota et al., 2018). Thus, with mixed findings in the literature (Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007; Jiang et al., 2015), the domain specificity or generality of pitch processing in ASD warrants further studies.

The development of prosodic abilities and its relationship with pitch sensitivity

The fourth aim of our study was to examine the relationship between psychophysical pitch thresholds and intonation perception/imitation. We found that children with ASD had elevated pitch thresholds relative to their typically developing counterparts with substantial support from Bayesian factors. Adults with ASD performed worse relative to adult controls, and adolescents with ASD performed comparably to their controls in the pitch thresholds task. However, these findings received weak support from Bayesian analyses. So it would be premature to draw strong conclusions specific to these two age cohorts. In addition, our results point to positive relationships between the pitch thresholds and intonation perception/imitation in both groups: the more sensitive to pitch, the better performance on intonation perception. This correlation was more pronounced in the control group suggested by Bayes factors. This finding is consistent with previous research showing an overall positive relationship between low-level and higher-level pitch processing (Germain et al., 2019). These findings likely reflect a bottom-up cascading in which the degree of low-level strength or impairment influences performance at later stages, such as language acquisition and communication (Bertone et al., 2010; Germain et al., 2019; Stevenson et al., 2014).

Finally, the major contribution of the current study relates to the effects of age on pitch thresholds and intonation processing. Both groups showed age-related improvements across all tasks with positive support from Bayes factors. In particular, adults consistently showed smaller pitch thresholds and better intonation perception and imitation than children, suggesting a developmental improvement in pitch perception and intonation processing. Interestingly, age-related changes across the lifespan from children to adolescents to adults were not identical across different tasks.

Specifically, there were no significant differences in pitch thresholds and intonation discrimination between the adult and adolescent cohorts, who performed significantly better than the child cohort on those tasks. In terms of intonation identification and imitation, however, there was a gradual improvement from children to adolescents to adults, with no significant difference between adjoining age cohorts, but the adult cohort was significantly better than the child cohort. These findings suggest that pitch processing ability may improve with age, and that although important developments in the understanding and use of prosody continue during the school years (Cruttenden, 1985; Lyons et al., 2014; Wells et al., 2004), it is not yet adultlike for both ASD and TD participants.

Our finding of similar developmental changes in pitch discrimination ability across the ASD and control groups is incompatible with the markedly different developmental trajectories described by Mayer et al. (2016), where pitch discrimination ability increased with age in the control group but remained stable and enhanced across age cohorts in ASD. The discrepancy between the studies may be explained in several ways. First, there were differences in the paradigms used between the studies. The present study used an adaptive-tracking pitch threshold task to measure participants' pitch sensitivity starting with a default excursion size of six semitones, while Mayer et al. (2016) used stimulus pairs with either the same pitch or at a distance corresponding to 2, 3 or 6 semitones. Thus, the pitch variability of the stimuli used in their study was coarse, resulting in the task being easier than the one in the current study. Second, in Mayer et al. (2016), participants' ages overlapped between the child cohort (between 6 years 11 months and 14 years 9 months) and the adolescent cohort (between 9 years 8 months and 16 years 5 months), with both cohorts including intellectually lower-functioning ASD individuals, while the adult groups were all intellectually high-functioning. In our study, in order to match groups for age, gender and cognitive capability, all participants with ASD were intellectually high-functioning individuals and our age cohorts were defined with adults >= 16, adolescents between 12-15, and children between 7-11 years. Finally, the data in Mayer et al. (2016) were collected and combined from three separate studies, which may have also affected the results.

Furthermore, in contrast to "language-normal" preadolescents with ASD, those with language impairment showed developmental delays in the perception and production of statements and questions (Lyons et al., 2014). Our finding of no group difference in intonation discrimination, identification and imitation across age cohorts is in line with the results from "language-normal" preadolescents and adolescents with ASD in Lyons et al. (2014). While there are many different ways to categorize age cohorts (Ahmad et al., 2009; Nithyashri & Kulanthaivel, 2012), the present study followed the division methods in the Autism-Spectrum Quotient (Auyeung et al., 2008; Baron-Cohen et al., 2001, 2006). One limitation of using this three-way split is that the age differences are not fine-grained, which may not be sufficient to detect subtle developmental changes over time. Hence, while we observed developmental changes from children to adults in both ASD and TD, further studies are required to use more fine-grained classifications of age cohorts together with larger sample sizes, in order to map detailed developmental trajectories of pitch and intonation processing abilities in both groups.

Implications of the current research findings

The current study has a basic research focus with the aim of determining whether, and how, sensitivity to pitch and prosody is affected in a group of participants with ASD when compared to a matched control group. In doing so, however, the study is also directly investigating the responses of individuals with ASD to social stimuli, as speech is an inherently social act. As social stimuli can sometimes be somewhat aversive to children with ASD, possibly by virtue of their unpredictability, nonverbal children with ASD typically require hundreds of hours of generalized imitation training (e.g., object imitation, gross-motor imitation, oral-motor imitation, vocal imitation) and echoic/vocal mand training to improve their spoken language skills (Hampton & Kaiser, 2016). Even so, interventions are not always successful, not least because of other comorbidities in such children which may interfere with their ability to perceive, acquire, or reproduce speech (Tager-Flusberg & Kasari, 2013). For such children, our results have a number of consequences.

Firstly, our data confirm that, absent such comorbidities and assuming sufficient general cognitive capability, there is no a priori reason to suppose that speech processing and production mechanisms are impaired when actively processing speech and therefore the perceptual, cognitive and motor mechanisms are likely to be in place to support any speech-therapeutic programme which may be indicated. Secondly, there is no evidence for a dissociation between pitch processing in music and speech in our investigations. The data therefore support the use of musical stimuli (which may be less aversive to some individuals) as scaffolding for training in pitch-based discrimination and imitation for individuals who may be more attracted to music than to language. Note however, both that our sample were not representative of children with ASD and learning difficulties, and that appropriate translational studies to confirm the generalisability of training in pitch in music to perception and production of pitch in speech are beyond the scope of the current investigation. Thirdly, despite the generally equivalent performance across groups in our tasks—and the lack of a response bias which might otherwise complicate interpretations on some tasks—pitch discrimination thresholds are elevated in the ASD group relative to the control group. This identifies and highlights a particular perceptual problem in this group although, somewhat surprisingly, not one that impacted upon performance in other tasks despite the overall negative correlations between pitch threshold and intonation discrimination in both music and speech.

2.5 Conclusion

In the present study, an experimental, acoustics-based approach was used to investigate perception and production of prosody in ASD, facilitating objective comparisons between the two modalities of intonation. In addition, we examined intonation processing in ASD and TD from the perspectives of response bias, task condition (discrimination, identification, imitation), stimulus type (speech, music), developmental changes, and its association with pitch thresholds. Our study revealed that intonation discrimination (in both speech and music conditions), identification and imitation abilities may be intact in some individuals with ASD across age cohorts (children, adolescents and adults), although children with ASD tend to have elevated pitch direction discrimination thresholds than their typically developing counterparts. The ASD and control groups also showed a similar developmental improvement in pitch thresholds, intonation discrimination and identification, as well as imitation. Furthermore, intonation identification and imitation are associated in both individuals with ASD and their peers, suggesting that improvements in intonation comprehension may also contribute to intonation production, and vice versa. We also found an association between low-level pitch threshold and high-level intonation processing across all participants, which reflects that the degree of strength or impairment in low-level pitch processing may influence performance on language acquisition and/or communication skills, whereas this bottom-up effect is more pronounced in controls than in individuals with ASD. In summary, our findings provide evidence for shared mechanisms in pitch processing between speech and music, as well as associations between low- and high-level pitch processing and between perception and production of pitch in prosody in individuals with and without ASD, who also show similar developmental trajectories for these abilities. Further studies with individuals with ASD from different cultures, particularly in other languages, would be helpful in obtaining a more comprehensive understanding of these shared mechanisms in ASD.

Chapter 3. Study 2: Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation

Abstract

Individuals with autism spectrum disorder (ASD) often exhibit atypical imitation. However, few studies have identified clear quantitative characteristics of vocal imitation in ASD. This study investigated imitation of speech and song in English-speaking individuals with and without ASD and its modulation by age. Participants consisted of 25 autistic children and 19 autistic adults, who were compared to 25 children and 19 adults with typical development matched on age, gender, musical training, and cognitive abilities. The task required participants to imitate speech and song stimuli with varying pitch and duration patterns. Acoustic analyses of the imitation performance suggested that individuals with ASD were worse than controls on absolute pitch and duration matching for both speech and song imitation, although they performed as well as controls on relative pitch and duration matching. Furthermore, the two groups produced similar numbers of pitch contour errors, pitch interval errors and time errors. Across both groups, sung pitch was imitated more accurately than spoken pitch, whereas spoken duration was imitated more accurately than sung duration. Whereas children imitated spoken pitch more accurately than adults when it came to speech stimuli, age showed no significant relationship to song imitation. These results reveal a vocal imitation deficit across speech and music domains in ASD that is specific to absolute pitch and duration matching. This finding provides evidence for shared mechanisms between speech and song imitation, which involves independent implementation of relative versus absolute features.

Keywords: ASD, Vocal imitation, Speech, Song, Pitch, Duration

3.1 Introduction

Imitation is a fundamental skill that emerges early in typical human development (Meltzoff, 2017). It is essential for learning of complex constructs, including language (McEwen et al., 2007; Rose et al., 2009) and social interaction (Kuhl, 2007; Masur, 2006; Vivanti & Hamilton, 2014). In particular, by imitating others or being imitated, individuals gradually become aware of the physical world, such as cause-effect relations (Meltzoff & Williamson, 2013); the mental states of other people, such as their intentions and feelings (Meltzoff & Keith Moore, 1994); and the sounds around them, such as languages (Charman et al., 2000; Young et al., 2011).

Imitation in individuals with autism spectrum disorder (ASD) is often described as atypical (J. H. G. Williams et al., 2001). Deficits in imitation skills in ASD have been reported for a variety of tasks, including action imitation, which involves using body and hands to act (Ham et al., 2011; Young et al., 2011); object-directed action imitation where actions involve objects (Cossu et al., 2012; Vivanti et al., 2014); facial imitation (Bernier et al., 2007) and vocal imitation (McCann et al., 2007). Specifically, when individuals with ASD are instructed to imitate an action or utterance, they imitate with lower levels of accuracy and do so less frequently than typically developing (TD) counterparts (Edwards, 2014; Turan & Okcun Akcamus, 2012; Vivanti & Hamilton, 2014). Imitation deficits in ASD mainly manifest in high fidelity imitation of form, rather than in emulation of function or end points (Edwards, 2014). Functional magnetic resonance imaging studies suggest dysfunction of the mirror neuron system during action imitation in ASD (Yang & Hofmann, 2016).

Compared with other areas of imitation (e.g., action, object and face), research on vocal imitation in ASD is relatively scarce and has only focused on either the speech or music domain. Although several studies have addressed vocal imitation of speech in ASD, results to date are mixed regarding whether and to what extent (pitch, duration, and/or the balance between the two) individuals with ASD are associated with speech imitation deficits. Some differences across studies may be due to the use of acoustic analyses versus the use of perceptual ratings. For instance, based on ratings by speech and language therapists, children with ASD had impaired imitation of various prosodic forms, including affect, intonation, chunking, and focus (McCann et al., 2007; McCann & Peppé, 2003; Peppé et al., 2007, 2011). By contrast, acoustic analyses of pitch range showed no difference across groups for imitation of stress, despite the fact that ASD participants received lower perceptual ratings of accuracy than TD controls (Paul et al., 2008). Thus, Van Santen et al. (2010) called attention to the unreliability and bias of clinicians' perceptual ratings (not strictly blind to participants' diagnostic status) and advocated the advantages and objectivity of instrumental methods. However, studies employing acoustic measures to assess imitation performance also produced divergent findings, with one study showing a group difference in duration only but not in mean pitch (Diehl & Paul, 2012) and other studies reporting both pitch and duration differences between groups (Fosnot & Jun, 1999; Hubbard & Trauner, 2007).

In contrast to speech, music has been seen as an area of exceptional skills in ASD (Molnar-Szakacs & Heaton, 2012; Ouimet et al., 2012). However, only one study has examined music imitation in ASD, and the results suggested that children with ASD showed comparable or better performance than controls when imitating pitch, rhythm, and duration of musical tones based on independent observers' judgment (Applebaum et al., 1979). Thus, regarding vocal imitation, ASD seems associated with atypical speech imitation but normal to superior music imitation. Given that vocal imitation is crucial for language acquisition (Kuhl, 2000; Kuhl & Meltzoff, 1996) and successful imitation requires sensorimotor, cognitive and social skills (Fridland & Moore, 2015; Heyes, 2001; N. Nguyen & Delvaux, 2015; Over & Carpenter, 2013; Pagliarini et al., 2020), a potential impairment in vocal imitation may be related to landmark deficits of ASD including social and communicative difficulties (American Psychiatric Association, 2013; J. J. Diehl et al., 2015; McCann & Peppé, 2003). Previous studies have suggested that musical training benefits speech processing (A. D. Patel, 2011, 2012) and similar acoustic cues are used in emotional communication in music and speech (Juslin & Laukka, 2003). In addition, vocal imitation mechanisms are likely shared between speech and song production in adults with typical development (Mantell & Pfordresher, 2013; Wisniewski et al., 2013). Thus, the intimate link between music and speech begs the question as to whether vocal imitation impairment in ASD is indeed domain specific, especially when there has only been one study examining music imitation in ASD (Applebaum et al., 1979).

The domain specificity or generality of vocal imitation impairment in ASD is particularly relevant to a longstanding debate about whether speech and music share the same underlying processing systems (Albouy et al., 2020; Norman-Haignere et al., 2015; Zatorre & Gandour, 2008). The modular or domain-specific framework proposes that speech and music may involve distinct modules or mechanisms that deal with a particular aspect of the input and its output representation, either exclusively or more effectively than any other mechanisms (Fodor, 1983, 2001; Peretz, 2009; Peretz et al., 2015; Peretz & Coltheart, 2003; Peretz & Zatorre, 2005). While speaking and singing involve multiple processing components, musical abilities depend, in part, on modular processes such as tonal encoding of pitch, which is music-specific and independent of spoken pitch processing (Peretz, 2009; Peretz & Coltheart, 2003). In contrast to this view, others have suggested that speech and music systems may not be entirely modular or independent (Kunert & Slevc, 2015; A. D. Patel, 2013). Rather, there are shared or domain-general mechanisms underlying the processing of information across both domains (Koelsch, 2011; Koelsch & Siebel, 2005; A. D. Patel, 2010; Sammler et al., 2009). Numerous studies have provided evidence in support of either the domain-specific or domain-general view (Kunert & Slevc, 2015; Peretz et al., 2015). In addition to comparing music with language processing in typical development (Slevc et al., 2009; Slevc & Miyake, 2006), neurodevelopmental disorders such as congenital amusia (Liu et al., 2010, 2013) and ASD (DePriest et al., 2017; J. Jiang et al., 2015) could offer special insight into this debate, particularly regarding whether deficits are only present in one domain (e.g., music), but not in the other (e.g., speech).

Specifically, as a functional output representation, vocal imitation of speech and song could inform the domain-specific vs. domain-general debate from a production perspective (Peretz, 2009). Using matched speech and song stimuli, Liu et al., (2013) directly compared speech with music imitation in individuals with and without congenital amusia, a disorder of music processing (Ayotte et al., 2002). Individuals with congenital amusia demonstrated impaired pitch and duration matching of speech and song in terms of both absolute and relative measures. These findings suggest that vocal imitation mechanisms are likely shared between speech and song production even in congenital amusia (Liu et al., 2013). Although prior findings on vocal imitation in individuals with ASD seem to favour the domain-specific model, there have been too few published studies especially in music imitation in ASD to draw valid conclusions to inform the theoretical debate of music and language processing in this clinical population. Also, no studies have directly compared imitation abilities in speech versus music in ASD using matched linguistic and musical tasks to address the question of whether impaired speech imitation but spared/enhanced music imitation would be present in the same sample of participants. Furthermore, the absolute measures on pitch and duration matching used in Liu et al. (2013) required higher fidelity imitation than the relative measures. It remains to be determined whether individuals with ASD would show worse performance on absolute feature matching than relative feature matching during vocal imitation, similar to other areas of imitation in ASD (Edwards, 2014).

Studies of speech imitation in typically developing children and adults suggest that speech imitation ability is influenced by age (Kent & Forner, 1980; Loeb & Allen, 1993; Snow, 1998). Specifically, young children (aged 4) tended to imitate speech segments with longer duration and greater variability than older children (aged 6 and 12) and adults (Kent & Forner, 1980). While 3- and 4-year-olds showed more difficulty in imitating rising intonation contours in questions than falling intonation contours in statements, 5-year-olds were able to imitate both types of contours (Loeb & Allen, 1993; Snow, 1998). Studies of music imitation have examined pitch matching of tones or melodies in children (Cooper, 1995; Geringer, 1983) and adults (Amir et al., 2003; Pfordresher & Brown, 2007). Among children, pitch matching accuracy increases with age, with fourth-graders (9-10 years old) performing significantly better than third-graders (8-9 years old) (Cooper, 1995). While more than half of fourth-graders could match pitch within 50 cents (0.5 semitones), pre-school children (4-5 years old) produced a median deviation of 2.5 semitones (Geringer, 1983). Results on tempo matching or rhythm reproduction during music imitation suggest that, at 6 years, both musicians and non-musicians were able to reproduce rhythmic patterns embedded in a string of syllables (Gérard & Auxiette, 1992; Reifinger, 2006), and rhythmic response ability increased with age from Grade 1 to Grade 3 students (Schleuter & Schleuter, 1985). In addition, adults outperformed 5- to 7-year-old children on rhythm repetition, melody repetition, prosody repetition, as well as a range of other language and music tasks (Cohrdes et al., 2016). Thus, similar to speech imitation, music imitation ability is also influenced by age in typical development. In ASD, the evidence from studies on non-vocal imitation (e.g., action, object and face) suggests that, although imitative abilities increase over time, impairments in imitation continue throughout the lifespan in ASD (Biscaldi et al., 2014; Vivanti & Hamilton, 2014; Young et al., 2011). However, how age influences speech and music imitation in ASD has not been systematically studied.

In the current study, we examined vocal imitation abilities in children and adults with and without ASD using matched speech and song stimuli, addressing three research questions: (1) Do imitation abilities of individuals with ASD differ from controls in terms of pitch and duration matching across speech and music domains? (2) Do individuals with ASD differ from controls with respect to relative and absolute feature matching in vocal imitation? (3) Do vocal imitation abilities in ASD and TD vary with age? Based on previous findings, we hypothesized that: (1) Participants with ASD would show impaired pitch and duration imitation in speech but not in song compared to controls; (2) Participants with ASD would show poorer performance on absolute feature matching than on relative feature matching as compared to controls; and (3) Across both groups, the adult cohort would perform better than the child cohort overall.

3.2 Method

Participants

A group of 44 individuals with ASD and 44 matched controls were recruited via a variety of methods including email lists, local social media advertisements, and local experimental participant databases. All were native British English speakers with no speech or hearing problems, and reported no history of other neurological or psychiatric disorders. Participants in the ASD group had a formal diagnosis of autism spectrum disorder by clinicians. Participants in the control group were included using the cut-off scores of 32 (adults), 30 (adolescents) or 76 (children) on the Autism-Spectrum Quotient (AQ) (Auyeung et al., 2008; Baron-Cohen et al., 2001, 2006). All participants had normal hearing in both ears, with pure-tone air conduction thresholds of 25 dB HL or better at frequencies of 0.5, 1, 2, and 4 kHz. Participants' nonverbal IQ was estimated using the Raven's Standard Progressive Matrices Test (Raven et al., 1998), and verbal IQ was estimated by the Receptive One Word Picture Vocabulary Test IV (ROWPVT-IV) (Martin & Brownell, 2011). The Corsi block-tapping task was used to assess participants' nonverbal short-term memory span (Kessels et al., 2000), and the forward digit span task was used to assess verbal short-term memory (Wechsler, 2003). Participants were further divided into two age cohorts, children (< 16) and adults (>= 16), based on the age cut-off of 16 years, following the definition of adults in the Autism-Spectrum Quotient (Baron-Cohen et al., 2001). The reason that we used a two-way rather than a three-way split of age cohorts (children, adolescents and adults) was to ensure that there were enough participants in each cohort. The age range of the child cohort was between 7.39 to 15.75 years and that of the adult cohort was between 16 to 56.75 years. Children's music perception skills were assessed using the Montreal

Battery of Evaluation of Musical Abilities (MBEMA), which consists of five subtests with 20 trials each measuring the perception of scale, contour, interval, rhythm and recognition memory of musical melodies (Peretz et al., 2013). Adults were assessed using the Montreal Battery of Evaluation of Amusia (MBEA), which contains six subtests with 30 trials each measuring the perception of scale, contour, interval, rhythm, meter and recognition memory of musical melodies (Peretz et al., 2003). All participants also completed a questionnaire about their musical, language and medical background, where they were also asked to report whether they possess absolute pitch or perfect pitch, the ability to identify a musical note without a reference tone (Deutsch, 2013). As can be seen in Table 1, the ASD and control groups were comparable on all background measures. The study was approved by the University of Reading Research Ethics Committee. Written informed consent/assent was obtained from the participants and/or their parents prior to the experiment.

Table 1. Characteristics of the ASD (*n* = 44) and control groups (*n* = 44).

