Accommodation to Vocal Pitch in Children with Autism

Anqi Hu[1], Zhenghan Qi[1, 2], Kathryn Franich[1, 3]

[1]University of Delaware, [2]Northeastern University, [3]Harvard University
anqihu@udel.edu

ABSTRACT

Speech accommodation is a common phenomenon in typically developing (TD) adults. However, little is known regarding whether TD children and autistic children, who have atypical prosodic profiles and social communication skills, also spontaneously accommodate to others' speech. Thirty 5-to-10-year-old autistic children and 30 age-matched TD children repeated sentences after two model talkers either speaking with their original pitch or with artificially raised pitch. Acoustic analysis revealed that only the autistic children displayed evidence of pitch accommodation. In a follow-up experiment, 25 TD children were explicitly asked to copy the voice of the talkers; here, they showed a similar degree of pitch accommodation to the autistic children. We discuss results in the context of developmental patterns of speech accommodation and suggest TD children were less likely to accommodate to pitch changes due to their communicative function, and to the slightly less typical-sounding voice of the higher-pitch model talker.

Keywords: speech accommodation, f0, pitch, convergence, imitation, autism, speech development

1. INTRODUCTION

Speech accommodation (also known as speech *entrainment, imitation,* or *convergence*) is a process of adapting to the speech characteristics of another talker. Robust evidence has demonstrated that adults can accommodate to different dimensions of speech including pitch, speech rate, vowel spectra, and voice onset time of model talkers or conversation partners in both experimental contexts and naturalistic speech settings [1].

Although speech accommodation has been under-investigated in children, existing studies in children found evidence of phonetic accommodation using measures of fundamental frequency (F0), voice onset time (VOT), and lexical stress [2]–[4]. For example, Ko and colleagues [2] examined pitch convergence between toddlers and their mothers over the course of one to two months and found greater similarities in pitch between mother-and-child dyads in their last conversational turn compared to the first one. Nielsen [3] examined accommodation to VOT in 4-year-old and 8-year-old children and adults. All age groups showed effect of accommodation and produced /p/ with lengthened VOT after exposure, but the effect was stronger in children than in adults, with comparable effects between younger and older children. These studies hinted at the spontaneous nature of phonetic accommodation in children and a potential reduced speech accommodation effect across development.

Children's ability to accommodate to a model has been studied as a broad cognitive phenomenon for decades under the umbrella term "imitation". Robust evidence has demonstrated that children can imitate gestures, actions, and facial expressions since early infancy [5]. Speech accommodation can be considered as an example of spontaneous imitation that occurs automatically without explicit instructions [6]. It has been proposed that social affiliations and liking towards a model talker, as well as the talker's proficiency, social status, and resemblance to oneself can all influence likelihood of imitation [7]. In the speech domain, children are found to imitate more when models are more friendly [8], speak with a native accent as opposed to a foreign accent [9], or are adults as opposed to children [10]. It has also been argued that the development of these social biases during imitation are contingent upon children's age and development of different cognitive skills [7]. Although there is still debate regarding what kind of contexts and type of biases can elicit different profiles of imitation, existing evidence demonstrates that typically developing (TD) children show spontaneous imitation across domains from an early age, and the fidelity of their imitation is guided by the development of different social biases towards the models.

Children with autism spectrum disorder (ASD) often display atypical prosodic profiles [11]. Yet, little is known regarding the processes of speech accommodation in autism, and whether speech accommodation is impacted by social biases in this population with social communication impairment. Studies of speech accommodation in autistic children could provide important information about the source of their prosodic differences.

Around 1 in 40 to 59 children in the US are diagnosed with ASD [12]. Among individuals with autism that do acquire spoken language, atypical speech prosody has been reported and included as a crucial diagnostic criterion [13]. In the domain of pitch, the focus of the present study, several studies have reported similar mean pitch or pitch variability in sentences between individuals with autism and controls [14], [15], though others have found a higher mean pitch and a wider pitch range in ASD [16], [17].

In one study with 9-to-15-year-old autistic children, Lehnert-LeHouillier and colleagues examined speech accommodation during an interactive picture description task led by a female experimenter [18]. Only 25% of autistic children showed increasing similarity in their mean F0 values to the experimenter over the course of the task, compared to 66% in the TD group. While the TD group converged to the pitch of the experimenter, the similarity in pitch found in the ASD group was mainly driven by speech accommodation of the experimenter.