Age cohort	**Background measures**	**ASD**	**Control**	***W***	***P***	**Rank-Biserial Correlation**
Children (*n* = 50)	Gender (F:M)	4:21	4:21			
	Age	11.41(2.64)	11.17(2.63)	335	0.67	0.07
	Musical training	2.18(2.71)	1.76(2.26)	335	0.66	0.07
	NVIQ	69.00(26.14)	79.40(23.06)	225	0.08	0.28
	VIQ	122.12(15.43)	126.20(14.96)	273	0.45	0.13
	Corsi	5.12(1.30)	5.60(1.38)	267.5	0.37	0.14
	Digit span	5.48(1.05)	5.80(0.91)	246.5	0.18	0.21
	Self-reported Absolute pitch	*n* = 2	*n* = 2			
	Scale	16.64(2.20)	17.29(17.29)	232	0.17	0.23

	Contour	15.04(3.41)	16.63(2.70)	218.5	0.10	0.27
	Interval	15.52(3.38)	16.50(3.60)	239	0.22	0.20
	Rhythm	16.16(3.38)	17.46(3.05)	232.5	0.18	0.23
	Memory	16.48(3.08)	17.75(2.42)	227.5	0.14	0.24
	Pitch composite	47.20(8.11)	50.42(8.11)	224.5	0.13	0.25
	MBEMA Global	79.84(13.63)	85.62(12.62)	216	0.10	0.28
Adults **(*n* = 38)**	Gender (F:M)	10:9	10:9			
	Age	34.51(13.55)	33.74(12.73)	184.5	0.92	0.02
	Musical training	4.26 (6.40)	5.50(6.72)	165	0.65	0.09
	NVIQ	53.95(28.07)	47.37(31.11)	202.5	0.52	0.12
	VIQ	110.72(13.75)	111.53(14.09)	153	0.59	0.11
	Corsi	5.74(1.45)	6.26(1.15)	140.5	0.22	0.22
	Digit span	7.05(1.62)	7.00(1.11)	180	1.00	0.003
	Self-reported Absolute pitch	*n* = 2				
	Scale	25.79(3.41)	26.16(2.57)	172	0.81	0.05
	Contour	25.11(3.43)	25.53(2.95)	177.5	0.94	0.02
	Interval	24.21(3.17)	25.53(2.95)	141	0.25	0.22
	Rhythm	26.00(3.87)	26.63(2.11)	186.5	0.87	0.03
	Meter	22.11(9.38)	25.26(6.98)	145.5	0.31	0.19
	Memory	27.11(1.88)	26.42(2.39)	205.5	0.47	0.14
	Pitch composite	75.11(9.04)	77.21(6.02)	163.5	0.63	0.09
	MBEA Global	83.5(10.5)	86.4(7.3)	154	0.45	0.15

Note: Musical training: years of musical training; NVIQ: percentile point of Raven's Standard Progressive Matrices Test; VIQ: standard score of Receptive One Word Picture Vocabulary Test; Corsi: raw score of nonverbal short-term memory; Digit span: raw score of verbal short-term memory; The child cohort used the MBEMA with 5 subtests (Scale, Contour, Interval, Rhythm, and Memory) with 20 trials each, the pitch composite is the sum of the scale, contour, and interval scores, and the MBEMA global score is the percentage of correct responses out of the total 100 trials; the adult cohort used the MBEA with 6 subtests (Scale, Contour, Interval, Rhythm, Meter, Memory) with 30 trials each, the pitch composite is the sum of the scale, contour, and interval scores, and the MBEA global score is the percentage of correct responses out of the total 180 trials; 2-tailed Mann-Whitney-Wilcoxon Test results were used to compare group difference and effect size was given by the rank biserial correlation in the Mann-Whitney test.

Stimuli

The target stimuli were 12 sentences either spoken or sung as statements or questions from Mantell and Pfordresher (2013), yielding 48 sentences with three to five syllables each. The speech stimuli were naturally spoken, and the pitch-time trajectory did not correspond to any diatonic scales. In order to create contour variation in the sequences, statements were produced with a falling contour and questions with a rising contour. The song stimuli comprised pitches from a major diatonic scale that approximated the global melodic contours of the speech stimuli. Each sung syllable had a roughly identical duration so as to invoke a metrical beat, resulting in the song stimuli being longer than the speech stimuli. Three versions of the speech/song stimuli were used for different age and gender groups. The adult male and female versions were taken from Mantell and Pfordresher (2013) and used for male/female participants >= 12 years old. For child participants < 12 years old, a child version was created by a child (female, 11-year-old, with five years of musical training) imitating the female version but in her own pitch range (see Figure 1; for more details, see Table 1S).

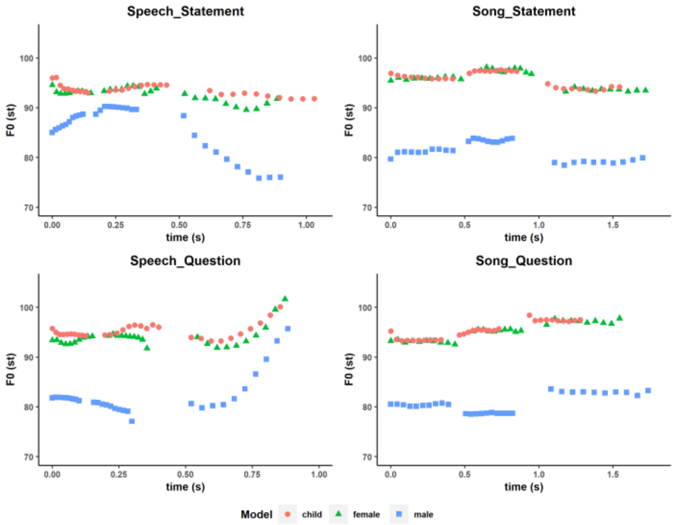


Figure 1. The pitch-time trajectory of the sentence “They went home” under different conditions by child/female/male target speakers.

Procedure

The presentation of the target stimuli and the recording of the imitations were both done using Praat (Boersma & Weenink, 2001). Participants were seated in a soundproof booth and presented with 4 practice trials (with items different from those in experimental trials: 2 speech vs. 2 song) to familiarize themselves with the task and the recording environment. Following the practice section, participants were presented with each of the 48 speech/song sentences one at a time in a pseudorandom order to ensure that different experimental conditions would alternate in an unpredictable manner and that long runs of the same condition (possible with true randomization) would not occur. Participants were instructed to imitate exactly the pitch and timing patterns of the sentences to the best of their ability, while their voices were recorded. Each sentence was played once and only replayed when participants failed to catch the words, and not when they wanted to listen to it again so they could imitate it better.

Data analysis

Recordings were analysed in Praat using ProsodyPro (Xu, 2013) to extract the pitch and duration of each syllable rhyme. The rhyme was defined as the vowel portion of the syllable plus any final voiced consonant (e.g., c**ar**, b**oo**k), which was done by the first author (a phonetician). Octave errors in pitch imitation were corrected, i.e., when imitated pitch was more than 6 semitones (half octave) apart from the model pitch, the value was adjusted as 12 – imitated pitch. In total, less than 3% of the data samples needed to be adjusted and most of these errors were caused by creaky voices, resulting in decreased fundamental frequency, F0 (Johnson, 2011). For accurate acoustic analysis of the data, we used ProsodyPro to manually add these missed vocal pulse marks for F0 based on the waveforms and spectrograms, to avoid having erroneous outliers misleading imitation results. Trials were not excluded when participants repeated the sentences slightly incorrectly but with the correct rhyme, e.g., substituting “he” for “she” or “brought” for “bought”. In the literature, pitch accuracy in singing and imitation has been analysed using a variety of measures, such as using median F0 (Dalla Bella, Giguère, et al., 2007; Dalla Bella et al., 2009; Liu et al., 2013) or mean F0 (Hutchins & Peretz, 2012) of the vowel or vocalic group to indicate pitch height of each note/syllable, or calculating mean absolute pitch error and pitch correlation across the entire pitch trajectories of the model and imitated sequences (Mantell & Pfordresher, 2013). For timing accuracy, either subjective ratings, e.g., 0 = “incorrect,” 0.5 = “partly correct,” and 1 = “correct” (Cohrdes et al., 2016), or objective acoustic analyses, e.g., number of time errors as determined by a 25% time deviation (Dalla Bella, Giguère, et al., 2007; Dalla Bella et al., 2009; Liu et al., 2013; Tremblay-Champoux et al., 2010) have been used. The pros and cons of these different methods and measures have been discussed (Dalla Bella, 2015). Since the ability to imitate/produce absolute versus relative features and pitch versus timing variables can dissociate in different “phenotypes” of poor singing (Berkowska & Dalla Bella, 2013; Dalla Bella & Berkowska, 2009), it is recommended that these dimensions be examined separately (Dalla Bella, 2015). Compared to mean F0, median F0 is a preferable measure of pitch height, since it is less affected by extreme or erroneous variation of F0 due to creaky voice (Dalla Bella, 2015). In contrast to the whole trajectory analysis of each sequence (Mantell & Pfordresher, 2013), measuring the median F0 of each note/syllable rhyme (or vowel group) makes the calculation of pitch interval and pitch contour (two critical components in memory for melodies) between consecutive notes/syllables possible (Dowling & Fujitani, 1971). Most importantly, similar to music, there are pitch targets in speech across tone and intonation languages, such as high, low, rising, and falling, and they are realised based on linguistic functions and articulatory constraints (Xu, 2005; Xu & Prom-on, 2014; Xu & Wang, 2001). With a tonal perception model, speech prosody can be transcribed using a stylization of pitch levels and movements coupled with vocalic segments (Mertens, 2004), enabling the comparison of spoken and musical rhythm and melody (A. D. Patel et al., 2006). Thus, taking a comparative approach to studying music and language (A. D. Patel et al., 2006) and balancing the advantages and disadvantages of different methods (Dalla Bella, 2015), we adapted the following absolute and relative pitch and time measures from earlier studies (Dalla Bella, Giguère, et al., 2007; Dalla Bella et al., 2009; Liu et al., 2013, 2016; Tremblay-Champoux et al., 2010) to compare imitation accuracy between music and speech in the current study.

The absolute pitch deviation (in cents): Median F0 was extracted from each syllable rhyme and then subtracted from that of their matched model to find the pitch deviation (in absolute value) for each imitated rhyme. The deviations were averaged over all syllables/notes in each utterance/melody, in order to control for the non-independence of the data points within in each utterance/melody (Mcdonald, 2014). The bigger the value, the less accurate the imitation in terms of absolute pitch matching.

The relative pitch deviation (in cents): Pitch interval was calculated as the absolute difference in median F0 between two consecutive syllables/notes, and then subtracted from their matched model speaker's pitch interval (in absolute value). The deviations were averaged over all intervals in each utterance/melody and the bigger the value, the less accurate the imitation in terms of relative pitch matching.

The number of contour errors: Contour errors were defined as imitated pitch intervals that differed from the corresponding target pitch intervals in regard to pitch directions (up, down, or level). Pitch direction was considered to be up or down if the difference in pitch interval was higher or lower by 50 cents (100 cents = 1 semitone) or more; otherwise (the difference was within 50 cents), the pitch intervals were considered to form a level/flat pitch direction. The number of contour errors was summed over each utterance/melody.

The number of pitch interval errors: Pitch interval errors were defined as imitated pitch intervals that were larger or smaller than the corresponding target pitch intervals by 100 cents without considering the pitch direction. Specifically, imitated and target pitch intervals were compared using absolute values. The number of pitch interval errors was summed over each utterance/melody.

The absolute duration difference (in milliseconds): Duration was extracted from each syllable rhyme and then subtracted from their matched model speaker's production to find the absolute difference for each rhyme. The differences were averaged over all rhymes in each utterance/melody and the larger the value, the less accurate the imitation in terms of absolute duration matching.

The relative duration difference (in milliseconds): Interonset interval (IOI) was calculated as the difference between two consecutive syllables/notes, and then subtracted from their matched model speaker's IOI (in absolute value). The differences were averaged over all IOIs in each utterance/melody and the larger the value, the less accurate the imitation in terms of relative duration matching.

The number of time errors: Time errors were defined as imitated syllables/notes that were more than 25% longer or shorter than the corresponding target syllables/notes

(Dalla Bella, Deutsch, et al., 2007; Dalla Bella et al., 2009; Prince & Pfordresher, 2012). This measure takes into account that in Western tonal music, event durations constitute simple integer ratio relationships, e.g., sixteenth notes (1/4 a beat), eighth notes (1/2 a beat), quarter notes (1 beat), etc., and counting time errors this way will capture the violation to the time signature (Drake & Palmer, 2000). Similarly, in stress-timed languages such as English, speech rhythm can also be measured in relative terms, making the comparison of spoken and musical rhythm possible (A. D. Patel et al., 2006; A. D. Patel & Daniele, 2003). The number of time errors was summed over each utterance/melody.

Statistical analyses were conducted using Rstudio (RStudio Team, 2020). We performed linear mixed effects analysis using the *lme4* (Bates et al., 2012; Brauer & Curtin, 2018) package with the above-mentioned pitch and time variables as the dependent variable and Diagnostic Group (effect-coded: Control vs. ASD), Age cohorts (effect-coded: Child vs. Adult), and Condition (effect-coded: Speech vs. Music) as well as all possible interactions as fixed effects. All models were fit using the maximal random effects structure that converged with two random factors (subject vs. file) (Barr, 2013; Barr et al., 2013). When the maximal model failed to converge, the random correlations were removed first. If the model still failed to converge, the random effect with the least variance was iteratively removed until the model converged. Statistical significance of the fixed effects was estimated using the *summary()* function of the *lmerTest* package (Kuznetsova et al., 2017), which provided *p* values for the corresponding *t* tests. Subsequent post-hoc comparisons, if any, were conducted using the *emmeans* package (Lenth et al., 2018).

3.3 Results

Absolute pitch deviation

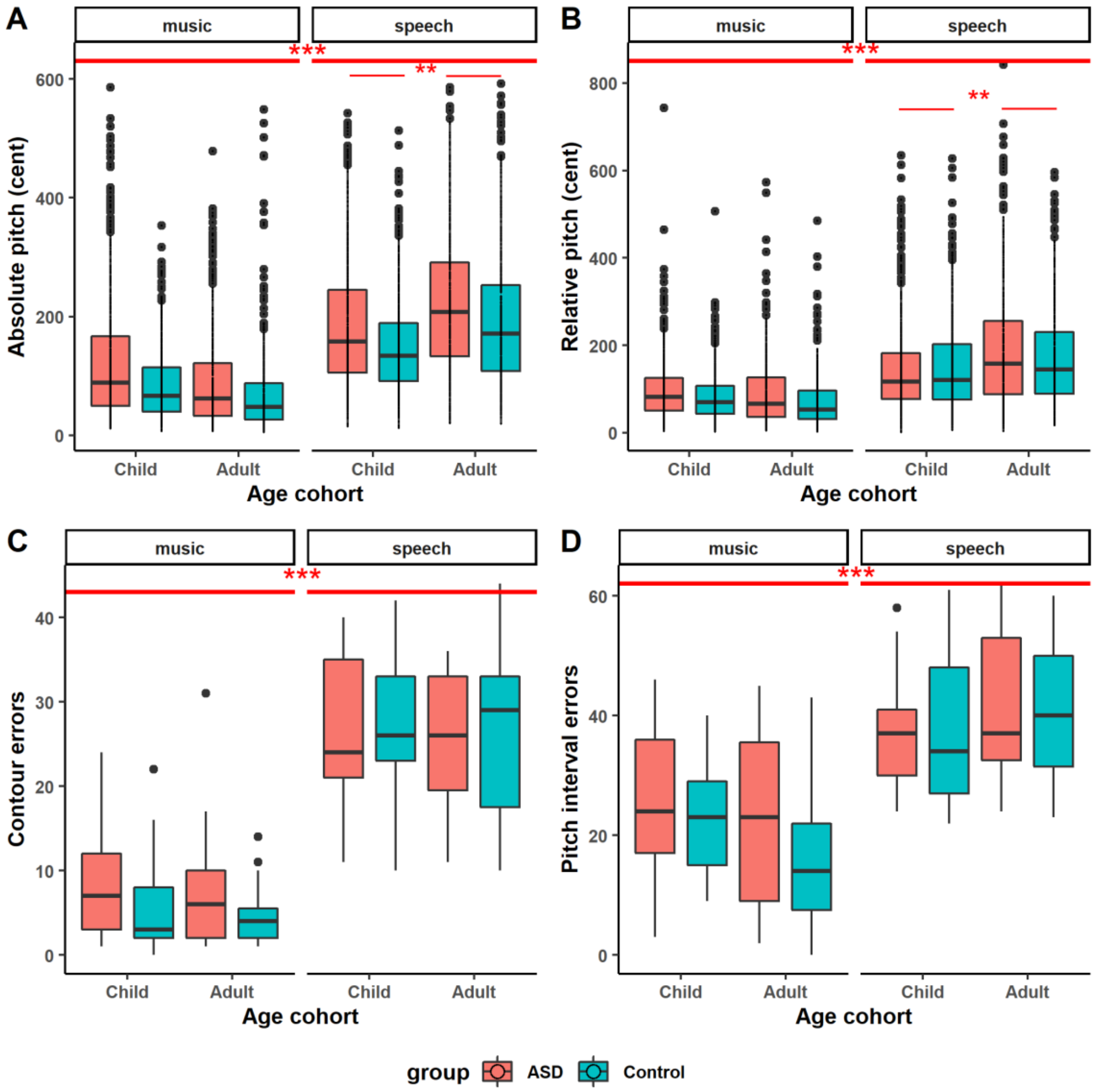


Figure 2. Boxplots of pitch-related measures for the ASD and control groups. (A) The absolute pitch deviation; (B) The relative pitch deviation; (C) The number of contour errors; (D) The number of pitch interval errors (Asterisks represent *p*-values between variables with $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$).

Figure 2(A) shows boxplots of the absolute pitch deviations for the ASD and control groups. Results from the linear mixed-effects model (Table 2) revealed a significant main effect of Diagnostic Group, as the ASD group (M(SD) = 156.17(116.08)) produced significantly larger absolute pitch deviations than did the Control group

(M(SD) = 124.48(97.45)). The main effect of Condition was also significant, with both groups showing better absolute pitch matching for the Music condition (M(SD) = 94.03(84.96)) than for the Speech condition (M(SD) = 186.62(109.39)). There was also a significant Condition by Age interaction and Post-hoc analyses with Bonferroni correction suggested that the child cohorts (M(SD) = 168.46(98.4)) showed better absolute pitch matching than the adult cohorts (M(SD) = 210.52(118.23)) in the speech condition (t(144) = 2.77, *p* = .006), whereas comparable performance (Child: M(SD) = 104.3(87.68); Adult: M(SD) = 80.54(79.29)) was observed in the music condition (t(144) = 1.56, *p* = .12). No other remaining main effects and interactions were significant.

Table 2. Results from the linear mixed-effects models for the pitch-related measures.

	Estimate	**Std. Error**	**df**	***t***	***p***
The absolute pitch deviations model					
Diagnostic Group	**15.57**	**6.44**	**84.00**	**2.42**	**0.02***
Age	4.60	6.44	84.00	0.71	0.48
Condition	**-48.53**	**4.16**	**94.87**	**-11.67**	**< .001*****
Group: Age	-2.37	6.44	84.00	-0.37	0.71
Group: Condition	-0.39	3.80	84.00	-0.10	0.92
Age: Condition	**-16.49**	**4.09**	**94.06**	**-4.04**	**< .001*****
Group:Age:Condition	-1.41	3.80	84.00	-0.37	0.71
The relative pitch deviations model					
Diagnostic Group	5.96	5.36	84.00	1.11	0.27
Age	6.30	5.36	84.00	1.17	0.24
Condition	**-39.97**	**3.74**	**67.65**	**-10.69**	**< .001*****
Group: Age	3.95	5.36	84.00	0.74	0.46
Group: Condition	2.72	2.90	84.02	0.94	0.35
Age: Condition	**-10.91**	**2.90**	**84.02**	**-3.76**	**< .001*****
Group:Age:Condition	-2.30	2.90	84.02	-0.79	0.43
The contour errors model					
Diagnostic Group	0.03	0.03	84.08	1.12	0.27
Age	-0.03	0.04	64.06	-0.74	0.46
Condition	**-0.42**	**0.04**	**31.98**	**-9.63**	**< .001*****
Group: Age	0.007	0.03	84.08	0.24	0.81

Group: Condition	0.02	0.02	84.17	0.91	0.36
Age: Condition	0.008	0.02	84.18	0.42	0.67
Group:Age:Condition	-0.004	0.02	84.17	-0.18	0.86
The pitch interval errors model					
Diagnostic Group	0.07	0.05	84.03	1.51	0.13
Age	-0.01	0.05	84.20	-0.23	0.82
Condition	**-0.38**	**0.03**	**52.38**	**-11.15**	**< .001*** **
Group: Age	0.02	0.05	84.03	0.37	0.71
Group: Condition	0.04	0.03	83.95	1.76	0.08
Age: Condition	**-0.10**	**0.03**	**57.66**	**-3.04**	**0.004*** *
Group:Age:Condition	0.001	0.03	83.95	0.05	0.96

Additionally, to evaluate the performance of those participants who self-reported possessing absolute pitch, we closely inspected the results of these participants (Table 3). Given that the ASD group showed impaired imitation of absolute pitch, we took the values from the control group as the “standard” (M(SD) = 124.48(97.45), Range: 63.36-298.26) and only two of them performed better than the average level.

Table 3. The results of absolute pitch deviation for participants who self-reported possessing absolute pitch.

ID	Diagnostic group	Age cohort	Absolute pitch deviation
Participant 1	ASD	Child	143.33
Participant 2	ASD	Child	76.70
Participant 3	ASD	Adult	173.24
Participant 4	ASD	Adult	156.85
Participant 5	Control	Child	78.43
Participant 6	Control	Child	180.60

Relative pitch deviation

Figure 2(B) shows boxplots of the relative pitch deviations for the ASD and control groups. Results from the linear mixed-effects model revealed a significant main effect of Condition with both groups showing better relative pitch matching for the Music condition (M(SD) = 86.5(67.1)) than for the Speech condition (M(SD) =

163.43(115.27)). The interaction between Condition and Age was also significant and Post-hoc analyses with Bonferroni correction suggested that the child cohorts (M(SD) = 148.75(103.56)) showed better relative pitch matching than the adult cohorts (M(SD) = 182.8(126.6)) in the speech condition (t(129) = 2.82, *p* = .006), while comparable performance (Child: M(SD) = 90.47(64.14); Adult: M(SD) = 81.27(70.49)) was observed in the music condition (t(129) = .76, *p* = .45). No other remaining main effects and interactions were significant (see Table 2).

Number of contour errors

Figure 2(C) shows boxplots of the number of contour errors for the ASD and control groups. Results from the linear mixed-effects model revealed, as shown in Table 2, a significant main effect of Condition, with both groups showing fewer contour errors with the Music condition (M(SD) = 6.49(5.84)) than the Speech condition (M(SD) = 25.83(8.95)). No other remaining main effects and interactions were significant (see Table 2).

Number of pitch interval errors

Figure 2(D) shows boxplots of the number of pitch interval errors for the ASD and control groups. As shown in Table 2, the linear mixed-effects model revealed a significant main effect of Condition, as both groups showed fewer pitch interval errors in the Music condition (M(SD) = 22.56(13.46)) than in the Speech condition (M(SD) = 39.28(11.2)). The interaction between Age and Condition was also significant, although Post-hoc analyses with Bonferroni correction revealed no significant difference between the child cohorts and adult cohorts in either condition (Speech: t(138) = -1.54, *p* = .13; Music: t(138) = 1.92, *p* = .06). No other remaining main effects and interactions were significant (see Table 2).

Absolute duration difference

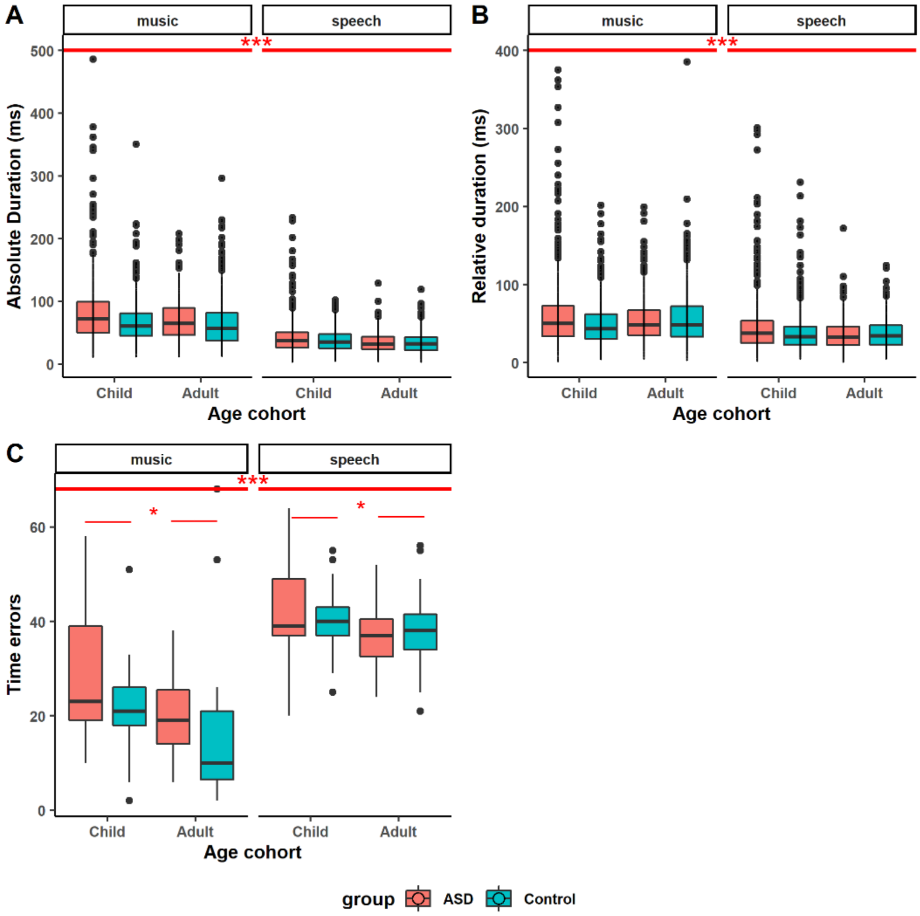


Figure 3. Boxplots of duration-related measures for the ASD and control groups. (A) The absolute duration difference; (B) The relative duration difference; (C) The number of time errors (Asterisks represent *p*-values between variables with $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$).

Figure 3(A) shows boxplots of the absolute duration differences for the ASD and control groups. The linear mixed-effects model revealed, as shown in Table 4, a significant main effect of Group, as the ASD group (M(SD) = 57.92(40.09)) produced significantly larger absolute duration differences than did the Control group (M(SD) = 51.48(32.55)). The main effect of Condition was also significant, with both groups showing larger absolute duration differences in the Music condition (M(SD) = 71.73(41.31)) than in the Speech condition (M(SD) = 37.67(20.01)). No other remaining main effects and interactions were significant (see Table 4).

Table 4. Results from the linear mixed-effects model for the duration-related measures.

	Estimate	**Std. Error**	**df**	***t***	***p***
The absolute duration differences model					
Diagnostic Group	**2.96**	**1.48**	**84.01**	**2.01**	**0.048***
Age	-2.73	1.51	87.62	-1.80	0.07
Condition	**17.00**	**1.46**	**62.19**	**11.64**	**< .001*****
Group: Age	-1.94	1.48	84.01	-1.31	0.19
Group: Condition	2.03	1.07	84.00	1.90	0.06
Age: Condition	-0.26	1.07	84.00	-0.25	0.81
Group:Age:Condition	-0.95	1.07	84.00	-0.89	0.38
The relative duration differences model					
Diagnostic Group	2.07	1.59	85.63	1.30	0.20
Age	-1.36	1.58	84.15	-0.86	0.39
Condition	**8.17**	**1.31**	**64.56**	**6.22**	**< .001*****
Group: Age	-2.95	1.57	83.87	-1.88	0.06
Group: Condition	0.28	1.04	78.87	0.27	0.79
Age: Condition	1.62	1.08	80.44	1.50	0.14
Group:Age:Condition	-0.93	1.01	83.79	-0.92	0.36
The time errors model					
Diagnostic Group	0.07	0.04	83.96	1.66	0.10
Age	**-0.10**	**0.05**	**87.60**	**-2.30**	**0.02***
Condition	**-0.39**	**0.04**	**48.47**	**-10.40**	**< .001*****
Group: Age	-0.02	0.04	83.96	-0.53	0.60
Group: Condition	0.05	0.03	84.07	1.97	0.052
Age: Condition	-0.04	0.03	57.96	-1.09	0.28
Group:Age:Condition	0.0009	0.03	84.07	0.03	0.97

Relative duration difference

Figure 3(B) shows boxplots of the relative duration differences for the ASD and control groups. The linear mixed-effects model revealed a significant main effect of Condition, with both groups showing bigger relative duration differences in the Music condition (M(SD) = 55.64(37.05)) than in the Speech condition (M(SD) = 39.75(27.2)). No other remaining main effects and interactions were significant (see Table 4).

Number of time errors

Figure 3(C) shows boxplots of the number of time errors for the ASD and control groups. The linear mixed-effects model revealed a significant main effect of Age, as the adult cohorts (M(SD) = 28.49(14.97)) produced fewer time errors than did the child cohorts (M(SD) = 33.19(12.98)). There was a main effect of Condition, as both groups made fewer time errors with the Music condition (M(SD) = 22.39(13.1)) than with the Speech condition (M(SD) = 39.93(8.26)). No other remaining main effects and interactions were significant (see Table 4).

3.4 Discussion

The present study investigated imitation of speech and song in English-speaking individuals with and without ASD and its modulation by age using absolute and relative pitch and duration measures. The main results showed that individuals with ASD were worse than controls on absolute pitch and duration matching, while performing as well as controls on relative pitch and duration matching in both speech and song imitation. In addition, the two groups produced similar numbers of pitch contour errors, pitch interval errors, and time errors. Furthermore, like the controls, individuals with ASD imitated sung pitch more accurately than spoken pitch, whereas spoken duration was imitated more accurately than sung duration. Across both groups, children tended to imitate pitch more accurately than adults when it came to speech stimuli rather than song stimuli, whereas adults made fewer time errors than did children in both stimulus types.

In terms of absolute feature matching during vocal imitation, we discovered impaired performance in the ASD group for both pitch and duration across both speech and song conditions as compared to the control group. This finding is in line with previous results showing impaired imitation of form in ASD (Edwards, 2014). A few previous studies also showed impaired pitch and duration imitation for speech in ASD (Fosnot & Jun, 1999; Hubbard & Trauner, 2007). However, other studies indicated that speech imitation deficits in ASD only manifested in duration (J. J. Diehl & Paul, 2012; Paul et al., 2008). The discrepancy may be related to the different methods used to measure imitation performance across the studies. While we compared group differences in imitation by measuring how well participants in each group matched the pitch and duration features of the model utterances, previous studies ignored the model but compared the pitch and duration patterns of the produced utterances across groups (J. J. Diehl & Paul, 2012; Paul et al., 2008). Thus, as in previous vocal imitation studies (Liu et al., 2013; Mantell & Pfordresher, 2013), we measured imitation abilities by comparing acoustic parameters between the model and imitated utterances, and the smaller the difference, the more accurate the imitation. Using this method, we were able to reveal differences in absolute feature matching during imitation between groups. However, previous studies only showed the differences in characteristics between the produced utterances of the two groups (J. J. Diehl & Paul, 2012; Paul et al., 2008), thus measuring speech production, rather than imitation accuracy.

In contrast to the intact musical imitation abilities reported in a previous study on children with ASD (Applebaum et al., 1979), our finding demonstrated that both children and adults with ASD were impaired in absolute pitch and duration matching for song imitation. One explanation for this discrepancy may be related to how the accuracy of imitation was calculated. Specifically, Applebaum et al. (1979) relied on subjective perceptual ratings of imitation accuracy by two independent observers, whereas the current study employed objective acoustic analyses. A second possible explanation relates to the difference in sample size. While 88 participants (44 per group) were involved in the present study, only 6 individuals participated in Applebaum et al.'s (1979) study (3 per group). Thus, the current results may be more reliable given the objective acoustic analyses and a larger sample size.

Despite impaired absolute pitch and duration matching, individuals with ASD showed comparable performance to controls on relative pitch and duration matching, as well as on other measures of relative-feature matching (e.g., number of pitch contour, pitch interval, and time errors). Our results are consistent with previous findings on poor singers (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009). For instance, Dalla Bella and Berkowska (2009) examined occasional singers’ pitch and duration accuracy in terms of both absolute and relative features when spontaneously producing well-known melodies, as well as when imitating these melodies with a metronome at a slower tempo. They found that poor singers performed less accurately in the absolute measures than the relative measures and suggested that the production of relative and absolute pitch and time features may be independent in the music domain. Our results extend those of Dalla Bella and Berkowska (2009), showing that the dissociation between relative and absolute pitch and duration matching is also the case for impaired vocal imitation in ASD and that the dissociation exists not only in music but also in speech.

However, to the best of our knowledge, no previous studies in ASD have examined relative versus absolute feature or relative feature matching alone in either speech or music imitation in ASD, which makes it difficult to find evidence to explain why individuals with ASD showed preserved relative but impaired absolute pitch and duration matching during vocal imitation. We propose two possibilities for the divergent results of absolute versus relative feature matching in ASD below, which would require further investigations by future studies. First, one possibility might relate to the differential requirement for fidelity of imitation between absolute and relative features. There has been extensive evidence from non-vocal studies (e.g., action, objects, and face) suggesting that individuals with ASD manifest impaired imitation ability in tasks that require high fidelity imitation, such as reproducing precisely both the form and the end result of a model (Edwards, 2014). However, tasks requiring lower fidelity, such as emulation that only requires reproducing the final result/goal without considering the forms needed to achieve the final goal, generally fail to observe deficits in the ASD group (Edwards, 2014; Hamilton, 2008). In our study, the absolute measures required higher fidelity imitation compared to the relative measures. In particular, absolute measures examined the exact matching of pitch and duration features for each syllable/note, while relative measures assessed the matching of the relative pitch and timing relationship between two consecutive syllables/notes. Thus, our current results indicate for the first time that, consistent with non-vocal imitation studies (Edwards, 2014; Hamilton, 2008), individuals with ASD show impaired vocal imitation ability in tasks requiring high fidelity (i.e., absolute feature matching), but not in tasks requiring lower fidelity (i.e., relative feature matching). Second, it is possible that the imitation mode (relative vs. absolute) that participants were experiencing during vocal imitation may account for the dissociation. Specifically, evidence from perception research in TD indicates that, as children mature from 3 to 6 years, there is a general developmental shift from an absolute to relative mode in pitch perception (Crozier, 1997; Saffran & Griepentrog, 2001; Sergeant & Roche, 1973; Takeuchi & Hulse, 1993). Studies also found that while adults relied primarily on relative pitch cues, they were able to access absolute cues under certain conditions (Saffran & Griepentrog, 2001), and both children and adults demonstrated absolute memory of familiar melodies (Levitin, 1994; Schellenberg & Trehub, 2003, 2008). Taking these findings together, it is possible that different participants may depend on different perception modes when imitating speech and song. The reason that the two groups did not differ significantly in relative pitch and duration matching may be because participants all tended to the relative cues. While controls also accessed absolute cues during the process, participants with ASD did not or were less capable of doing so. Future studies are required to test this possibility by examining the relationship between perception and production during vocal imitation.

When using acoustically matched speech and song stimuli testing the same sample of participants, we observed impairments (i.e., absolute pitch and duration matching) as well as preserved skills (i.e., relative pitch and duration matching) in ASD not only in speech but also in music. Hence, compatible with the findings of vocal imitation in people with typical development (Mantell & Pfordresher, 2013) and those with congenital amusia (Liu et al., 2013), vocal imitation also constitutes domain-general mechanisms in individuals with ASD. These findings provide support for using music therapy to improve speech for those individuals with ASD who manifest deficits in language (James et al., 2015). In addition, successful imitation requires sensorimotor, cognitive and social skills (Fridland & Moore, 2015; Heyes, 2001; N. Nguyen & Delvaux, 2015; Over & Carpenter, 2013; Pagliarini et al., 2020). Thus, the benefit of music imitation may extend to improving cognitive and social skills in ASD (Boster et al., 2020).

It has been reported that absolute pitch (AP) ability is more common among individuals with ASD than in non-clinical populations (Heaton et al., 1998; Mottron et al., 1999; Stanutz et al., 2014). However, the present imitative results were not in line with these findings. Rather, we found that individuals with ASD showed impaired absolute pitch and duration matching. While we did not test our participants' receptive AP ability in the current study, we did ask whether they have absolute pitch (or perfect pitch, the ability to identify a musical note without a reference tone) in a questionnaire. According to the self-reports, two children with ASD (out of 25) and two adults with ASD (out of 19), as well as two control children (out of 25) possessed AP. However, they did not perform exceptionally when imitating absolute pitch, which suggests that receptive AP may not transfer to expressive AP in imitation. This finding is consistent with the dual-route model, which posits that vocal stimuli are processed for motor-relevant features and conscious, symbolic representations along two different, independent pathways (Hutchins & Moreno, 2013). Thus, vocal perception and production abilities could be uncorrelated, and each can be intact or impaired without affecting the other (Griffiths, 2008; Hutchins & Moreno, 2013; Loui, 2015). Notably, our findings are based on self-reports rather than experimental testing of AP. Further studies are needed to clarify the nature of receptive and productive AP in ASD.

Moreover, the current study examined whether speech and song imitation abilities vary with age in ASD and controls. Across both groups, adults made fewer time errors in both speech and song imitation relative to the child cohort. Time errors were defined as deviation from the target duration by 25%, and this is the only measure where the two age cohorts differed significantly, but not in other timing matching measures (e.g., absolute and relative duration matching). These results suggest that while both children and adults can imitate the duration of speech and song segments comparably, children may have greater duration variability than adults when errors were measured relative to the duration of each segment. The findings are in agreement with previous research indicating that there is a developmental decrease in duration variability (Kent & Forner, 1980; Munson, 2004; B. L. Smith, 1978). Indeed, children possess less refined neuromotor capabilities than adults (B. L. Smith, 1978), and they are unable to exert adult-like control of speech production mechanisms. Hence, children’s output reflects greater variability of phonetic segments compared to adults (Kent & Forner, 1980; Koenig et al., 2008; Munson, 2004; B. L. Smith et al., 1996).

Conversely to what was observed in timing matching, across both groups, children tended to imitate absolute and relative pitch more accurately than adults when it came to speech stimuli rather than song stimuli. This result may be due to children attending to speech pitches more readily than adults. Speech imitation is based on intentional understanding (Over & Gattis, 2010). Individuals thus tend to imitate the functional goal (e.g., statements with falling pitch contours vs. questions with rising pitch contours) rather than copying the exact form of the utterances (Liu et al., 2010, 2013). In the present study, this tendency appeared more pronounced in adults than in children. Indeed, we did not find any differences between the child and adult cohorts in the pitch contour imitation, as they all preferred to and were able to imitate the functional goals (rising vs. falling). However, adults neglected form-related information in speech more saliently, resulting in poorer performance than children on exact matching of absolute and relative pitch. On the other hand, the results could also mean that children do not make as strong distinctions between speech and song as adults do. Studies have shown that, unlike musical communication, speech comprehension is remarkably robust to lack of detail in pitch variation (Liu et al., 2015; A. D. Patel, 2011; A. D. Patel et al., 2010). This is because the need for pitch precision in speech can be relaxed by integrating multiple context-based cues (including the voice onset time, vowel length, fundamental frequency, and first and second formant patterns) and knowledge sources (including semantics, syntax, and pragmatics) (Mattys et al., 2005; Toscano et al., 2010; Toscano & McMurray, 2010). However, since the integrating abilities in children are not as mature as adults (McCreery & Stelmachowicz, 2011; Stelmachowicz et al., 2000), they may still mainly rely on pitch cues in speech imitation as they do in music imitation.

Generally speaking, we did not observe the developmental increase in imitative abilities that has been suggested by previous studies in speech (Kent & Forner, 1980; Loeb & Allen, 1993; Snow, 1998) and music (Cohrdes et al., 2016; Cooper, 1995; Geringer, 1983), except in the duration variability. One explanation for this discrepancy may be related to differences in age of the participants among the studies. The youngest child participant in the present study was 7.39 years old, whilst several previous studies examined the development from 3 to 5 years (Loeb & Allen, 1993; Snow, 1998) or 5-7 years (Cohrdes et al., 2016). Thus, it is possible that the present task was too simple to reveal the developmental change for participants beyond 7 years old, since 5-year-olds were already able to imitate falling versus rising contours (Loeb & Allen, 1993; Snow, 1998). In addition, the different grouping of age cohorts between studies might also account for the discrepancy in findings. Specifically, we grouped participants below 16 into the child cohort and those above 16 into the adult cohort, and age-related differences were then examined by comparing these two age cohorts. However, previous studies compared age-related differences at year-level (Cooper, 1995; Geringer, 1983; Kent & Forner, 1980; Loeb & Allen, 1993; Snow, 1998), e.g., comparing 5 years with 4 years (Loeb & Allen, 1993; Snow, 1998). Thus, subtle developments over time may be masked in the present study, given the wide age range within each age cohort. Across our pre-defined age cohorts, however, there was no significant age × group interaction on any of the absolute or relative pitch or duration measures we examined. This suggests that age (>= 16 or < 16 years) influences speech and music imitation similarly across ASD and TD. Thus, our results on vocal imitation corroborate previous findings of persistent impairments in other areas of imitation across the lifespan in ASD (Biscaldi et al., 2014; Vivanti & Hamilton, 2014; Young et al., 2011).

Consistent with previous studies (Liu et al., 2013; Mantell & Pfordresher, 2013; Wisniewski et al., 2013), both the ASD and control groups imitated song more accurately than did speech across all pitch-related measures. Several possibilities may explain this result. First, the reason for the enhanced pitch imitation in songs may be because, in order to achieve adequate communication, a higher degree of pitch precision is required for conveying musical meaning than speech meaning (A. D. Patel, 2010, 2011). Indeed, even individuals with congenital amusia imitated musical pitch better than linguistic pitch, since music is form-driven and speech is function-driven (Liu et al., 2013). Studies of intonation imitation among typically developing adults also suggest that English speakers tended to imitate the phonological structure (e.g., pitch accent, intonational phrase boundary), rather than the phonetic details (e.g., pause duration, irregular pitch periods), of intonation (Cole & Shattuck-Hufnagel, 2011). Thus, the worse pitch matching in speech imitation compared to song imitation may be because people tend to imitate the functional goal, rather than the exact form, of speech utterances (Liu et al., 2010). Secondly, one may argue that the slower tempo in songs might have positively affected pitch imitation, since singing accuracy improves considerably when people sing at slower as opposed to faster tempos (Dalla Bella, Deutsch, et al., 2007). However, even when durations were equated across speech and song stimuli, the pitch imitation advantage for singing remained (Mantell & Pfordresher, 2013). Thus, the enhanced sung pitch imitation cannot simply be attributed to differences in the rate of speech versus song stimuli in the present study.

Interestingly, our results on duration matching across speech and song imitation indicate that the effect of domain is not equivalent across different duration measures. Whereas both groups achieved better absolute and relative duration matching in speech than in song, they made fewer time errors in song than in speech. The fewer time errors in song than in speech may not necessarily imply that duration matching during song imitation was superior to speech imitation for both groups. This is because the more time errors in speech than in song may be caused by the higher time precision required in speech. Specifically, given that our time errors were defined using deviation from 25% of the target duration and that song stimuli contained longer target durations than speech stimuli (see Figure 1), the accuracy requirement was higher for speech than for song imitation. The results of more accurate imitation for both absolute and relative duration in speech than in song are consistent with previous findings (Albouy et al., 2020; Liu et al., 2013; Mantell & Pfordresher, 2013). Overall, both groups tended to be particularly sensitive to the appropriateness of duration in speech compared to that in song, and were more sensitive to pitch in song than in speech, suggesting that pitch imitation is independent from the imitation of duration (Dalla Bella, Deutsch, et al., 2007; Dalla Bella et al., 2009; Drake & Palmer, 2000; Mantell & Pfordresher, 2013).

A caveat about the design of the current study is the stimuli we used for participants to imitate in our experiment. Taken from Mantell and Pfordresher (2013), the stimuli for the adult male and female versions were produced in American English with a midland dialect (male speaker) and an inland North dialect (female speaker). The child version was created by a child imitating the female model in British English. Since all our participants were British English speakers, one may wonder whether or to what extent the different dialects have affected imitative performance. However, research has shown that speakers can imitate detailed intonational patterns of a different variety of their language (D'Imperio et al., 2014). Similar to controls, individuals with ASD who have good language abilities can perceive acoustic differences in dialects as well as use these cues to group the dialects into the areas they come from (Clopper et al., 2012). Consistent with these findings, children, female and male adults with ASD and their TD counterparts in our study performed comparably in imitation of relative pitch and duration, suggesting that different dialects may not have affected imitation performance of our participants. In addition, impairments of absolute pitch and duration matching in ASD were observed not only in the adult cohort but also in the child cohort who imitated British English. Taken together, while the effect of dialect is unlikely to have influenced the current results, further studies are required to examine this specific hypothesis.