Given the role that social dynamics can play in accommodation, it isn't yet clear from these results whether autistic children cannot accommodate to adults as well as TD children, or whether they may have failed to accommodate due to some aspect of the social dynamic in the study. Autistic children are known to display greater speech and language deficits in contexts involving higher cognitive and social demand [19]; therefore, a socially-demanding interactive task may have made speech processing—and, hence, accommodation—more challenging for the ASD group. In the present work, we explore speech accommodation among children with ASD in a less socially demanding context, that of a sentence repetition task. This type of task—where a participant listens to the speech of a model talker and repeats what has been said—have been used to successfully elicit accommodation in adults and children in previous work [20]. We hypothesize that, if children with ASD are going to display accommodation to a model talker, they may be more likely to do so in this type of simple repetition context.

2. METHODS

2.1. Participants

Thirty children with ASD (6.42 – 9.50 years; 3 girls) and 30 age-matched TD children (4.67 – 10.92 years; 14 girls) participated in this web-based study. Children with ASD were recruited by SPARK from nationwide clinical sites [21]. The autism diagnoses of children in SPARK were validated by medical records [22] and confirmed using Social Communication Questionnaire [23] and Repetitive Behavior Scale – Revised (RBS-R) [24]. All children with ASD were native speakers of English with no history of hearing loss, mutism, or (non-verbal) learning disabilities based on parental questionnaires. As will be discussed, an additional 25 TD children of comparable age (9 girls) were recruited for a follow-up experiment. TD children were neurotypical and had no history of speech or language impairments.

2.2. Stimuli and Design

The auditory Sentence Recall task (RSR) [25] was used to assess children's ability to accommodate to a model talker. This task has been validated in children from ages 5 to 9 and showed high sensitivity and specificity for detecting language impairment in this age range [25]. Sixteen sentences were included. Each sentence contained on average 11 words and 13.38 syllables (*SD* = 1.31). All sentences were monoclausal and contained a single intonational phrase and all discourse-new information.

In the **lower-pitch condition**, the auditory stimulus used was the original audio recording from Redmond [25]. A female speaker of standard American English pre-recorded all the sentences with a naturalistic and child-friendly prosody. In the **higher-pitch condition**, a different new female speaker of standard American English recorded the 16 sentences (mean duration = 4 seconds). To examine whether children will accommodate to the pitch of the model talkers, we artificially manipulated the pitch of the new model talker to be 50 Hz higher than the original talker in Redmond [25] using the Time-Domain Pitch-Synchronous Overlap-and-Add (TD-PSOLA) method in Praat [26], [27]. A norming study eliciting impressions from 11 adults showed that most listeners (n = 9) considered the speech of the higher-pitch model talker natural, clear, and child-friendly, but two adults mentioned the speech sounded slightly monotone. Other dimensions of the two speakers' speech (e.g., vowel spaces and volume) were similar, so that the primary dimension on which the two model talkers differed was pitch. The order of the 16 sentences was the same across conditions.

2.3. Procedure

The sentence repetition tasks were completed entirely through a video conference. The lower-pitch and higher-pitch conditions were arranged in a block design and the two blocks had a minimum of 50-minute interval between them to prevent interference. To avoid the confounding effect of fatigue on pitch lowering, the lower-pitch condition was always completed before the higher-pitch condition.

Upon starting the tasks, a pre-recorded audio instruction by a female native speaker of English told the child participants that they would play a talking game during which they would listen to some sentences from a girl, and they should "repeat exactly what you hear." In the higher-pitch condition, children were given the same instruction to repeat the sentences. Here, no instruction was given regarding the pitch differences between conditions. As will be discussed, only the ASD group showed the expected pattern of spontaneous accommodation. In order to verify that TD participants could show similar patterns to the ASD group, a follow-up experiment was conducted with 25 TD participants which differed only in the instructions given: here, children were explicitly instructed to copy exactly both the voice and the words from the model talkers in both conditions.

2.4. Analyses

Mean F0 values were extracted from each unreduced vowel in the corpus using the pitch tracking algorithm in ProsodyPro [28]. To quantify pitch accommodation, we followed Babel & Bulatov [29] and calculated the absolute distance between the F0 of the model talkers and that of the participants. For each vowel and each participant, the ERB-transformed F0 of the vowel produced by the participant was deducted from that of the corresponding vowel produced by the model talker (F0 distance = |vowel F0 model talker – vowel F0 participant|). The pairs of vowels being compared here were matched on the word and sentence levels so that the same vowels are being compared between the participants and the model talkers. The distance in F0 was calculated separately for the lower-pitch and for the higher-pitch condition.