3.5 Conclusion

Using sentences and melodies that shared critical features, this study revealed for the first time that vocal imitative skills in individuals with ASD are impaired in absolute pitch and duration matching but intact in relative pitch and duration matching across speech and music domains. From children to adults, vocal imitation showed an improvement in the number of time errors across speech and song, but a decrease in pitch imitation accuracy in the speech condition only. These findings support the idea that speech and song imitation may involve shared cognitive and motor mechanisms, which may have implications for the development of language in individuals with ASD (Stone & Yoder, 2001).

Chapter 4. Study 3: Mental representations of speech and music pitch contours in individuals with and without autism spectrum disorder

Abstract

Studies using standard behavioural measures to investigate pitch processing across different auditory domains (speech versus music) in individuals with ASD have produced mixed results. Here, we investigate a different aspect of pitch processing ability—mental representation of pitch. By using a novel data-driven method, the reverse-correlation paradigm, we examined the mental representation of pitch contours in speech and music in 13 children with ASD and 17 matched controls. The results indicate that the two groups exhibited similar representations in both speech and music conditions, though there were some slight differences in the overall shape, particularly for the speech condition. The two groups did not differ significantly in their internal noise (a measure of the robustness of participant responses to external variability), indicating that our findings of their pitch contour representations reflect their genuine ability in representing pitch. The current findings uncover for the first time how ASD affects mental representations of pitch patterns in both speech and music, and reveal that individuals with ASD represent pitch similarly to those without ASD.

Keywords: ASD, Pitch, Speech, Music, Mental representation, Reverse-correlation

4.1 Introduction

Autism spectrum disorder (ASD) is a neuro-developmental disorder defined by atypical development in the areas of social communication, social interaction, and restricted and repetitive behaviours and interests (American Psychiatric Association, 2013). ASD is typically diagnosed in childhood and has a wide range of symptoms, each of which may differ in their severity across different individuals. One instance of this is communication: some children with ASD communicate verbally, some nonverbally, and others a combination of both (e.g., they have very few words and may supplement their verbal communication using specialist language assistance software or the Picture Exchange Communication System). Impairments in language and communication may have varying manifestations (Eigsti et al., 2011), but critically, far from impairment, many individuals with ASD demonstrate exceptional musical abilities, including extraordinary musical memory, and increased sensitivity to musical pitch (Heaton, 2009).

The dissociable ability to process language and music in some individuals with ASD has attracted considerable attention from researchers in an attempt to understand whether this dissociation is a general characteristic of the ASD population (J. Jiang et al., 2015; Lai et al., 2012; Sharda et al., 2015). In addition, the findings of whether individuals with ASD show impaired speech but enhanced music processing are particularly relevant to an ongoing debate about whether speech and music share the same underlying processing systems (Albouy et al., 2020; Norman-Haignere et al., 2015; Zatorre & Gandour, 2008). Some researchers have proposed a modular or domain-specific framework (Fodor, 1983, 2001; Peretz, 2009; Peretz et al., 2015; Peretz & Coltheart, 2003; Peretz & Zatorre, 2005), which emphasizes that music processing utilizes modules that are not shared with speech processing. Others have suggested that there are shared or domain-general mechanisms underlying the processing of information across both domains (Koelsch, 2011; Koelsch & Siebel, 2005; A. D. Patel, 2010; Sammler et al., 2009).

Pitch, as a salient acoustic feature shared between the two domains, provides a natural laboratory for comparative studies of language and music. Specifically, pitch not only is a foundational building block of what constitutes melodies and chords (Krumhansl, 2004; Sadakata et al., 2020) but also carries important prosodic and semantic (e.g., in tone languages) information in languages (Bidelman et al., 2011). Most studies, especially the earlier ones, have suggested that individuals with ASD exhibit exceptional musical pitch sensitivity compared with their typically developing counterparts (TD) (O'Connor, 2012; Ouimet et al., 2012). For example, individuals with ASD show enhanced abilities to discriminate (e.g., same or different) and identify (e.g., low or high) pitch in pure tones (Bonnel et al., 2003, 2010; Heaton et al., 1998; O'Riordan & Passetti, 2006). Beyond these simple stimuli, enhanced pitch processing in ASD has also been observed in musical melodies, including identification of melodic pitch contour (J. Jiang et al., 2015) and detection of a local pitch change in a melody (Heaton, 2005; Heaton, Pring, et al., 1999; Mottron et al., 2000; Stanutz et al., 2014). Therefore, enhanced musical pitch processing has been viewed as a characteristic of many individuals with ASD. Nevertheless, typical or even impaired musical pitch processing has also been reported in the literature (Cheng et al., 2017; Germain et al., 2019; Heaton, Williams, et al., 2008; Jamey et al., 2019; Jones et al., 2009; Kargas et al., 2015; Schelinski et al., 2017). For instance, more recent studies investigating pitch discrimination between two tones (e.g., which one is higher) found that after controlling for age and IQ, individuals with ASD performed either similarly to (Jones et al., 2009) or worse than TD (Kargas et al., 2015) at the group level. Still, enhanced pitch discrimination was found in a subgroup of participants with ASD, e.g., 20% in Jones et al. (2009) and 9% in Kargas et al. (2015). Hence, while enhanced musical pitch processing has been observed in ASD, it may only be evident among a subgroup of individuals with ASD.

In contrast to musical pitch processing, pitch-mediated speech processing ability, especially prosodic pitch processing, is typically viewed as a skill that individuals with ASD have difficulty with, including identifying and discriminating questions and statements (J. Jiang et al., 2015), distinguishing lexical stress contrasts (Paul et al., 2005), as well as encoding lexical tones (Lau et al., 2020; Wang et al., 2017). Given that semantic and prosodic information play a crucial role in speech communication and interaction, this atypical pitch processing in speech may hinder language acquisition and development in ASD (Schreibman et al., 1986). However, some studies showed enhanced identification of pitch contours (e.g., rising, failing, rising-falling, falling-rising) and discrimination of pitch differences (e.g., are these two sounds the same?) across speech and musical conditions, suggesting that superior pitch processing in individuals with ASD is not limited to music but also extends to speech (Heaton, Hudry, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007).

In summary, although pitch processing in ASD has been studied widely across music and speech, the findings are mixed at best since enhanced, intact, and impaired pitch processing have been found across domains. Although some apparently contradictory findings may be due to sampling variability, since ASD is a heterogeneous group and high variability in the ASD sample has been found in several areas, including pitch processing (Kargas et al., 2015; Milne, 2011; Valla & Belmonte, 2013), inherent shortcomings in the methods used in previous studies might also contribute to the mixed findings regarding pitch processing abilities in ASD.

Specifically, previous studies only employed standard behavioural measures (e.g., discrimination, identification, etc.), which are, while informative, normally considered as hypothesis-driven studies. In those studies, the signal features (e.g., pitch) that drive judgments are limited regarding diversity across individuals and are postulated by the experimenter before being tested experimentally, which is very likely to create a variety of confirmation biases (Burred et al., 2019). For instance, prosodic cues in speech stimuli are normally generated by one or a few individuals (speakers) and then selected/confirmed by experimenters, but these cues may not exhaust the many other ways in which individuals express prosody and may not match individuals' internal representations. Additionally, the stimuli used in these studies generally stem from a limited number of utterances or corpora (e.g., from 20 to 100 stimuli), which may not be sufficient to capture sensitivity to subtle signal changes due to the likelihood of producing coarse variation (Ponsot et al., 2018). For instance, the stimuli used in one of the aforementioned studies for discrimination and identification tasks consisted of four standard tones at 500 Hz, 700 Hz, 1000 Hz, and 1500 Hz respectively, and the comparison tones were presented with either the same frequency (i.e., the same condition) or 1%, 2% and 3% higher frequency (i.e., the different condition), in which case, each standard tone, such as 1000 Hz, only had three comparison tones with 1010, 1020, and 1030 Hz (Bonnel et al., 2003).

To overcome these shortcomings, the present study explored a novel data-driven method—the reverse-correlation paradigm. In this paradigm, participants' mental representations of signal features (i.e., pitch) that drive judgments can be mathematically determined and visually presented by analysing participant responses to large sets of systematically varied stimuli (Burred et al., 2019). In particular, participants were presented with a large number of pairs of stimuli (e.g., 800 stimuli), with each stimulus having a randomly manipulated pitch contour (e.g., adding Gaussian pitch noise with a normal distribution), and they were asked to make a judgment (e.g., "Which of the two sounds more like a rising tone?") on each pair through a forced-choice task. Through the use of many stimuli and by obtaining the difference in pitch contours between the stimuli that were chosen and not chosen, the tool of reverse correlation can be used to determine participants' mental representations of pitch contours from the pattern of stimulus noise (the stimuli were not chosen) and the associated responses (the ones that were chosen) (Adolphs et al., 2016; Jack & Schyns, 2017; Ponsot, Burred, et al., 2018).

To the best of our knowledge, no published studies have used a reverse-correlation paradigm exploring how pitch is represented in individuals with ASD. To fully understand pitch processing ability in individuals with ASD and tease apart the mixed findings in the literature, it is important to examine the mental representation of pitch, since there seems to be a top-down bias in which representations affect perception (e.g., discrimination, and identification). Studies using the reverse-correlation paradigm from the visual domain have found evidence of how top-down biases affect social perception (Brinkman et al., 2017; Imhoff et al., 2013; Ratner et al., 2014). For instance, social attitudes biased Dutch participants' decisions in evaluating faces of Moroccans, as the more negative the attitude of the participants to Moroccans, the more criminal and less reliable their representations of Moroccan faces (Dotsch et al., 2008). In addition to discovering how pitch is represented in individuals with ASD, examining mental representations of pitch in speech and music also helps inform the theoretical debate on whether speech and music share underlying mechanisms.

Therefore, the present study used matched music and speech stimuli to explore how the pitch contour in each domain was represented by individuals with and without ASD. Given that there is no objective criterion on which to base their representations (i.e., there is no correct definition of what constitutes a 'correct' rising tone), it is not possible to compare the groups on how accurate their representations are. We can, nonetheless, compare whether the groups differ in their representations across domains. Based on previous research suggesting that individuals with ASD show atypical speech pitch perception and enhanced musical pitch perception relative to individuals without ASD, we hypothesised that there will be subtle group differences in how they represent pitch in both speech and music domains.

4.2 Methods

Participants

A group of 13 children with ASD and 17 matched controls were recruited via a variety of methods including email lists, local social media advertisements, and local experimental participant databases. All were native British English speakers. Participants in the ASD group had a formal diagnosis of ASD by professional clinicians, which was also confirmed using the Autism Spectrum Quotient (AQ) cut-off score of 76 for children (Auyeung et al., 2008; Baron-Cohen et al., 2001, 2006). Participants in the control group scored below the cut-off score. All participants had normal hearing in both ears, with pure-tone air conduction thresholds of 25 dB HL or better at frequencies of 0.5, 1, 2, and 4 kHz. Participants completed a nonverbal IQ test using the Raven's Standard Progressive Matrices Test (Raven et al., 1998) and a receptive vocabulary test using the Receptive One Word Picture Vocabulary Test IV (ROWPVT-IV) (Martin & Brownell, 2011). The Corsi block-tapping task was used to assess participants' nonverbal short-term memory span (Kessels et al., 2000) and the forward digit span task was used to assess verbal short-term memory (Wechsler, 2003). As shown in Table 1, the groups did not differ in their age, years of musical training, or cognitive abilities.

Table 1 Characteristics of the ASD (n = 13) and TD groups (n = 17)

Variables	**ASD**	**TD**	***W***	***P***	***Rank-Biserial Correlation***
Age					
Mean (SD)	12.36(2.65)	11.85 (2.14)	122	0.65	0.10
Musical training					
Mean (SD)	1.73 (1.75)	2.88 (2.91)	86.5	0.32	0.22
NVIQ					
Mean (SD)	65.39 (28.39)	76.77 (24.11)	82	0.23	0.26
ROWPVT-IV					
Mean (SD)	123.62 (13.84)	125.24 (16.43)	98.5	0.63	0.11
Corsi					
Mean (SD)	4.92 (0.86)	5.53 (1.23)	90.5	0.34	0.18
Digit span					
Mean (SD)	5.31(0.75)	5.82 (0.88)	72	0.09	0.35

Note: Musical training: years of musical training; NVIQ: percentile Point of Raven's Standard Progressive Matrices Test; ROWPVT-IV: standard score of Receptive One Word Picture Vocabulary Test; Corsi: raw score of nonverbal short-term memory, Digit span: raw score of verbal short-term memory; 2-tailed Mann-Whitney U Test results were used to compare group difference and effect size was given by the rank biserial correlation in the Mann-Whitney test.

Stimuli

There were two types of auditory stimuli: speech and music. For the speech stimuli, a single word /mi/, which sounds like "me" in English, with a level tone was recorded by a female adult speaker. The original sound was manipulated to last 250 ms and the intensity set at 80 dB using Praat. Finally, the original pitch contour was flattened to its mean pitch (210 Hz). Following a previous study, a Python-based toolbox (CLEESE; see Burred et al., 2019 for details) was used to generate variations of the sound with randomly manipulated pitch contours while maintaining a constant amplitude and duration. Specifically, Gaussian pitch noise (i.e., pitch-shifting) was added to the contour by sampling pitch values at eight successive time-points, using a normal distribution (SD = 70 cents; clipped at ± 2.2 SD). These values were linearly interpolated between time points. After piloting (see below), a total of 800 speech stimuli were synthesized.

For the music stimuli, Praat was used to generate a complex tone analogue of /mi/. The complex tone comprised of F0 (fundamental frequency) and its seven odd harmonics, of the same amplitude and with sine phase, which leads to a clarinet sound quality (Liu et al., 2010; A. D. Patel et al., 1998, 2005, 2008). In keeping with the acoustic characteristics of the speech sound /mi/, the pitch value of the complex tone was set at 210 Hz, intensity at 80 dB, and duration at 250 ms. Thereafter the same procedure was applied as with the speech stimuli to manipulate the pitch contour of the complex tone. After piloting, 800 complex tones with different pitch contours were generated.

Piloting

Reverse correlation experiments typically have many trials in order to obtain reliable results (e.g., Ponsot et al. (2018) had 500 trials and Park et al. (2017) used 480 trials).

Given that participants would need to complete both conditions (speech and music), in order to obtain reliable results while keeping the duration of each condition feasible, a pilot study was conducted to determine the optimal number of trials for each condition. Burred et al. (2019) showed that reliable results (defined as r = 0.8 and above, for the correlation of pitch values between a subset of trials and the full set of trials) were obtained by most participants within 100 trials, though some needed up to 300 trials. Based on that, we conducted our pilot study using 1000 stimuli in each condition. These stimuli were randomly paired in each condition, resulting in 500 trials per condition, and three typically developing children made judgments on each pair per the requirement of the condition (e.g., in the music condition, participants judged which of the pair had a rising pitch). Each task took approximately 40 mins with short breaks. Using a similar approach to Burred et al. (2019), preliminary data from these three participants showed that reliable results were obtained with approximately 300 trials for both conditions (see Figure 1). Given these findings, it was determined that 400 trials in each would be appropriate for the main experiment.

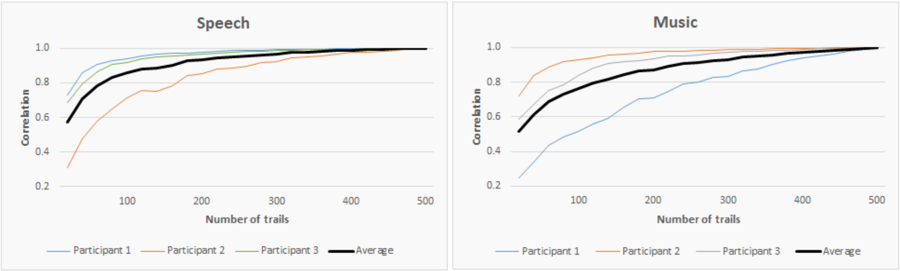


Figure 1. Correlations between the contours derived using the n first trials and the kernels derived using all trials (n = 500) in speech and music tasks.

Procedure

As in the pilot, the main experiment consisted of two separate conditions. In the speech condition, participants listened to pairs of randomly modulated /mi/ and were asked to indicate (using a button press) which of the two sounded the most interrogative. The music condition was similar to the speech condition, except that participants heard complex tones and chose the one that best matched a rising tone, which was indicated by an arrow on the screen (↗). The presentation order of both conditions was counterbalanced across participants, and the stimuli pairings in each condition were randomized across participants. The inter-stimulus interval in each trial was 500 ms, and the inter-trial interval was 1s. Participants were given a choice to take a self-timed break after every 100 trials in each condition.

Data analysis

For each participant, we computed a first-order temporal kernel (Ahumada & Lovell, 1971) separately for each condition. The kernel was defined as the mean pitch contour difference between the stimuli that were chosen as the best match in each pair and those that were not chosen, at each time point across trials for each condition. For example, in a speech trial of two "/mi/"-like sounds with different pitch contours, the pitch values (e.g., measured at eight successive time points) of the chosen stimulus minus the corresponding pitch values of the one not chosen were calculated for this trial. Then the kernel was obtained by averaging these values across all trials.

Using these kernels, group differences in participants' mental representations of the pitch contours were compared in several different ways. First, a two-way repeated measure ANOVA was employed to compare group differences at each time point of the temporal kernel. Second, in order to capture any potential differences in shape of the kernel between groups for each condition, linear and nonlinear (quadratic, cubic, and quartic) models were fit to the kernels of each group and condition. The models were compared using likelihood-ratio tests to determine which model/shape provided the best fit for the data. This was done to determine whether the groups differed in the overall shape of their temporal kernels for each condition. For example, for the music condition, if a linear model fitted the data of the ASD group best whereas a cubic model fitted the TD data best, then this would suggest a group difference in the overall shape of the kernels that might not have been revealed in the relatively conservative ANOVA model. Finally, after determining the best model for each group and condition, the appropriate model for each participant by group and condition was obtained by computing two individual-level parameters—y-intercept and slope. Group differences in these parameters were then compared using two sample *t*-tests.

To assess the energy of the kernels obtained, two further parameters were computed by participants and conditions. First, the root-mean-square (RMS) value was calculated for each participant’s kernel under each condition, which is a scalar of pitch perceptual filter that reflects how much people weight the different pitch portions in one direction or the other (Ponsot et al., 2021). The higher the RMS, the more sensitive the participant is to the pitch. Groups were then compared on their RMS values for each condition using two sample *t*-tests. Second, internal noise, which reflects non-systematic variations in participants' perceptual responses, was measured to rule out the possibility that any group difference in their mental representations was due to random variations in their responses (Burred et al., 2019; Venezia et al., 2019). For each participant, it was first determined which of the pair on each trial was the most similar to the individual's kernel (i.e., the individual's mental representation) for each condition using correlation analysis. The stimulus with the higher correlation coefficient in each pair was defined as the objectively correct response on each trial. Next, the percentage of the participant's actual responses that agreed with those of the objectively correct responses using correlation analysis was computed, with the assumption that this percentage would be an index of internal noise (the higher the value, the less the internal noise). Groups were then compared on their internal noise for each condition using two sample *t*-tests. These two parameters were used to obtain how sensitive individuals were to pitch while also estimating the extent to which random responses contribute to representation. All statistical analyses were conducted using RStudio (RStudio Team, 2020). Subsequent post-hoc comparisons, if any, were conducted using the *emmeans* package (Lenth et al., 2018).

4.3 Results

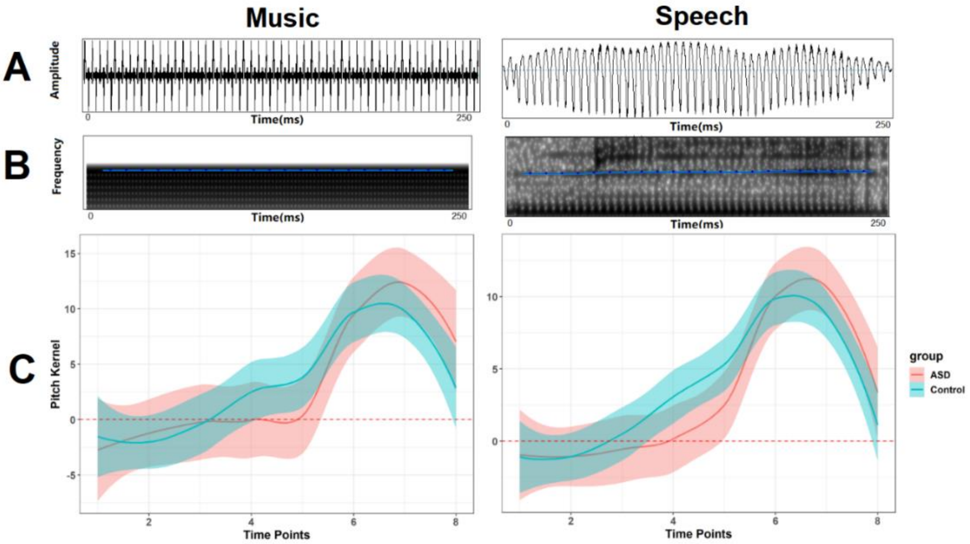


Figure 2. The waveform (A), spectrogram and level contour (blue line) of the original sound (B) and the pitch kernels of the rising pitch contour (C) for music and speech among ASD and TD (Shaded areas show 95% Confidence Interval).

Comparison at each time point

Figure 2 displays the kernels for ASD and TD at the group level. The pitch values of the kernels were subjected to a three-way repeated measure ANOVA using Group (ASD vs. TD) as a between-subjects factor, and Timepoint (8 successive time-points) and Stimulus type (speech vs. music) as within-subjects factors. The model revealed a significant main effect of Timepoint (F(7,196) = 23.68, $p < .001$) as well as a significant interaction between Timepoint and Stimulus type (F(7,196) = 7.87, $p < .001$). To explore this interaction, paired *t*-tests with Bonferroni correction between the pitch values of speech vs. music conditions at each of the eight timepoints were conducted. Results indicated that the pitch values in the speech condition were significantly higher than those in the music condition at timepoint 6 (Speech: M = 11.08, SD = 12.59 vs. Music: M = 6.58, SD = 7.26; t(224) = -3.32, $p = 0.001$) whereas significantly lower than those in the music condition at timepoint 7 (Speech: M = 9.8, SD = 11.24 vs. Music: M = 17.95, SD = 16.27; t(224) = 5.92, $p < 0.001$) and timepoint 8 (Speech: M = -1.1, SD = 3.84 vs. Music: M = 2.21, SD = 6.03; t(224) = 2.44, $p = 0.02$). There were no significant main effect or interactions involving Group.

Comparison of the overall shape

Linear and nonlinear polynomial models (quadratic, cubic, and quartic) were fitted to each group and condition and the best-fitting model was selected using likelihood ratio tests. We found that a quartic model was the best-fitting model in most cases except the speech data for the TD group, and the results are shown in Table 2.

Table 2. Results of likelihood-ratio tests to determine the best-fitting model by group and condition.

Condition	**Group**	**Comparison**	**DF**	**F**	***P***	**The better fitting model**	**The best fitting model**
Speech	ASD	Linear & Quadratic	1	1.75	0.19	Linear	**Quartic model**
		Linear & Cubic	2	11.69	< .001	Cubic	
		Cubic & Quartic	1	4.27	0.04	Quartic	
	TD	Linear & Quadratic	1	12.61	< .001	Quadratic	**Cubic model**
		Quadratic & Cubic	1	17.35	< .001	Cubic	
		Cubic & Quartic	1	2.65	0.11	Cubic	
Music	ASD	Linear & Quadratic	1	0.06	0.81	Linear	**Quartic model**
		Linear & Cubic	2	3.29	0.04	Cubic	
		Cubic & Quartic	1	19.22	< .001	Quartic	
	TD	Linear & Quadratic	1	2.54	0.11	Linear	**Quartic model**
		Linear & Cubic	2	10.97	< .001	Cubic	
		Cubic & Quartic	1	5.04	0.03	Quartic	

After determining the best models to fit the data by condition and group, the R^2 was computed for each model to ensure that the groups did not differ significantly in the proportion of the variance explained by the model fit for each condition for a fair comparison in their mental representation. Results showed no significant difference between the ASD and TD groups in their R^2 either in the speech condition (t(25.89) = 1.34, *p* = 0.19) or the music condition (t(27.97) = 0.79, *p* = 0.44). To compare whether there were any subtle differences in the shapes between groups for each condition, two model parameters—y-intercept (b0) and slope of tangent at midpoint—were extracted from the models fitted to each participant. In the speech condition, no significant group differences in either y-intercept (ASD: M = 3.01, SD = 4.57 vs. Control: M = 3.61, SD = 3.57; t(22.2) = -0.39, *p* = 0.70) or slope (ASD: M = -653.89, SD = 677.72 vs. Control: M = -514.26, SD = 725.10; t(26.80) = -0.55, *p* = 0.59) were found, which indicates that the two groups exhibited similar morphology for the speech kernels. However, in the music condition, there was a significant group difference in slope (t(16.36) = 4.38, *p* < 0.001) but not in y-intercept (t(25.40) = 0.16, *p* = 0.88). The ASD participants (M =

19.43, SD = 48.53) showed a positive-going slope of the tangent at midpoint for the music kernels, whereas controls showed a negative-going tangent slope (M = -541.97, SD = 526).

RMS values and internal noise

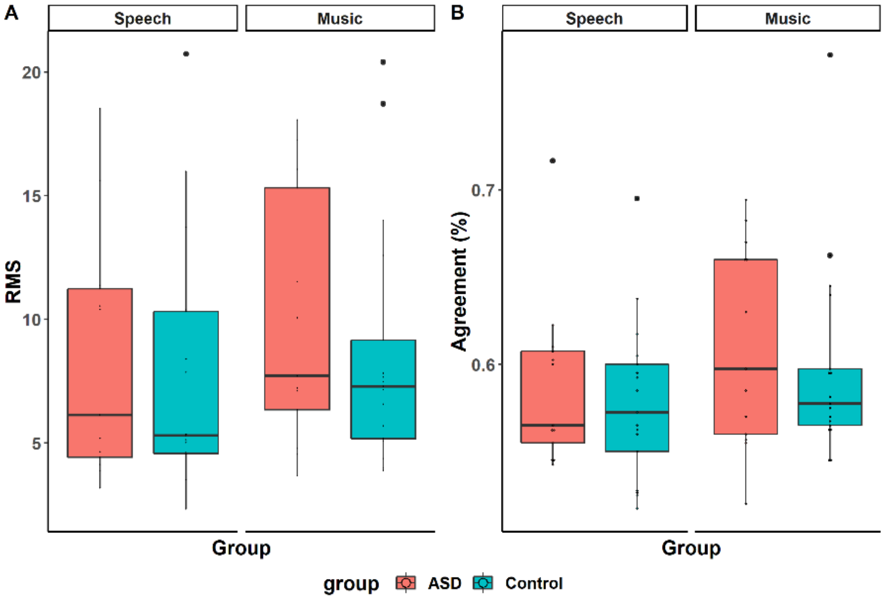


Figure 3. The result of RMS-values of perceptual filters (A) and agreement percentage, an index of internal noise (B).

Figure 3 displays the results of RMS values (as measured using the kernel of each participant by conditions) and the internal noise (as measured using the percentage of agreement between the objectively correct response as dictated by the participant's kernel and their actual response on each trial) by each group and condition. Two sample *t*-tests showed that neither the RMS values nor the internal noise were significantly different between the groups in each of the conditions (all *p*s > .05, see table 3). There were significant positive correlations between the RMS values and the agreement percentage in both groups under speech and music conditions (all *p*s < .05, see Fig. 4), which is in line with theoretical expectations (R. F. Murray, 2011; Ponsot et al., 2021) indicating that the more random the response variation, the less sensitive the person is to pitch in the estimated kernel.

Table 3. The results of two sample t-tests.

		Mean (ASD:TD)	**SD** (ASD:TD)	***t***	**df**	***p***
Speech (ASD vs. TD)	RMS	8.4:7.78	4.92:5	0.34	26.17	0.73
	Agreement	0.59:0.58	0.05:0.05	0.51	25.35	0.61
Music (ASD vs. TD)	RMS	9.97:8.64	5.14:4.95	0.71	25.46	0.48
	Agreement	0.61:0.60	0.06:0.06	0.61	25.85	0.55

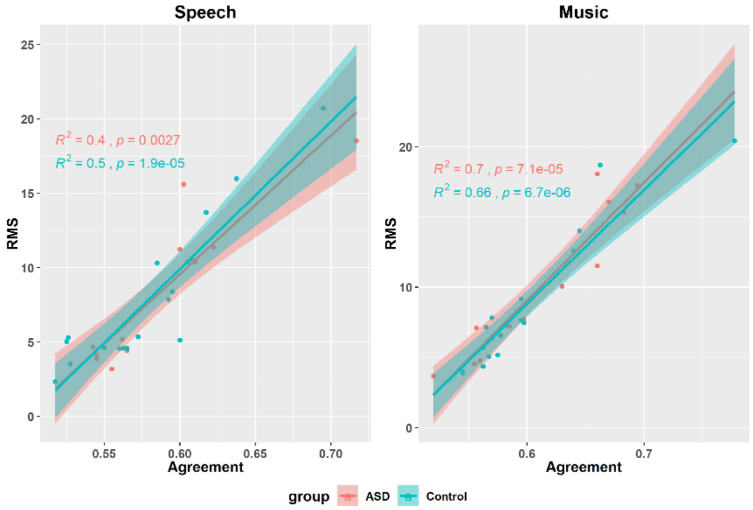


Figure 4. Relationship between RMS-values of the perceptual filters and the agreement percentage (Shaded areas show 95% Confidence Interval).

4.4 Discussion

The present experiment utilized a reverse-correlation paradigm to test a different aspect of pitch processing ability: the mental representation of pitch contour in different auditory domains (speech vs. music) in individuals with and without ASD. Groups were compared in terms of their representation of speech and musical pitch in two ways: (i) by examining group differences at each time point, and (ii) by examining the global shape. The two groups did not differ significantly in their mental representations of pitch contour at any time point across the two domains. However, when considering the global shape of the mental representations, some group differences were observed. Specifically, while participants with and without ASD exhibited similar shape in their representation of musical pitch contours, participants with ASD had positive-going tangent slopes at midpoint (timepoint 4.5) whereas participants without ASD had negative-going tangent slopes, which reflects a later point of inflection within the midpoint region by the latter (see Figure 3). Concerning the speech domain, the two groups had a different overall shape in their representation of the rising pitch contours (quartic shape for the ASD group and cubic shape for the TD group) but they did not differ in the shape parameters (y-intercept and slope). The two groups did not differ significantly in RMS values (pitch perceptual filter) across speech and music, providing evidence for similar pitch sensitivity in both groups across domains. Given that the two groups did not differ significantly in measures of their internal noise, these results are likely to reflect their genuine ability in representing pitch, rather than random variations in their responses.

A comparable mental representation of pitch contour in the music condition between the ASD and control groups is in line with previous findings of unimpaired musical pitch perception in individuals with ASD (Bonnel et al., 2003, 2010; Cheng et al., 2017; Germain et al., 2019; Heaton, 2005; Heaton et al., 1998; Heaton, Pring, et al., 1999; Heaton, Williams, et al., 2008; Jamey et al., 2019; J. Jiang et al., 2015; Jones et al., 2009; Mottron et al., 2000; O'Riordan & Passetti, 2006; Stanutz et al., 2014). As mentioned in the Introduction, previous studies examined response accuracy in musical pitch perception using standard behavioural measures, including discrimination (e.g., same or different) and identification tasks (e.g., low or high). In those tasks, the higher the response accuracy, the better the pitch processing ability. However, these measures cannot reveal how participants represent pitch when processing auditory stimuli. By investigating participants' mental representation of pitch contour, this study found that some individuals with ASD represent musical pitch in a similar way to controls. Additionally, as suggested by the measure of slope at the midpoint, individuals with ASD seem to have a slightly earlier inflection point than controls when responding to rising tones, a finding that is impossible to observe using traditional behavioural tasks.

Consistent with our second hypothesis regarding speech pitch processing, the two groups exhibited different overall shapes of mental representations of a rising pitch contour. However, these differences were not driven by impaired speech pitch processing ability in participants with ASD, given that the two groups had similar perceptual sensitivity to pitch as measured by the RMS. In addition, as shown in Figure 3, it is clear that the two groups can perceive and represent a rising intonation at the end of an utterance, which functions as a question requesting a response (Cruttenden, 1997). It has been reported that when asking questions, speakers do have variable pitch contours in English (Banuazizi & Creswell, 1999; Grabe, 2002; Gussenhoven & Chen, 2000; Kügler, 2003). For example, the rise can be pronounced in as brief as one single syllable or as long as one word or more (Warren, 2005), and the overall shape can be Low-High, High-Low-High, or Low-High-High, and so on (Grabe, 2002). Thus, the differences between groups are more about variations in how they represent questions, rather than driven by impaired pitch processing or question intonation perception in the ASD group.

The intact pitch-related speech intonation processing ability is compatible with previous findings suggesting that individuals with ASD may have no impairment in processing question intonation (Filipe et al., 2014; Paul et al., 2005; Peppé et al., 2007, 2011). However, this stands in contradiction to other findings that individuals with ASD had difficulties identifying questions and statements (J. Jiang et al., 2015). These discrepancies may be explained by several possible reasons, the first being the fundamental differences in methodologies used between the studies. Jiang et al. (2015) used standard behavioural tasks (e.g., discrimination and identification) to investigate whether individuals with ASD can perceive statement-question intonation, whilst the present study used a reverse correlation paradigm to explore the mental representation of pitch contour when individuals with ASD process question intonation. Thus, these two studies were designed to address two different research questions, and the correlation between auditory pitch representation and perception is still unstudied. As mentioned previously, studies that used the same paradigm within the visual domain suggested a top-down bias between representation and perception (Brinkman et al., 2017). We therefore speculated that the fidelity of how well one represents pitch might be associated with behavioural judgements of pitch. For example, a better pitch representation might relate to a more accurate judgement of pitch as measured by standard behavioural tasks. Still, given that the present study did not test behavioural outcomes of pitch processing (e.g., pitch identification/discrimination accuracy or pitch threshold, etc.), the speculation needs to be investigated by future studies applying both the reverse-correlation paradigm and standard behavioural measures on the same participants. In addition, the stimuli used in the present study and Jiang et al. (2015) were also different. The present study used a single word /mi/ in English, whereas Jiang et al. (2015) used disyllabic words in Mandarin. Since Mandarin is a tone language, where the processing of intonation is further complicated by tones embedded in the speech stimuli (Liu & Xu, 2005), it may result in increased task difficulty and lead to poor performance in ASD compared to controls in Jiang et al. (2015). It is worth noting that the difference in results between the present study and Jiang et al. (2015) is unlikely to be the result of a smaller sample size. While the ASD sample size for this study is smaller than that of Jiang et al. (2015) (13 ASD vs. 17 controls in this study, and 17 ASD vs. 17 controls in Jiang et al. (2015)), the number of stimuli used in the present study was much larger than those in Jiang et al. (2015). We utilized 800 distinct stimuli consisting of 400 pairs, which was much more than the 40 pairs with half the same and half different conditions used in Jiang et al. (2015). Nevertheless, the relatively small sample size in general requires us to be cautious and not to over-interpret these results until the present findings are replicated with a larger sample of ASD individuals.

In summary, we assessed how well individuals with ASD are able to represent pitch using acoustically matched speech sounds and musical tones. The primary interest in doing this was to determine whether individuals with ASD would perform differently across domains and to inform the theoretical debate about whether speech and music share the same underlying mechanisms. The two groups did not differ significantly in pitch sensitivity (suggested by RMS values) across both speech and music conditions and the two groups also represented pitch contours as either a rising tone (music) or question intonation (speech) in a similar way. Thus, these results support the view that pitch processing constitutes domain-general mechanisms in ASD, which provides theoretical implications for using musical therapy to improve speech for those individuals with ASD who manifest deficits in speech communication (Eigsti et al., 2011).

4.5 Conclusion

In the present study, a novel reverse-correlation paradigm was used to investigate for the first time how ASD affects mental representations of pitch contours in speech and music. Our findings revealed that the representations were similar across individuals with and without ASD in both domains for the most part, though there were some subtle group differences in their global shape, particularly for the speech domain. This study extends our understanding of pitch processing ability in speech and music, and demonstrates a novel and promising way to investigate auditory processing in ASD. Future studies could explore other populations such as musicians vs. non-musicians, those with and without congenital amusia, or tone vs. non-tonal language speakers.

Chapter 5. General discussion and conclusion

Using both traditional and novel methods, this thesis investigated how individuals with and without ASD decode and encode pitch information across auditory domains (speech vs. music), modalities (perception vs. production) and processing levels (low-level vs. high-level). While a growing number of studies have investigated pitch processing in individuals with ASD, mixed results have been reported with no substantial evidence to inform the ongoing theoretical debates in ASD, such as whether speech and music processing share underlying mechanisms; whether perception ability correlates with production ability; and whether high-level information processing ability is intact in individuals with ASD. Given the current literature, well-designed comparative studies considering not only different sections (e.g., speech vs. music, etc) but also developmental stages (e.g., children vs. adults) might be particularly helpful to inform these debates and to obtain a better understanding of the phenotypes of pitch processing ability in ASD. Therefore, three comparative studies were conducted in this thesis.

Study 1 investigated the ability to process pitch-related statement-question intonation under both speech and music conditions in individuals with and without ASD across developmental stages. The findings suggested that individuals with ASD had intact abilities to differentiate intonation in both conditions across the lifespan, as compared with controls. Study 1 also examined the ability to identify and imitate these intonations in natural speech only. Similarly, these abilities were also intact in individuals with ASD relative to controls, and the abilities to identify and to imitate intonation were associated with each other in both groups. In addition, across both groups, their abilities to process high-level information, including intonation discrimination, identification and imitation, were predicted by their sensitivity to pitch direction. In terms of the age effect, individuals with ASD exhibited similar developmental changes as controls, with performance on both intonation and pitch thresholds increasing with age.

Study 2 examined the ability to imitate speech and song in children and adults with and without ASD. Acoustic analyses of the imitation performances showed that individuals with ASD were worse than controls on absolute pitch and duration matching, while performing as well as controls on relative pitch and duration matching in both speech and song imitation. In addition, the two groups produced similar numbers of pitch contour errors, pitch interval errors and time errors. Across both groups, children tended to imitate pitch more accurately than adults when it came to speech stimuli rather than song stimuli, whereas adults made fewer time errors than did children in both speech and music.

Study 3, using a data-driven method called the reverse-correlation paradigm, investigated pitch perception in speech (representing questioning intonation) and music (with a rising tone) in children with and without ASD. The results indicated that the ASD and control groups showed comparable sensitivity to pitch (as suggested by RMS values) and displayed similar representations of pitch contours in the music condition. Although there were some slight differences between groups in the overall shape in the speech condition, the differences were unlikely to be driven by impaired speech pitch processing ability in participants with ASD, since participants with ASD showed similar sensitivity to speech pitch relative to controls, and their speech pitch contours also exhibited a final raised shape similar to those of controls to represent questioning intonation.

Implications of these results will be discussed in more detail along with the four specific aims that the present thesis sought to address.

5.1 The processing of pitch in speech versus music in ASD

As introduced in Chapter 1, there is a longstanding debate about whether speech and music share the same underlying processing systems. The domain-specific framework emphasizes that music processing utilizes modules that are not shared with speech processing. In contrast, the domain-general framework argues that there are shared or domain-general mechanisms underlying the processing of information across both domains.

5.1.1 Pitch perception in speech and music

From a perception perspective, Studies 1 and 3 investigated how individuals with ASD perceive pitch in speech and music. While Study 1 and Study 3 applied different research methods, these two studies complemented each other and achieved similar results with individuals with ASD showing intact pitch processing abilities in both speech and music. Specifically, Study 1 used a discrimination task to compare the performances between groups according to the accuracy of participant responses. The ASD and control groups did not differ significantly in terms of their response accuracy when processing pitch-related statement-question intonation in either musical glides or speech utterances. These findings suggest that individuals with ASD are able to discriminate pitch-related statement-question intonation under both speech and music conditions. Study 3 utilized the reverse-correlation paradigm to determine the mental representation of pitch contours in speech and music. To mathematically compare the pitch contours, Study 3 applied multiple analysis methods. In particular, when examining group differences at each time point of the pitch contours, no significant group effects or group related interactions were found for both speech and music conditions. When inspecting group differences regarding overall shapes of pitch contours, while the two groups exhibited a similar overall shape under the music condition, different overall shapes were shown between groups under the speech condition. However, the between-group difference regarding overall shape of speech pitch contour was unlikely driven by impaired speech pitch processing ability in the ASD group due to several reasons. Firstly, as suggested by RMS values, the two groups had similar perceptual sensitivity to pitch in both speech and music. Secondly, the two groups did not differ significantly at any time point of the speech pitch contours. Finally, both groups showed a final increase at the end of the syllable, which functions as a question requesting a response (Cruttenden, 1997). Taken together, the results from Study 3 suggest that both groups can use pitch cues to process questioning intonation.

The findings of intact pitch perception across domains in individuals with ASD from Studies 1 and 3 are also compatible with previous studies examining pitch perception using matched music and speech stimuli (Cheng et al., 2017). However, they are inconsistent with others showing enhanced identification/discrimination of pitch contours across domains (Heaton, Hudry, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008; Järvinen-Pasley & Heaton, 2007). This discrepancy may be due to the different stimuli used between these studies. Unlike the present study using stimuli containing prosodic cues (statement vs. question), previous studies used stimuli either pronounced in a neutral way or produced to match four visual shapes (rising, falling, rising-falling and falling-rising). Since prosody reflects elements beyond the literal meaning of the speech itself (Boutsen, 2003), the present study may be relatively more difficult than previous studies and this might be the reason why we fail to observe enhanced pitch processing across domains.

Study 1 found that both groups were better at discriminating between musical glides than between speech utterances. The poorer performance in speech may be caused by the semantic information in the speech, which might hamper individuals' judgment (Järvinen-Pasley & Heaton, 2007). Indeed, Study 1 followed a previous study (Liu et al., 2010) which used the same stimuli to investigate statement-question perception abilities associated with congenital amusia. Consistent with current findings, individuals with congenital amusia also performed better on musical glides than on speech utterances (Liu et al., 2010). However, Liu et al. (2010) failed to observe such differences in their controls, possibly because of ceiling effects within that study. Therefore, this finding may indicate that when differentiating statements and questions, there might be an advantage in using musical glides relative to speech utterances, since the semantic information plus prosodic cues in speech might hinder participants' perception accuracy.

The consistent findings from Study 1 and Study 3 provide an answer to the question we raised in Study 3 about whether there is an association between mental representation of pitch and subsequent behavioural outcomes of pitch processing (e.g., better or worse pitch identification/discrimination accuracy). The findings support the speculation we made in Study 3 that the fidelity of how well one represents pitch might be associated with one's subsequent behavioural judgment of the pitch. Nevertheless, further studies are required to examine the relationship between pitch representations and behavioural outcomes in the same sample. In short, the findings of intact pitch perception across speech and music domains in individuals with ASD reported by both Studies 1 and 3 lent support for shared mechanisms between speech and music pitch perception as suggested by the domain-general framework.

5.1.2 Pitch production in speech and music

From a production perspective, Study 2 used an imitation task where participants were asked to accurately replicate the pitch and timing patterns of the sentences that were either spoken or sung. The findings indicated that individuals with ASD were impaired in absolute pitch and duration matching but intact in relative pitch and duration matching across both speech and music domains. Study 1 also tested pitch imitation ability in individuals with ASD but only used natural speech alone. The results, however, showed that the ASD group performed comparably to controls, suggesting that individuals with ASD could imitate speech intonation as accurately as controls.

The impaired speech imitation regarding absolute pitch and duration as reported in Study 2 at first blush appears to be inconsistent with the findings from Study 1. However, it is worth noting that different methods were used to measure imitative ability between Study 1 and Study 2. The method used in Study 1 was a relative measure (which is similar to how contour errors were measured in Study 2) rather than an absolute measure. Specifically, the correct imitation in Study 1 was defined by the glide size of the final word without considering the pitch range or variability. If participants' imitated utterances had final falls (negative glide sizes), they would be coded as statements; otherwise, final rises (positive glide sizes) would be coded as questions. By comparing the match between imitated utterances and sounds presented to participants (statements were imitated as statements, and questions were imitated as questions), imitation accuracy was calculated. The imitation accuracy did not differ significantly between the ASD and control groups. However, in Study 2, both absolute and relative measures were used to gauge the imitative ability in comparison to the original model. For example, the absolute pitch deviation was extracted as the absolute median F0 difference between the original model played to participants and the sounds produced by participants on a syllable basis across sentences. Similarly, pitch interval was calculated as the absolute difference in median F0 between two consecutive syllables/notes, and then subtracted from the pitch interval of the model played to participants. Concerning contour error, this was defined as pitch intervals produced by participants that differed from the corresponding pitch intervals of the original model in regard to pitch directions (up, down, or level). For example, if “up” pitch directions were imitated as “up”, imitation of pitch contour was counted as correct, and otherwise as contour error. Thus, the acoustic measures in Study 2 were more detailed and precise than those from Study 1, thus leading to the observed impaired imitation in absolute pitch matching, which was absent in Study 1. Regarding relative pitch imitation, the findings were consistent between Studies 1 and 2, showing that individuals with ASD performed as well as controls. Moreover, the results of Study 2 also suggest that - compared with controls - vocal imitative skills in individuals with ASD, whether impaired (i.e., absolute pitch and duration) or intact (i.e., relative pitch and duration), did not show differences between speech and music. The findings provide evidence for shared mechanisms between speech and song imitation.

5.1.3 The summary of pitch processing in speech and music

The current thesis from both perception and production views tested whether speech and music share the same underlying processing systems. The findings tend to favour the domain-general framework across perception (Studies 1 and 3) and production (Study 2). In addition to the association between speech and music, this thesis also found that there is a processing advantage to musical materials, as opposed to speech intonation, for both individuals with and without ASD. These findings provide a theoretical basis in shared cognitive resources for music therapy aimed at improving language understanding and comprehension in individuals with ASD, especially for those who suffer severe language deficits. A word of caution is required, however. It is known that language profiles vary dramatically among individuals with ASD. Approximately 30% of children with ASD have no or minimal language (e.g., fewer than 30 words) (Tager-Flusberg & Kasari, 2013). While others are verbal and some even have normal language abilities, many of them have notable problems with meaning and comprehension of words and sentences (Eigsti et al., 2011; Wittke et al., 2017). However, the present thesis focused solely on a subgroup in the population with ASD with cognitive abilities comparable to those of typically developing peers, including receptive verbal ability and non-verbal IQ, verbal and non-verbal memory. Further studies testing less linguistically capable individuals would be necessary to confirm the current results.

5.2 The relationship between perception and production in ASD

As previously mentioned, there are three models that predict the relationship between perception and production. Both the perception-based model and the motor model predict that there exists a very close relationship between vocal production and perception. Conversely, the dual-route model claims no correlation between perception and production (Hutchins & Moreno, 2013).

In study 1, participants were asked to imitate the stimuli they heard first, and then identify whether the stimuli being played were questions or statements. The correlation analysis between performance in imitation and identification suggested that identification abilities were positively correlated with imitation abilities in both groups. Thus, the results from Study 1 are consistent with previous studies using subjective ratings (Peppé et al., 2007) and favour the perception-based model and motor model, indicating that the more accurate the perception and understanding, the better the imitation performance, and vice versa.

However, if we view this relationship across these three studies, it is worth noting that Study 1 and Study 3 indicate intact pitch perception in both speech and music domains, whereas Study 2 reveals a partially impaired pitch imitation (intact relative pitch matching but impaired absolute pitch matching) in speech and music. These findings seem to suggest that there may be a dissociation between perception and production. However, the relationship between perception and production within current studies has been complicated by absolute and relative features. While pitch production capabilities between absolute and relative features were distinguished in Study 2, the differences between absolute and relative pitch perception were not tested in Studies 1 and 3. Thus, the current findings are not sufficient to answer the question of whether the relationship between pitch perception and production is modulated by absolute and relative features in individuals with ASD. In study 2, I individually inspected absolute pitch imitation performance for participants who self-reported possessing absolute pitch. Surprisingly, these participants did not perform exceptionally when imitating absolute pitch, as compared to others who reported that they did not possess absolute pitch, which might suggest a dissociation between absolute pitch perception and production in imitation. However, these findings are based on self-reports about whether participants possess absolute pitch, which have not be experimentally confirmed. In addition, a previous case study reported an association between the abilities to produce and perceive absolute pitch in an adolescent with ASD, who showed both absolute pitch identification and production (Mottron et al., 1999). Thus, the relationship between absolute pitch perception and production remains unresolved, and warrants further investigation.

5.3 Low-level versus high-level processing of pitch in ASD

Recall that while WCC and EPF propose that individuals with ASD tend to focus on low-level information, whether their higher-level information processing capacity is intact and the relationship between low-level and high-level processing remain open to debate. In particular, the WCC predicts a dissociation between low-level and high-level information processing in ASD, with enhanced/intact low-level but impaired high-level processing, whereas EPF proposes that enhanced or intact low-level processing does not cause a deficiency in the ability to process high-level information in ASD.

Using low-level pitch direction discrimination thresholds and high-level intonation perception tasks, study 1 showed that both low-level and high-level pitch processing was intact in individuals with ASD across age cohorts (children, adolescents, and adults), except that ASD children exhibited elevated pitch thresholds than their typically developing counterparts. There were also positive relationships between low-level and high-level processing in both groups, with higher sensitivity to pitch correlated with better performance in intonation perception, which likely reflects a bottom-up cascade.

Study 1 observed elevated pitch thresholds in children with ASD compared to their typically developing counterparts. The worse performance in children with ASD may be caused by an attention effect. Pitch threshold tasks generally require consistent attention from participants during tests and inattention leads to poor performance. Children with ASD are generally considered to have poor attention (Morgan et al., 2003; M. J. Murray, 2010), which could have contributed to the present elevated pitch thresholds in children with ASD. Although we have taken measures to sustain participants' attention (e.g., monitoring the whole session by inputting participants' oral responses into the computer) and excluded inattentive performers by inspecting the visual tracks on an individual basis, we do not rule out this possibility. In order to minimize the impact of inattention (e.g., attracting participants' attention consistently throughout the process), child-friendly studies, such as those using cartoons or fun games, are required to investigate the pitch threshold in children with ASD. Indeed, Jones et al. (2009) examined 72 adolescents with ASD and 57 IQ and age matched controls using a cartoon-based pitch threshold task (e.g., dinosaur). The results suggested that the two groups had comparable pitch thresholds. However, to the best of my knowledge, no child-friendly paradigm has been used for young children with ASD, which should be examined in future research.

Based on the intact high-level processing ability in individuals with ASD across all age cohorts, and the intact low-level processing in adolescents and adults with ASD, I argue that the data in Study 1 favour the EPF theory, according to which the ability to process low-level information in individuals with ASD is intact or enhanced in the absence of impairments in high-level information processing. This is consistent with previous studies (Germain et al., 2019; Mottron et al., 2000) and a systematic review (Haesen et al., 2011), which state that the entire empirical findings regarding auditory processing in ASD appear to be more favourable to EPF than the WCC theory. In addition, the positive relationships between low- and high-level pitch processing are also in line with Germain et al. (2019), indicating that the impairment of low-level pitch perception influences performance at subsequent stages, such as melody recognition, language acquisition and communication (Bertone et al., 2010; Stevenson et al., 2014).

5.4 Developmental changes of pitch processing in ASD

This thesis also attempted to identify whether, or to what extent, age influences pitch processing in individuals with ASD, and whether the effects differ from those of controls. The findings from Study 1 revealed that, similar to controls, individuals with ASD showed gains in pitch and pitch-related statement-question intonation processing across developmental stages. It has been found that pitch sensitivity continues to show improvement as age increases in typically developing individuals (Mayer et al., 2016; Stalinski et al., 2008), which is consistent with current results for both groups. In addition, the current findings of similar developmental changes between ASD and control groups in intonation perception and production are also compatible with the results between "language-normal" preadolescents (9-12 years old) and adolescents (13-17 years old) with and without ASD that were reported in Lyons et al. (2014). We further extend the findings from children and adolescents to adults, suggesting that similar age effects were also observed between controls and ASD individuals with normal-language abilities in the adult cohort. Lyons et al. (2014) also observed that, in contrast to “language-normal” preadolescents and adolescents with ASD, those with language impairments exhibited developmental delays in perception and production of intonation. However, the present thesis did not recruit ASD individuals with relatively low language abilities, leaving the question of whether the developmental delays in ASD individuals with language impairments persist into adulthood to future studies.

In contrast, when imitating speech pitch in Study 2, there was a trend for decreasing imitation accuracy in absolute and relative pitch as age increased. The decreased pattern of speech pitch imitation over age was observed in both individuals with ASD and controls. However, the same children and adults produced similar amounts of pitch contour errors and pitch interval errors. It has been suggested that in typically developing children, 5 year olds are able to imitate both the rising contours in question and the falling contours in statement intonation (Loeb & Allen, 1993; Snow, 1998). On the other hand, adults tend to imitate the functional goal (e.g., statements with falling pitch contours vs. questions with rising pitch contours) rather than copying the exact form of the utterances (Liu et al., 2010, 2013). Taken all these findings together and considering the similar amount of pitch contour/interval errors made between the child and adult cohorts, it can be concluded that both children (age range: 7.39-15.75) and adults (age range: 16-56.75) in Study 2 were able to imitate the overall pitch patterns of the speech utterances. Whereas children paid more attention to imitating the exact form, adults paid more attention to the functional goals. That is, the decreased accuracy of speech pitch imitation across development may be due to differences in the imitative attention to form or functional goals between children and adults, rather than to reduced pitch production ability in adults.

In summary, the findings across studies indicated that age did not affect individuals with and without ASD differently when processing pitch information. However, following Mayer et al. (2016), the current age effect was examined using arbitrary age groups, which, while informative, may not be fine-grained enough to detect subtle changes over time across individuals with and without ASD. Further studies using larger sample sizes with different ages across the lifespan might be helpful to map a clearer developmental trajectory in pitch processing in individuals with and without ASD.

5.5 Future directions

While the need for experimental designs to be rigorous was known from the start, some limitations were revealed during the process and could be addressed in future research.

First, while Study 1 attempted to explore how individuals with ASD process low-level and high-level information, particularly using pitch cues, the results are not clear as to whether the impaired low-level pitch processing in children with ASD was due to their pitch processing ability per se or was caused by inattention while doing threshold tasks. Thus, young child-friendly studies, such as those using interfaces with cartoons or fun games, are required to investigate the low-level pitch threshold in children with ASD.

Also, there is a clear follow-up research question that has been raised regarding whether the relationship between perception and production is mediated by absolute vs. relative features. As suggested by the present Study 2, there was a dissociation between the production/imitation of absolute and relative features of sound. However, whether this dissociation influences the relationship between perception and production remains unresolved. Thus, in order to address the question, well-matched studies in terms of absolute vs. relative features and perception vs. production need to be considered. This not only improves our understanding of the relationship between perception and production in individuals with ASD, but also has implications for how absolute vs. relative features contribute to the relationship between perception and production.

In addition, as mentioned, there is limited literature that uses data-driven methods to investigate how individuals with ASD process pitch information across domains and levels, or even other aspects of the auditory system. More data-driven studies are warranted to tease apart the existing mixed results reported in ASD in auditory contexts. Also, while the current findings across Study 1 and Study 3 appear to suggest that representation and behavioural outcomes of pitch are correlated, these findings need to be replicated in studies that used matched stimuli or matched designs between tasks to test representations and to examine behavioural outcomes. By doing so, the findings could provide a more well-founded answer to address the question of whether the relationship between the two is indeed associated. Given that the ability to process pitch improved with age, as suggested by Study 1, it is also crucial to recruit participants across different age cohorts, i.e., children, adolescents, and adults to reveal how age affects mental representations of pitch in individuals with ASD.

Moreover, the current work has focused mainly on English-speakers with and without ASD. It has been suggested that individuals who acquire a tone language, such as Mandarin, tend to have enhanced pitch perception and production than speakers of intonation languages, such as English (Pfordresher & Brown, 2009). Further studies investigating mental representation of pitch across domains in individuals with ASD from other language backgrounds are required to consolidate the current results, which will help us to better understand how tone-language exposure affects pitch representation in individuals with ASD and to comprehensively understand the phenotypes of pitch processing ability in the general ASD population.

5.6 Conclusion

This thesis aims to investigate how individuals with ASD decode and encode pitch information across different auditory domains (speech vs. music), processing levels (low-level vs. high-level), and modalities (perception vs. production) to inform several ongoing theoretical debates. Also, the developmental changes of pitch processing have been mapped in both individuals with and without ASD. Three studies reported in this thesis provide evidence for shared pitch processing mechanisms between speech and music, and associations between low-level and high-level processing, as well as the correlations between perception and production in individuals with ASD. In particular, the three studies all examined the processing of pitch in speech and music, with Study 1 and Study 3 using a traditional discrimination task and a novel reverse-correlation paradigm respectively, suggesting pitch perception shares underlying mechanisms between speech and music in individuals with ASD, and Study 2 using an imitation task revealing that pitch production mechanisms are also shared between speech and music. These findings are consistent with the domain-general framework. Additionally, the correlation between perception and production supports the perception-based model and the motor model. Furthermore, the intact high-level pitch processing ability and the association between low-level and high-level processing are in line with the EPF theory. However, in contrast to increased pitch perception ability over age, the pitch imitation ability tends to decrease with age in both individuals with and without ASD. Taken together, the overall findings have tested several theoretical debates and have provided evidence for the domain-general framework between speech and music, the EPF theory of intact high-level information processing ability, and the closely linked perception and production that has been proposed by the perception-based model and the motor model. These findings not only improve our understanding of how individuals with ASD decode and encode pitch information, but also have theoretical implications for the use of music to improve language comprehension, and the practice of imitation to improve language perception, as well as the use of low-level information to improve higher-level cognitive understanding in individuals with ASD who suffer language problems.

References

Acton, G. S., & Schroeder, D. H. (2001). Sensory discrimination as related to general intelligence. *Intelligence, 29*(3), 263–271. https://doi.org/10.1016/S0160-2896(01)00066-6

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150367. https://doi.org/10.1098/rstb.2015.0367

Ahmad, O. B., Boschi-pinto, C., Lopez, A. D., Lozano, R., & Inoue, M. (2009). Age standardization of rates: A new WHO standard. *Geneva: World Health Organization, 9(10).*

Ahumada, A., & Lovell, J. (1971). Stimulus Features in Signal Detection. *The Journal of the Acoustical Society of America*, *49*(6B), 1751–1756. https://doi.org/10.1121/1.1912577

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*(6481), 1043–1047. https://doi.org/10.1126/science.aaz3468

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

Altgassen, M., Kliegel, M., & Williams, T. I. (2005). Pitch perception in children with autistic spectrum disorders. *British Journal of Developmental Psychology*, *23*(4), 543–558. https://doi.org/10.1348/026151005X26840

American Psychiatric Association. (1994). *American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)*. American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association. http://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America*, *113*(2), 1102–1108. https://doi.org/10.1121/1.1536632

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

Arciuli, J., & Ballard, K. J. (2017). Still not adult-like: Lexical stress contrastivity in word productions of eight- to eleven-year-olds*. *Journal of Child Language*, *44*(5), 1274–1288. https://doi.org/10.1017/S0305000916000489

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00321

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of Autism and Developmental Disorders*, *38*(7), 1230–1240. https://doi.org/10.1007/s10803-007-0504-z

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251. https://doi.org/10.1093/brain/awf028

Ballard, K. J., Djaja, D., Arciuli, J., James, D. G. H., & Van. (2012). Developmental Trajectory for Production of Prosody: Lexical Stress Contrastivity in Children Ages 3 to 7 Years and in Adults. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1822–1835. https://doi.org/10.1044/1092-4388(2012/11-0257)

Banuazizi, A., & Creswell, C. (1999). Is that a real question? Final rises, final falls, and discourse function in yes-no question intonation. *CLS*, *35*, 1–14.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The Autism-Spectrum Quotient (AQ)—Adolescent Version. *Journal of Autism and Developmental Disorders*, *36*(3), 343–350. https://doi.org/10.1007/s10803-006-0073-6

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0.

Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circulation Research*, *116*(1), 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69–83. https://doi.org/10.2478/v10053-008-0068-2

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The Sung Performance Battery (SPB). *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00714

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain and Cognition*, *64*(3), 228–237. https://doi.org/10.1016/j.bandc.2007.03.004

Bertone, A., Hanck, J., Kogan, C., Chaudhuri, A., & Cornish, K. (2010). Using Perceptual Signatures to Define and Dissociate Condition-Specific Neural Etiology: Autism and Fragile X Syndrome as Model Conditions. *Journal of Autism and Developmental Disorders*, *40*(12), 1531–1540. https://doi.org/10.1007/s10803-010-1109-5

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–434. https://doi.org/10.1162/jocn.2009.21362

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and

Music. *PLoS ONE*, *8*(4), e60676. https://doi.org/10.1371/journal.pone.0060676

Bijou, S. W., & Ghezzi, P. M. (1999). The behavior interference theory of autistic behavior in young children. *Autism: Behavior Analytic Perspectives*, 33–43.

Biscaldi, M., Rauh, R., Irion, L., Jung, N. H., Mall, V., Fleischhaker, C., & Klein, C. (2014). Deficits in motor abilities and developmental fractionation of imitation performance in high-functioning autism spectrum disorders. *European Child & Adolescent Psychiatry*, *23*(7), 599–610. https://doi.org/10.1007/s00787-013-0475-x

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

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal Speech Spectrum and Increased Pitch Variability in Young Autistic Children. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00237

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

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced Pitch Sensitivity in Individuals with Autism: A Signal Detection Analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Boster, J. B., Spitzley, A. M., Castle, T. W., Jewell, A. R., Corso, C. L., & McCarthy, J. W. (2020). Music Improves Social and Participation Outcomes for

Individuals With Communication Disorders: A Systematic Review. *Journal of Music Therapy, thaa015*. https://doi.org/10.1093/jmt/thaa015

Boutsen, F. (2003). Prosody: The music of language and speech. *The ASHA Leader*, *8*(4), 6–8.

Bouvet, L., Simard-Meilleur, A.-A., Paignon, A., Mottron, L., & Donnadieu, S. (2014). Auditory local bias and reduced global interference in autism. *Cognition*, *131*(3), 367–372. https://doi.org/10.1016/j.cognition.2014.02.006

Bradley, E. D. (2012). Tone language experience enhances sensitivity to melodic contour. *LSA Annual Meeting Extended Abstracts*, *3*(0), 40-1–5. https://doi.org/10.3765/exabs.v0i0.612

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Brown, W. A., Cammuso, K., Sachs, H., Winklosky, B., Mullane, J., Bernier, R., Svenson, S., Arin, D., Rosen-Sheidley, B., & Folstein, S. E. (2003). Autism-Related Language, Personality, and Cognition in People with Absolute Pitch: Results of a Preliminary Study. *Journal of Autism and Developmental Disorders*, *33*(2), 163–167. https://doi.org/10.1023/A:1022987309913

Burred, J. J., Ponsot, E., Goupil, L., Liuni, M., & Aucouturier, J.-J. (2019). CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. *PLoS ONE*, *14*(4). https://doi.org/10.1371/journal.pone.0205943

Busby, P. A., Roberts, S. A., Tong, Y. C., & Clark, G. M. (1991). Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant. *British Journal of Audiology*, *25*(5), 291–302. https://doi.org/10.3109/03005369109076601

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

Chan, K. K. L., & To, C. K. S. (2016). Do Individuals with High-Functioning Autism Who Speak a Tone Language Show Intonation Deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792. https://doi.org/10.1007/s10803-016-2709-5

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481–498. https://doi.org/10.1016/S0885-2014(01)00037-5

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch Perception in Tone Language-Speaking Adults With and Without Autism Spectrum Disorders. *I-Perception*, *8*(3), 2041669517711200. https://doi.org/10.1177/2041669517711200

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

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*, 0301006617718715. https://doi.org/10.1177/0301006617718715

Clopper, C. G., Rohrbeck, K. L., & Wagner, L. (2012). Perception of Dialect Variation by Young Adults with High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *42*(5), 740–754. https://doi.org/10.1007/s10803-011-1305-y

Cohrdes, C., Grolig, L., & Schroeder, S. (2016). Relating Language and Music Skills in Young Children: A First Approach to Systemize and Compare Distinct Competencies on Different Levels. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01616

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? *Proceedings of Interspeech 2011*, 969–972. https://experts.illinois.edu/en/publications/the-phonology-and-phonetics-of-perceived-prosody-what-do-listener

Cooper, N. A. (1995). Children’s Singing Accuracy as a Function of Grade Level, Gender, and Individual versus Unison Singing. *Journal of Research in Music Education*, *43*(3), 222–231. https://doi.org/10.2307/3345637

Cossu, G., Boria, S., Copioli, C., Bracceschi, R., Giuberti, V., Santelli, E., & Gallese, V. (2012). Motor Representation of Actions in Children with Autism. *PLOS ONE*, *7*(9), e44779. https://doi.org/10.1371/journal.pone.0044779

Crozier, J. B. (1997). Absolute Pitch: Practice Makes Perfect, the Earlier the Better. *Psychology of Music*, *25*(2), 110–119.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds*. *Journal of Child Language*, *12*(3), 643–661. https://doi.org/10.1017/S030500090000670X

Cruttenden, A. (1997). *Intonation* (Second Edition). Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. CUP Archive.

Dahan, D. (2015). Prosody and language comprehension. *WIREs Cognitive Science*, *6*(5), 441–452. https://doi.org/10.1002/wcs.1355

Dalla Bella, S. (2015). Defining poor-pitch singing: A problem of measurement and sensitivity. *Music Perception*, *32*(3), 272–282. https://doi.org/10.1525/mp.2015.32.3.272

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

Dalla Bella, S., Deutsch, D., Giguère, J.-F., Peretz, I., & Deutsch, D. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America*, *121*(2), 1182–1189. https://doi.org/10.1121/1.2427111

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America*, *121*(2), 1182–1189.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America*, *126*(1), 414–424. https://doi.org/10.1121/1.3132504

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

DePape, A.-M. R., Chen, A., Hall, G. B. C., & Trainor, L. J. (2012). Use of Prosody and Information Structure in High Functioning Adults with Autism in Relation to Language Ability. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00072

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

Deutsch, D. (2002). The Puzzle of Absolute Pitch. *Current Directions in Psychological Science*, *11*(6), 200–204. https://doi.org/10.1111/1467-8721.00200

Deutsch, D. (2013). Absolute Pitch. In *The Psychology of Music* (pp. 141–182). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00005-5

Diehl, J. J., Friedberg, C., Paul, R., & Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*, *27*(3), 867–884. https://doi.org/10.1017/S0954579414000741

Diehl, J. J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(1), 123–134. https://doi.org/10.1016/j.rasd.2011.03.012

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

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

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, *55*(1), 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00781

D’Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01226

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic Out-Group Faces Are Biased in the Prejudiced Mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Dowling, W. J., & Fujitani, D. S. (1971). Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America*, *49*(2B), 524–531. https://doi.org/10.1121/1.1912382

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, *74*(1), 1–32. https://doi.org/10.1016/S0010-0277(99)00061-X

Edwards, L. A. (2014). A Meta-Analysis of Imitation Abilities in Individuals With Autism Spectrum Disorders. *Autism Research*, *7*(3), 363–380. https://doi.org/10.1002/aur.1379

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

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cerebral Cortex*, *22*(3), 650–658. https://doi.org/10.1093/cercor/bhr142

Fancourt, A., Dick, F., & Stewart, L. (2013). Pitch-change detection and pitch-direction discrimination in children. *Psychomusicology: Music, Mind, and Brain*, *23*(2), 73.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

Fodor, J. A. (2001). *The Mind Doesn’t Work that Way: The Scope and Limits of Computational Psychology*. MIT Press.

Fosnot, S. M., & Jun, S.-A. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. *Proceedings of the 14th International Congress of Phonetic Sciences*, 1925–1928.

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *The Journal of the Acoustical Society of America, 114*(3), 1611–1621.

Fridland, E., & Moore, R. (2015). Imitation reconsidered. *Philosophical Psychology*, *28*(6), 856–880. https://doi.org/10.1080/09515089.2014.942896

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

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. https://doi.org/10.3758/BF03193857

Gérard, C., & Auxiette, C. (1992). The Processing of Musical Prosody by Musical and Nonmusical Children. *Music Perception, 10*(1), 93–125. https://doi.org/10.2307/40285541

Geringer, J. M. (1983). The Relationship of Pitch-Matching and Pitch-Discrimination Abilities of Preschool and Fourth-Grade Students. *Journal of Research in Music Education*, *31*(2), 93–99. https://doi.org/10.2307/3345213

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

Gold, B., Morgan, N., & Ellis, D. (2011). *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons.

Grabe, E. (2002). Variation Adds to Prosodic Typology. *Proceedings of the Speech Prosody 2002 Conference*, 127–132.

Griffiths, T. D. (2008). Sensory systems: Auditory action streams? *Current Biology: CB*, *18*(9), R387-388. https://doi.org/10.1016/j.cub.2008.03.007

Grossman, R. B., Bemis, R. H., Skwerer, D. P., & Tager-Flusberg, H. (2010). Lexical and Affective Prosody in Children With High-Functioning Autism. *Journal of Speech, Language, and Hearing Research*, *53*(3), 778–793. https://doi.org/10.1044/1092-4388(2009/08-0127)

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. *6th International Conference on Spoken Language Processing (ICSLP 2000)*, 91–94.

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

Hall, D., Huerta, M. F., McAuliffe, M. J., & Farber, G. K. (2012). Sharing Heterogeneous Data: The National Database for Autism Research. *Neuroinformatics*, *10*(4), 331–339. https://doi.org/10.1007/s12021-012-9151-4

Ham, H. S., Bartolo, A., Corley, M., Rajendran, G., Szabo, A., & Swanson, S. (2011). Exploring the Relationship Between Gestural Recognition and Imitation: Evidence of Dyspraxia in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *41*(1), 1–12. https://doi.org/10.1007/s10803-010-1011-1

Hamilton, A. F. de C. (2008). Emulation and Mimicry for Social Interaction: A Theoretical Approach to Imitation in Autism. *Quarterly Journal of Experimental Psychology*, *61*(1), 101–115. https://doi.org/10.1080/17470210701508798

Hampton, L. H., & Kaiser, A. P. (2016). Intervention effects on spoken-language outcomes for children with autism: A systematic review and meta-analysis. *Journal of Intellectual Disability Research*, *60*(5), 444–463. https://doi.org/10.1111/jir.12283

Happé, F., & Booth, R. D. L. (2008). The Power of the Positive: Revisiting Weak Coherence in Autism Spectrum Disorders. *Quarterly Journal of Experimental Psychology*, *61*(1), 50–63. https://doi.org/10.1080/17470210701508731

Happé, F., & Frith, U. (2006). The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

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. (2009). Assessing musical skills in autistic children who are not savants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1443–1447. https://doi.org/10.1098/rstb.2008.0327

Heaton, P., Allen, R., Williams, K., Cummins, O., & Happé, F. (2008). Do social and cognitive deficits curtail musical understanding? Evidence from autism and Down syndrome. *British Journal of Developmental Psychology*, *26*(2), 171–182. https://doi.org/10.1348/026151007X206776

Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and Pitch Processing: A Precursor for Savant Musical Ability? *Music Perception: An Interdisciplinary Journal*, *15*(3), 291–305. https://doi.org/10.2307/40285769

Heaton, P., Hermelin, B., & Pring, L. (1999). Can children with autistic spectrum disorders perceive affect in music? An experimental investigation. *Psychological Medicine*, *29*(06), 1405–1410. https://doi.org/null

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

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/10.1080/13554799908402745

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

Heaton, P., Williams, K., Cummins, O., & Happé, F. G. E. (2007). Beyond Perception: Musical Representation and On-line Processing in Autism.

Journal of Autism and Developmental Disorders, *37*(7), 1355–1360. https://doi.org/10.1007/s10803-006-0283-y

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

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261. https://doi.org/10.1016/S1364-6613(00)01661-2

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

Horváth, J., Czigler, I., Birkás, E., Winkler, I., & Gervai, J. (2009). Age-related differences in distraction and reorientation in an auditory task. *Neurobiology of Aging*, *30*(7), 1157–1172. https://doi.org/10.1016/j.neurobiolaging.2007.10.003

Howell, D. C. (2009). *Statistical Methods for Psychology* (7 edition). Cengage Learning.

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

Hübscher, I., & Prieto, P. (2019). Gestural and Prosodic Development Act as Sister Systems and Jointly Pave the Way for Children’s Sociopragmatic Development. *Frontiers in Psychology*, *10*, 1259. https://doi.org/10.3389/fpsyg.2019.01259

Hutchins, S., & Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00825

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

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00386

Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, *68*(1), 269–297. https://doi.org/10.1146/annurev-psych-010416-044242

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

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

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

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

Järvinen-Pasley, A., Wallace, G. L., Ramus, F., Happé, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Developmental Science*, *11*(1), 109–121. https://doi.org/10.1111/j.1467-7687.2007.00644.x

JASP Team. (2020). *JASP (Version 0.13.1) [Computer software]*.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*(9), 2630–2639. https://doi.org/10.1016/j.neuropsychologia.2010.05.009

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

Johnson, K. (2011). *Acoustic and Auditory Phonetics*. John Wiley & Sons.

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

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., Sauter, D., Tregay, J., Phillips, R. J., Baird, G., Simonoff, E., & Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(3), 275–285. https://doi.org/10.1111/j.1469-7610.2010.02328.x

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kanner, L. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia*, *1*(2), 119–145.

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

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, *8*(2), 157–168. https://doi.org/10.1016/S0095-4470(19)31460-3

Kessels, R. P. C., Zandvoort, M. J. E. van, Postma, A., Kappelle, L. J., & Haan, E. H. F. de. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.-L., Barthélémy, C., & Collet, L. (2004). Increased perception of loudness in autism. *Hearing Research*, *198*(1), 87–92. https://doi.org/10.1016/j.heares.2004.07.006

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An Investigation of Language Impairment in Autism: Implications for Genetic Subgroups. *Language and*

Cognitive Processes, *16*(2–3), 287–308. https://doi.org/10.1080/01690960042000058

Koelsch, S. (2011). Toward a Neural Basis of Music Perception – A Review and Updated Model. *Frontier in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00110

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584. https://doi.org/10.1016/j.tics.2005.10.001

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, *124*(5), 3158–3170. https://doi.org/10.1121/1.2981639

Krumhansl, C. L. (2004). The Cognition of Tonality – as We Know it Today. *Journal of New Music Research*, *33*(3), 253–268. https://doi.org/10.1080/0929821042000317831

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.

Kügler, F. (2003). Do we know the answer? Variation in yes-no-question intonation. *Linguistics in Potsdam*, *21*, 9–29.

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

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*(1), 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

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

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425–2438. https://doi.org/10.1121/1.417951

Kujala, T., Lepistö, T., Nieminen-von Wendt, T., Näätänen, P., & Näätänen, R. (2005). Neurophysiological evidence for cortical discrimination impairment of prosody in Asperger syndrome. *Neuroscience Letters*, *383*(3), 260–265. https://doi.org/10.1016/j.neulet.2005.04.048

Kunert, R., & Slevc, L. R. (2015). A Commentary on: “Neural overlap in processing music and speech.” *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00330

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Kwok, E. Y. L., Brown, H. M., Smyth, R. E., & Oram Cardy, J. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *9*, 202–222. https://doi.org/10.1016/j.rasd.2014.10.008

Lai, G., Pantazatos, S. P., Schneider, H., & Hirsch, J. (2012). Neural systems for speech and song in autism. *Brain*, *135*(3), 961–975. https://doi.org/10.1093/brain/awr335

Lau, J. C. Y., To, C. K. S., Kwan, J. S. K., Kang, X., Losh, M., & Wong, P. C. M. (2020). Lifelong Tone Language Experience does not Eliminate Deficits in Neural Encoding of Pitch in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04796-7

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means* (1(2)) [Computer software]. R package version. https://github.com/rvlenth/emmeans

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

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

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/BF03206733

Levitt, A. G. (1993). The Acquisition of Prosody: Evidence from French- and English-Learning Infants. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 385–398). Springer Netherlands. https://doi.org/10.1007/978-94-015-8234-6_31

Liberman, A. M. (1996). *Speech: A Special Code*. MIT Press.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Lieberman, P. (1960). Some Acoustic Correlates of Word Stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454. https://doi.org/10.1121/1.1908095

Liu, F., Chan, A. H. D., Ciocca, V., Roquet, C., Peretz, I., & Wong, P. C. M. (2016). Pitch perception and production in congenital amusia: Evidence from Cantonese speakers. *The Journal of the Acoustical Society of America*, *140*(1), 563–575. https://doi.org/10.1121/1.4955182

Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., & Stewart, L. (2013). Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing. *Attention, Perception, & Psychophysics*, *75*(8), 1783–1798. https://doi.org/10.3758/s13414-013-0506-1

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PloS One*, *7*(2), e30374. https://doi.org/10.1371/journal.pone.0030374

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia*, *66*, 111–118. https://doi.org/10.1016/j.neuropsychologia.2014.11.001

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*(6), 1682–1693. https://doi.org/10.1093/brain/awq089

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

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

Loeb, D. F., & Allen, G. D. (1993). Preschoolers' Imitation of Intonation Contours. *Journal of Speech, Language, and Hearing Research*, *36*(1), 4–13. https://doi.org/10.1044/jshr.3601.04

Loui, P. (2015). A Dual-Stream Neuroanatomy of Singing. *Music Perception*, *32*(3), 232–241. https://doi.org/10.1525/mp.2015.32.3.232

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology : CB*, *18*(8), R331–R332. https://doi.org/10.1016/j.cub.2008.02.045

Lyons, M., Simmons, E. S., & Paul, R. (2014). Prosodic Development in Middle Childhood and Adolescence in High-Functioning Autism. *Autism Research*, *7*(2), 181–196. https://doi.org/10.1002/aur.1355

Magne, C., Schon, D., & Besson, M. (2006). *Musician Children Detect Pitch Violations in Both Music and Language Better than Nonmusician Children: Behavioral and Electrophysiological Approaches*. *18*(2), 21.

Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, *127*(2), 177–202. https://doi.org/10.1016/j.cognition.2012.12.008

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715. https://doi.org/10.1162/jocn.2010.21585

Marin, O. S., & Perry, D. W. (1999). Neurological aspects of music perception and performance. In *The psychology of music* (pp. 653–724). Elsevier.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463. https://doi.org/10.1162/jocn.2007.19.9.1453

Martin, N. A., & Brownell, R. (2011). *ROWPVT-4: Receptive One-Word Picture Vocabulary Test*.

Masur, E. F. (2006). Vocal and Action Imitation by Infants and Toddlers during Dyadic Interactions: Development, Causes, and Consequences. In *Imitation and the social mind: Autism and typical development* (pp. 27–47). The Guilford Press.

MATLAB. (2010). *Version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500.

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

McAlpine, A., Plexico, L. W., Plumb, A. M., & Cleary, J. (2014). Prosody in Young Verbal Children With Autism Spectrum Disorder. *Contemporary Issues in Communication Science and Disorders, 41*(Spring), 120–132. https://doi.org/10.1044/cicsd_41_S_120

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

McCann, J., Peppé, S., Gibbon, F. E., O’Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, *42*(6), 682–702. https://doi.org/10.1080/13682820601170102

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America*, *130*(6), 4070–4081. https://doi.org/10.1121/1.3658476

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. https://doi.org/10.1016/j.conb.2008.09.005

Mcdonald, J. H. (2014). *Handbook of Biological Statistics* (3rd ed.). Sparky House Publishing.

McEwen, F., Happé, F., Bolton, P., Rijsdijk, F., Ronald, A., Dworzynski, K., & Plomin, R. (2007). Origins of Individual Differences in Imitation: Links With Language, Pretend Play, and Socially Insightful Behavior in Two-Year-Old Twins. *Child Development*, *78*(2), 474–492. https://doi.org/10.1111/j.1467-8624.2007.01010.x

Meltzoff, A. N. (2017). Elements of a comprehensive theory of infant imitation. *Behavioral and Brain Sciences, 40*. https://doi.org/10.1017/S0140525X1600193X

Meltzoff, A. N., & Keith Moore, M. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, *17*(1), 83–99. https://doi.org/10.1016/0163-6383(94)90024-8

Meltzoff, A. N., & Williamson, R. A. (2013). Imitation: Social, cognitive, and theoretical perspectives. In *The Oxford handbook of developmental psychology (Vol 1): Body and mind* (pp. 651–682). Oxford University Press.

Mendez, M. F. (2001). Generalized Auditory Agnosia with Spared Music Recognition in a Left-Hander. Analysis of a Case with a Right Temporal Stroke. *Cortex*, *37*(1), 139–150. https://doi.org/10.1016/S0010-9452(08)70563-X

Mertens, P. (2004). The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. *Proceedings of Speech Prosody 2004*, 23–26. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.6769&rep=rep1&type=pdf

Miller, L. K. (1999). The Savant Syndrome: Intellectual impairment and exceptional skill. *Psychological Bulletin*, *125*(1), 31–46. https://doi.org/10.1037/0033-2909.125.1.31

Milne, E. (2011). Increased Intra-Participant Variability in Children with Autistic Spectrum Disorders: Evidence from Single-Trial Analysis of Evoked EEG. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00051

Molnar-Szakacs, I., & Heaton, P. (2012). Music: A unique window into the world of autism. *Annals of the New York Academy of Sciences*, *1252*, 318–324. https://doi.org/10.1111/j.1749-6632.2012.06465.x

Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research*, *238*(1), 147–154. https://doi.org/10.1016/j.heares.2007.11.013

Morgan, B., Maybery, M., & Durkin, K. (2003). Weak central coherence, poor joint attention, and low verbal ability: Independent deficits in early autism. *Developmental Psychology*, *39*(4), 646–656.

Möttönen, R., Järveläinen, J., Sams, M., & Hari, R. (2005). Viewing speech modulates activity in the left SI mouth cortex. *Neuroimage*, *24*(3), 731–737.

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In *The development of autism: Perspectives from theory and research* (pp. 131–148). Lawrence Erlbaum Associates Publishers.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Mottron, L., Peretz, I., Belleville, S., & Rouleau, N. (1999). Absolute pitch in autism: A case study. *Neurocase*, *5*(6), 485–501. https://doi.org/10.1080/13554799908402744

Mottron, L., Peretz, I., & Ménard, E. (2000). Local and Global Processing of Music in High-functioning Persons with Autism: Beyond Central Coherence? *Journal of Child Psychology and Psychiatry*, *41*(8), 1057–1065. https://doi.org/10.1111/1469-7610.00693

Munson, B. (2004). Variability in /s/ Production in Children and Adults. *Journal of Speech, Language, and Hearing Research*, *47*(1), 58–69. https://doi.org/10.1044/1092-4388(2004/006)

Murray, M. J. (2010). Attention-deficit/Hyperactivity Disorder in the Context of Autism Spectrum Disorders. *Current Psychiatry Reports*, *12*(5), 382–388. https://doi.org/10.1007/s11920-010-0145-3

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5), 2–2. https://doi.org/10.1167/11.5.2

Nadig, A., & Shaw, H. (2012). Acoustic and Perceptual Measurement of Expressive Prosody in High-Functioning Autism: Increased Pitch Range and What it Means to Listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-Level Information and High-Level Perception: The Case of Speech in Noise. *PLOS Biology*, *6*(5), e126. https://doi.org/10.1371/journal.pbio.0060126

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

Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, *53*, 46–54. https://doi.org/10.1016/j.wocn.2015.08.004

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*(1), 490–493.

Nithyashri, J., & Kulanthaivel, G. (2012). Classification of human age based on Neural Network using FG-NET Aging database and Wavelets. *2012 Fourth International Conference on Advanced Computing (ICoAC)*, 1–5. https://doi.org/10.1109/ICoAC.2012.6416855

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, *88*(6), 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035

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

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute Pitch—Functional Evidence of Speech-Relevant Auditory Acuity. *Cerebral Cortex*, *20*(2), 447–455. https://doi.org/10.1093/cercor/bhp113

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

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

Over, H., & Carpenter, M. (2013). The Social Side of Imitation. *Child Development Perspectives*, *7*(1), 6–11. https://doi.org/10.1111/cdep.12006

Over, H., & Gattis, M. (2010). Verbal imitation is based on intention understanding. *Cognitive Development*, *25*(1), 46–55. https://doi.org/10.1016/j.cogdev.2009.06.004

Pagliarini, S., Leblois, A., & Hinaut, X. (2020). Vocal Imitation in Sensorimotor Learning Models: A Comparative Review. *IEEE Transactions on Cognitive* *and Developmental Systems*, 1–1. https://doi.org/10.1109/TCDS.2020.3041179

Parish-Morris, J., Liberman, M., Ryant, N., Cieri, C., Bateman, L., Ferguson, E., & Schultz, R. T. (2016). Exploring autism spectrum disorders using HLT. *Proceedings of the Conference. Association for Computational Linguistics. Meeting*, *2016*, 74.

Park, W. J., Schauder, K. B., Zhang, R., Bennetto, L., & Tadin, D. (2017). High internal noise and poor external noise filtering characterize perception in autism spectrum disorder. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-17676-5

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press, USA.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. *Annals of the New York Academy of Sciences*, *1252*, 124–128. https://doi.org/10.1111/j.1749-6632.2011.06426.x

Patel, A. D. (2013). Sharing and nonsharing of brain resources for language and music. In *Language, Music, and the Brain: A Mysterious Relationship (M. Arbib, Ed.)* (pp. 329–355). MIT Press. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.716.9857

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35–B45. https://doi.org/10.1016/S0010-0277(02)00187-7

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*(3), 310–313. https://doi.org/10.1016/j.bandc.2004.10.003

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, *119*(5), 3034–3047. https://doi.org/10.1121/1.2179657

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing Prosodic and Musical Patterns: A Neuropsychological Investigation. *Brain and Language*, *61*(1), 123–144. https://doi.org/10.1006/brln.1997.1862

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). SPEECH INTONATION PERCEPTION DEFICITS IN MUSICAL TONE DEAFNESS (CONGENITAL AMUSIA). *Music Perception: An Interdisciplinary Journal*, *25*(4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357

Patel, A. D., Xu, Y., & Wang, B. (2010). The role of F0 variation in the intelligibility of Mandarin sentences. *Speech Prosody*, 4.

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4, 7 and 11 year-old children. *Speech Communication*, *48*(10), 1308–1318. https://doi.org/10.1016/j.specom.2006.06.007

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. (2008). Production of Syllable Stress in Speakers with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124. https://doi.org/10.1016/j.rasd.2007.04.001

Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jääskeläinen, I. P., Kujala, T., & Sams, M. (2006). Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: An fMRI study at 3 T. *Neuroimage*, *29*(3), 797–807.

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

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, *17*(4–5), 345–354. https://doi.org/10.1080/0269920031000079994

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and Expressive Prosodic Ability in Children With High-Functioning Autism. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Peretz, I. (2009). Music, Language and Modularity Framed in Action. *Psychologica Belgica*, *49*(2–3), 157. https://doi.org/10.5334/pb-49-2-3-157

Peretz, I., Champod, A. S., & Hyde, K. L. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. https://doi.org/10.1038/nn1083

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00030

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

Peretz, I., & Zatorre, R. (2005). Brain Organization for Music Processing. *Annual Review of Psychology*, *56*, 89–114.

Pfordresher, P. Q., & Brown, S. (2007). Poor-Pitch Singing in the Absence of "Tone Deafness." *Music Perception*, *25*(2), 95–115. https://doi.org/10.1525/mp.2007.25.2.95

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

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

Plack, C. J., Barker, D., & Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research*, *307*, 53–64. https://doi.org/10.1016/j.heares.2013.07.020

Plack, C. J., & Oxenham, A. J. (2005). Overview: The present and future of pitch. *Pitch*, 1–6.

Plack, C. J., Oxenham, A. J., & Fay, R. R. (2006). *Pitch: Neural Coding and Perception*. Springer Science & Business Media.

Ponsot, E., Arias, P., & Aucouturier, J.-J. (2018). Uncovering mental representations of smiled speech using reverse correlation. *The Journal of the Acoustical Society of America*, *143*(1), EL19–EL24. https://doi.org/10.1121/1.5020989

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J.-J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences*, *115*(15), 3972–3977. https://doi.org/10.1073/pnas.1716090115

Ponsot, E., Varnet, L., Wallaert, N., Daoud, E., Shamma, S. A., Lorenzi, C., & Neri, P. (2021). Mechanisms of Spectrotemporal Modulation Detection for Normal- and Hearing-Impaired Listeners. *Trends in Hearing*, *25*, 2331216520978029. https://doi.org/10.1177/2331216520978029

Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, *141*(2), 184–198. https://doi.org/10.1016/j.actpsy.2012.07.013

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111–163. JSTOR. https://doi.org/10.2307/271063

Ratner, K. G., Dotsch, R., Wigboldus, D. H., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*(6), 897.

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

Reifinger, J. L. (2006). Skill Development in Rhythm Perception and Performance: A Review of Literature. *Update: Applications of Research in Music Education*, *25*(1), 15–27. https://doi.org/10.1177/87551233060250010103

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.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A Cognitive Approach to the Development of Early Language. *Child Development*, *80*(1), 134–150. https://doi.org/10.1111/j.1467-8624.2008.01250.x

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Sadakata, M., Weidema, J. L., & Honing, H. (2020). Parallel pitch processing in speech and melody: A study of the interference of musical melody on lexical pitch perception in speakers of Mandarin. *PLOS ONE*, *15*(3), e0229109. https://doi.org/10.1371/journal.pone.0229109

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74.

Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *Neuroreport*, *17*(18), 1893–1897.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., Grigutsch, M., Huppertz, H.-J., Knösche, T. R., Wellmer, J., Widman, G., &

Schulze-Bonhage, A. (2009). Overlap of Musical and Linguistic Syntax Processing: Intracranial ERP Evidence. *Annals of the New York Academy of Sciences*, *1169*(1), 494–498. https://doi.org/10.1111/j.1749-6632.2009.04792.x

Schelinski, S., Roswandowitz, C., & Kriegstein, K. von. (2017). Voice identity processing in autism spectrum disorder. *Autism Research*, *10*(1), 155–168. https://doi.org/10.1002/aur.1639

Schelinski, S., & von Kriegstein, K. (2020). Brief Report: Speech-in-Noise Recognition and the Relation to Vocal Pitch Perception in Adults with Autism Spectrum Disorder and Typical Development. *Journal of Autism and Developmental Disorders*, *50*(1), 356–363. https://doi.org/10.1007/s10803-019-04244-1

Schellenberg, E. G., & Trehub, S. E. (2003). Good Pitch Memory Is Widespread. *Psychological Science*, *14*(3), 262–266. https://doi.org/10.1111/1467-9280.03432

Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Perception*, *25*(3), 241–252. https://doi.org/10.1525/mp.2008.25.3.241

Schleuter, S. L., & Schleuter, L. J. (1985). The Relationship of Grade Level and Sex Differences to Certain Rhythmic Responses of Primary Grade Children. *Journal of Research in Music Education*, *33*(1), 23–29. https://doi.org/10.2307/3344755

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schreibman, L., Kohlenberg, B. S., & Britten, K. R. (1986). Differential responding to content and intonation components of a complex auditory stimulus by nonverbal and echolalic autistic children. *Analysis and Intervention in Developmental Disabilities*, *6*(1–2), 109–125. https://doi.org/10.1016/0270-4684(86)90009-1

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354. https://doi.org/10.1016/j.jneuroling.2009.12.004

Sergeant, D., & Roche, S. (1973). Perceptual Shifts in the Auditory Information Processing of Young Children. *Psychology of Music*, *1*(2), 39–48.

Sharda, M., Midha, R., Malik, S., Mukerji, S., & Singh, N. C. (2015). Fronto-Temporal Connectivity is Preserved During Sung but Not Spoken Word Listening, Across the Autism Spectrum. *Autism Research*, *8*(2), 174–186. https://doi.org/10.1002/aur.1437

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

Shriberg, L. D., Paul, R., Black, L. M., & van Santen, J. P. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(4), 405–426. https://doi.org/10.1007/s10803-010-1117-5

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and Prosody Characteristics of Adolescents and Adults With

High-Functioning Autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Simmons, J. Q., & Baltaxe, C. (1975). Language patterns of adolescent autistics. *Journal of Autism and Childhood Schizophrenia*, *5*(4), 333–351.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, *16*(2), 374–381. https://doi.org/10.3758/16.2.374

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, *6*(1), 37–67. https://doi.org/10.1016/S0095-4470(19)31084-8

Smith, B. L., Kenney, M. K., & Hussain, S. (1996). A longitudinal investigation of duration and temporal variability in children’s speech production. *The Journal of the Acoustical Society of America*, *99*(4), 2344–2349. https://doi.org/10.1121/1.415421

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research*, *18*(4), 795–811.

Snow, D. (1998). Children’s imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language, and Hearing Research*, *41*(3), 576–587.

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

Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: Distinguishing up from down. *The Journal of the Acoustical Society of America*, *124*(3), 1759–1763. https://doi.org/10.1121/1.2956470

Stanutz, S., Wapnick, J., & Burack, J. A. (2014). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *18*(2), 137–147. https://doi.org/10.1177/1362361312462905

Steffens, T., Steffens, L. M., & Marcrum, S. C. (2020). Chance-level hit rates in closed-set, forced-choice audiometry and a novel utility for the significance test-based detection of malingering. *PLOS ONE*, *15*(4), e0231715. https://doi.org/10.1371/journal.pone.0231715

Steinke, W. R., Cuddy, L. L., & Jakobson, L. S. (2001). Dissociations among functional subsystems governing melody recognition after right-hemisphere damage. *Cognitive Neuropsychology*, *18*(5), 411–437. https://doi.org/10.1080/02643290125702

Stelmachowicz, P. G., Hoover Brenda M., Lewis Dawna E., Kortekaas Reinier W. L., & Pittman Andrea L. (2000). The Relation Between Stimulus Context, Speech Audibility, and Perception for Normal-Hearing and Hearing-Impaired Children. *Journal of Speech, Language, and Hearing Research*, *43*(4), 902–914. https://doi.org/10.1044/jslhr.4304.902

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory Temporal Integration in Autism Spectrum Disorders. *The Journal of Neuroscience*, *34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Stone, W. L., & Yoder, P. J. (2001). Predicting Spoken Language Level in Children with Autism Spectrum Disorders. *Autism*, *5*(4), 341–361. https://doi.org/10.1177/1362361301005004002

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

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Aud. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (1st ed., pp. 335–364). Wiley. https://doi.org/10.1002/9780470939345.ch12

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*(2), 345.

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640*, 251–263. https://doi.org/10.1016/j.brainres.2015.10.035

Toscano, J. C., & McMurray, B. (2010). Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. *Cognitive Science*, *34*(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous Perception and Graded Categorization: Electrophysiological Evidence for a Linear Relationship Between the Acoustic Signal and Perceptual Encoding of Speech. *Psychological Science, 21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Trehub, S. E., Cohen, A. J., Thorpe, L. A., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 295.

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M.-A., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive Neuropsychology*, *27*(6), 463–476. https://doi.org/10.1080/02643294.2011.567258

Turan, F., & Okcun Akcamus, M. C. (2012). An Investigation of The Development of Imitation Skill in Children with Autism and Its Relations with Receptive-Expressive Language Development and The Development of Symbolic Play. *Turkish Journal of Psychiatry*. https://doi.org/10.5080/u7007

Valla, J. M., & Belmonte, M. K. (2013). Detail-oriented cognitive style and social communicative deficits, within and beyond the autism spectrum: Independent traits that grow into developmental interdependence. *Developmental Review*, *33*(4), 371–398. https://doi.org/10.1016/j.dr.2013.08.004

Van Bourgondien, M. E., & Woods, A. V. (1992). Vocational possibilities for high-functioning adults with autism. In *High-functioning individuals with autism* (pp. 227–239). Springer.

Van Santen, J. P. H., Prud'hommeaux, E. T., Black, L. M., & Mitchell, M. (2010). Computational prosodic markers for autism. *Autism: The International Journal of Research and Practice, 14*(3), 215–236. https://doi.org/10.1177/1362361309363281

Venezia, J. H., Martin Allison-Graham, Hickok Gregory, & Richards Virginia M. (2019). Identification of the Spectrotemporal Modulations That Support Speech Intelligibility in Hearing-Impaired and Normal-Hearing Listeners. *Journal of Speech, Language, and Hearing Research*, *62*(4), 1051–1067. https://doi.org/10.1044/2018_JSLHR-H-18-0045

Vivanti, G., & Hamilton, A. (2014). Imitation in Autism Spectrum Disorders. In *Handbook of Autism and Pervasive Developmental Disorders, Fourth Edition*. American Cancer Society. https://doi.org/10.1002/9781118911389.hautc12

Vivanti, G., Hamner, T., & Lee, N. R. (2018). Neurodevelopmental Disorders Affecting Sociability: Recent Research Advances and Future Directions in Autism Spectrum Disorder and Williams Syndrome. *Current Neurology and Neuroscience Reports*, *18*(12), 94. https://doi.org/10.1007/s11910-018-0902-y

Vivanti, G., Trembath, D., & Dissanayake, C. (2014). Mechanisms of Imitation Impairment in Autism Spectrum Disorder. *Journal of Abnormal Child Psychology*, *42*(8), 1395–1405. https://doi.org/10.1007/s10802-014-9874-9

Vuvan, D. T., Nunes-Silva, M., & Peretz, I. (2015). Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia. *Cortex*, *69*, 186–200. https://doi.org/10.1016/j.cortex.2015.05.002

Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, *7*(1), 43254. https://doi.org/10.1038/srep43254

Warren, P. (2005). Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change*, *17*, 209–230. https://doi.org/10.1017/S095439450505009X

Wechsler, D. (2003). *Wechsler intelligence scale for children–Fourth Edition (WISC-IV)*. TX: The Psychological Corporation.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, *31*(4), 749–778. https://doi.org/10.1017/S030500090400652X

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

Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). Neuropsychologic Functioning in Children with Autism: Further Evidence for Disordered Complex Information-Processing. *Child Neuropsychology*, *12*(4–5), 279–298. https://doi.org/10.1080/09297040600681190

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, *25*(4), 287–295. https://doi.org/10.1016/S0149-7634(01)00014-8

Williamson, V. J., Liu, F., Peryer, G., Grierson, M., & Stewart, L. (2012). Perception and action de-coupling in congenital amusia: Sensitivity to task demands. *Neuropsychologia*, *50*(1), 172–180. https://doi.org/10.1016/j.neuropsychologia.2011.11.015

Wisniewski, M., Mantell, J., & Pfordresher, P. (2013). Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind, and Brain*, *23*, 82. https://doi.org/10.1037/a0033299

Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical Language Impairment in Autism Spectrum Disorder: Exploring Language Phenotypes Beyond Standardized Testing. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00532

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(04), 565–585.

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

Xie, X., Buxó-Lugo, A., & Kurumada, C. (2021). Encoding and decoding of meaning through structured variability in intonational speech prosody. *Cognition*, *211*, 104619. https://doi.org/10.1016/j.cognition.2021.104619

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*(3–4), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

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

Xu, Y., & Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication*, *57*, 181–208. https://doi.org/10.1016/j.specom.2013.09.013

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337.

Yang, J., & Hofmann, J. (2016). Action observation and imitation in autism spectrum disorders: An ALE meta-analysis of fMRI studies. *Brain Imaging and Behavior*, *10*(4), 960–969. https://doi.org/10.1007/s11682-015-9456-7

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

Young, G. S., Rogers, S. J., Hutman, T., Rozga, A., Sigman, M., & Ozonoff, S. (2011). Imitation from 12 to 24 months in autism and typical development: A longitudinal Rasch analysis. *Developmental Psychology*, *47*(6), 1565–1578. https://doi.org/10.1037/a0025418

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

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161

Zhang, J., Meng, Y., Wu, C., Xiang, Y.-T., & Yuan, Z. (2019). Non-speech and speech pitch perception among Cantonese-speaking children with autism spectrum disorder: An ERP study. *Neuroscience Letters*, *703*, 205–212. https://doi.org/10.1016/j.neulet.2019.03.021

Appendices for Study 1

Appendix A:

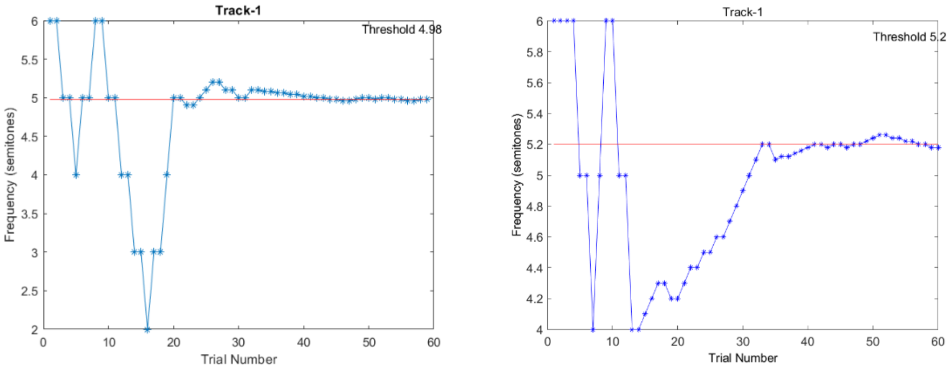


Figure 1S. Visual tracks of two non-compliant performers on the pitch direction discrimination task

Appendix B:

Table 1S. Characteristics of existing participants on the pitch direction discrimination task

Age group	Diagnostic group	Age	Music training	NVIQ	ROWPVT-IV	Corsi	Digit span	AQ	EQ	SQ
Children ASD = 12 control = 10	ASD	9.28(1.23)	1.54(2.42)	72.68(27.59)	123.00(10.11)	4.67(0.99)	5.42(1.00)	96.58(24.55)	19.42(6.30)	30.08(11.21)
	control	9.71(1.48)	1.50(0.97)	80.00(23.92)	124.80(12.79)	5.50(1.43)	5.80(0.79)	44.20(19.60)	39.90(9.70)	26.40(4.86)
	Comparison statistics: Bayesian									
	W	72.5	74.5	66.5	50	76	78	9	112	46
	BF_{01}	1.96	2.11	2.29	2.46	1.58	1.75	**0.12**	**0.16**	2.06
	Median	-0.26	-0.18	-0.11	0.07	-0.36	-0.30	1.09	-1.18	0.21
	95%CI	[-1.06,0.42]	[-0.97,0.53]	[-0.90,0.60]	[-0.65,0.82]	[-1.23,0.37]	[-1.15,0.41]	[0.20,2.11]	[-2.18,-0.21]	[-0.49,1.02]
Adolescents ASD = 9 control = 10	ASD	13.87(1.44)	3.56(3.21)	51.11(30.19)	115.78(19.80)	5.78(1.64)	5.67(1.12)	37.22(6.02)	13.00(6.67)	47.60(11.71)
	control	13.77(1.09)	3.10(3.00)	77(20.17)	134.3(12.46)	6.20(1.40)	6.20(0.79)	15.40(7.31)	45.63(12.46)	34.88(13.94)
	Comparison statistics: Bayesian									
	W	44.5	39.5	69.5	69.5	52	62	0	40	9
	BF_{01}	2.46	2.13	**0.58**	**0.83**	2.13	1.55	**0.07**	**0.28**	1.15
	Median	-0.002	0.16	-0.68	-0.59	-0.16	-0.36	1.29	-1.13	0.51
	95%CI	[-0.77,0.77]	[-0.60,0.94]	[-1.67,0.14]	[-1.55,0.22]	[-0.97,0.61]	[-1.28,0.43]	[0.28,2.43]	[-2.56,0.00]	[-0.36,1.69]
Adults ASD = 12 control = 18	ASD	37.68(14.45)	5.83(7.50)	50.00(32.12)	107.83(11.55)	5.58(1.68)	7.25(1.96)	36.67(10.33)	20.92(5.55)	79.33(33.86)
	control	35.34(12.88)	5.14(7.03)	41.94(29.06)	108.78(13.61)	6.06(1.00)	7.06(1.11)	15.06(6.53)	48.22(13.74)	49.89(16.31)
	Comparison statistics: Bayesian									
	W	97.5	94	94.5	115.5	134	104	12	215	55
	BF_{01}	2.45	2.41	2.56	2.76	1.86	2.66	**0.03**	**0.02**	**0.44**
	Median	0.17	0.16	0.10	-0.05	-0.29	0.03	1.21	-1.42	0.66
	95%CI	[-0.44,0.81]	[-0.47,0.86]	[-0.53,0.79]	[-0.67,0.61]	[-1.03,0.39]	[-0.62,0.71]	[0.38,2.05]	[-2.33,-0.52]	[-0.02,1.43]

Note: Age and Musical training are in years; NVIQ and ROWPVT-IV are percentile points of nonverbal IQ and standard scores of receptive verbal ability respectively; Corsi and Digit span are the raw scores of nonverbal and verbal short-term memory respectively; AQ, EQ and SQ are the scores of Autism Spectrum, Empathy and Systemizing Quotient respectively. Bayes factors from a default prior 2-tailed Bayesian Mann-Whitney-Wilcoxon Test are expressed in terms of the Bayes factor in favour of the null hypothesis of no difference (BF01). The delta effect size in these Bayesian comparisons is given by the median of a posterior distribution and 95% credible intervals.

Appendices for Study 2

Appendix A:

Table 1S:

Stimuli	**Female model**							
	Median F0 (st)		**Pitch interval (st)**		**Duration (ms)**		**IOI (ms)**	
	Speech	**Music**	**Speech**	**Music**	**Speech**	**Music**	**Speech**	**Music**
She was here./?	93.89	95.38	1.51	2.6	197.32	429.1	220.87	501.8
They went home./?	93.25	95.48	1.05	2.41	246.3	481.6	266.36	558.2
He ate it all./?	92.98	96.09	2.12	2.24	215.48	529.5	233.96	553
He ran a mile./?	93.93	95.71	1.37	2.94	216.39	526.4	187.47	595.7
She bought apples./?	93.76	95.6	3.2	2.59	175.76	563.6	260.22	567.3
She wrote a book./?	93.17	95.57	0.95	2.79	166.36	472.6	218.06	575.9
He lost his boots./?	93.56	95.39	1.39	2.8	160.64	317.2	265.67	502.8
She parked the car./?	93.75	95.75	0.97	2.83	157.75	434.6	232.13	540.5
He washed the dishes./?	93.3	96.7	2.07	2.28	130.96	442.3	216.74	569.2
They forgot her name./?	94.22	96.06	1.9	2.83	178.1	456.2	236.24	514.6
They went to the store./?	92.91	96.22	0.99	2.1	147.85	411.5	202.8	518.2
They finished the test./?	94.05	96.15	1.53	2.15	110.5	329	198.45	500.5
Paired t-test (two-tailed)	***p* < .001*****		***p* < .001*****		***p* < .001*****		***p* < .001*****	
	Male model							
She was here./?	86.69	81.04	2.62	2.65	166.51	549.9	196.27	632.4
They went home./?	83.54	80.98	5.28	3.21	234.64	494.6	259.68	546.4
He ate it all./?	83.19	82.29	3.59	3.31	212.31	472.7	159.3	516.8
He ran a mile./?	83.32	81.89	4.79	2.82	201.13	534.5	188.28	562.4
She bought apples./?	82.92	82.59	5.39	3.81	123.52	424.5	214.47	543
She wrote a book./?	85.46	82.18	3.46	2.84	121.85	391.6	183.26	510.5
He lost his boots./?	83.73	82.96	5.56	3.58	142.03	318.3	264.78	502.9
She parked the car./?	84.1	83	3.3	4.19	123.91	443.5	217.11	553.3
He washed the dishes./?	84.19	82.84	3.9	3.38	116.19	374.5	226.59	536
They forgot her name./?	84.28	82.75	3.55	2.94	183.57	450.7	208.89	535.4

They went to the store./?	84.08	82.93	4.77	3.58	153.79	446.4	212.95	579.1
They finished the test./?	84.1	82.89	4.16	3.72	86.91	351.2	190.57	562.7
Paired t-test (two-tailed)	***p* < .001*****		***p* < .001*****		***p* < .001*****		***p* < .001*****	
	Child model							
She was here./?	94.63	95.69	1.75	2.27	180.91	372.9	230.19	492.9
They went home./?	94.22	95.56	1.55	2.28	261.74	401.5	284.94	498.7
He ate it all./?	94.69	96.27	1.68	2.08	208.69	439.2	274.82	491
He ran a mile./?	94.44	96	1.12	2.45	199.46	383.4	197.72	510.5
She bought apples./?	95.22	95.71	2.4	3.11	190.27	414.7	323.97	493.5
She wrote a book./?	94.75	95.29	1.64	3.18	167.23	351.6	213.32	479.2
He lost his boots./?	94.99	95.86	2.27	2.97	141.66	290.6	261.37	467.1
She parked the car./?	95.05	95.68	1.46	3.23	173.39	398.7	251.5	536.8
He washed the dishes./?	93.6	96.62	2.6	2.19	138.58	388.3	225.38	535
They forgot her name./?	94.4	96.17	1.66	2.41	203.12	403.3	269.82	468.6
They went to the store./?	94.84	96.17	1.42	2.01	155.89	356.3	202.69	482.3
They finished the test./?	95.07	96.26	1.62	1.85	116.44	296	190.48	479.9
Paired t-test (two-tailed)	***p* < .001*****		***p* < .001*****		***p* < .001*****		***p* < .001*****	