3. RESULTS

3.1. Group Differences in Pitch Accommodation

3.1.1. Mean F0

We first compared the F0 values between groups to determine the direction of pitch shifts in each condition. We analysed the fixed effects of *Group* (TD vs. pooled ASD and TD Follow-up; ASD vs. TD Follow-up), *Condition* (lower-pitch vs. higher-pitch), and their interactions on ERB-scaled F0 values using a mixed-effects linear regression model [30]. The model also included age, sentence recall accuracy scores, the total number of vowel tokens, and baseline pitch, extracted from the mean F0 values of the first 5 sentences in the lower-pitch condition, as covariates. Random intercepts for participants were included.

Collapsed across groups, we found a main effect of *Condition* ($\beta = 0.08$, $SE = 0.02$, $t = 3.98$, $p < 0.001$), indicating a significantly higher pitch in the higher-pitch than lower-pitch condition across groups. There was a significant interaction between *Condition* and *TD vs. pooled ASD and TD Follow-up group* ($\beta = -0.20$, $SE = 0.04$, $t = -5.43$, $p < 0.001$), reflected as a larger difference in mean F0 values between conditions in ASD and TD Follow-up groups than the TD group. The interaction between *Condition* and *ASD vs. TD Follow-up group* was not significant. Post-hoc *t*-tests within each group revealed a significantly higher mean F0 in the higher-pitch than the lower-pitch condition in the ASD ($p < 0.001$, $d = 0.13$) and the TD Follow-up groups ($p = 0.01$, $d = 0.08$; **Fig. 1**). No significant difference was found between conditions in the TD group.

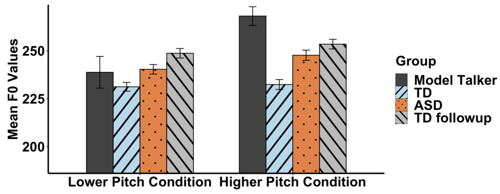

Figure 1: Mean F0 values in model talkers and children. The error bars represent 95% confidence interval.

3.1.2. F0 Distance

To examine the group differences in pitch accommodation, we analysed the fixed effects of *Group*, *Condition*, and their interactions on F0 distance using a similar model in the Mean F0 analysis. We found significant interactions between *Condition* and *Group* (*TD vs. pooled ASD and TD Follow-up*: $\beta = 0.18$, $SE = 0.03$, $t = 5.50$, $p < 0.001$; *ASD vs. TD Follow-up*: $\beta = 0.13$, $SE = 0.04$, $t = 3.47$, $p = 0.001$), suggesting condition difference was largest in the TD Follow-up and smallest in the TD group with the ASD group in between. Post-hoc *t*-tests within each group confirmed a significantly smaller mean F0 distance to the model talker in the higher-pitch than the lower-pitch condition in the ASD ($p = 0.003$, $d = 0.09$) and the TD Follow-up groups ($p < 0.001$, $d = 0.13$), indicating pitch accommodation to the higher-pitch model talker in the ASD and TD Follow-up groups (**Fig. 2**). The TD group showed a significantly larger F0 distance in the higher-pitch than the lower-pitch condition ($p < 0.001$), consistent with a lack of pitch accommodation to the model talker.

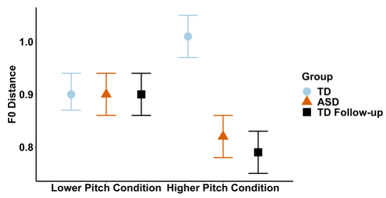

Figure 2: Mean F0 distance across groups. The error bars represent 95% confidence interval.

3.2. Impressions of Model Talker

We further asked all children in the TD Follow-up group about their impressions of the model talkers in order to examine the effect of social biases on speech accommodation. Children answered three questions: 1). Do the two talkers sound different? 2). How are their voices different? 3). Based on the voices, what do you think each one of them is like? All children reported the voices sounded different. 60% of the children (n = 15) provided specific descriptions on the vocal characteristics (e.g., "higher", "lower", "louder", "deeper voice", "squeaky voice") of the model talkers in question 2. In particular, three children noted that the lower-pitch model talker sounded more "normal" while the higher-pitch model talker sounded "unreal". In question 3, 64% of the children (n = 16) associated the voice of the model talkers with different social attributes (e.g., "more shy", "really active", "more relaxed", "chattier and outgoing", "not as much excited", "happy", "sad", "nicer"). Nine children noted the higher-pitch model talker may be more shy, quiet, or less excited.

4. DISCUSSION

The current study compared speech accommodation between autistic and TD children. Children with autism were able to spontaneously accommodate to the pitch of the model talker even when explicit imitation of pitch was not instructed. The TD children, however, only showed pitch accommodation when they were explicitly asked to copy the voice of the model talker.

One reason that may explain the group differences could be related to the aspect of speech children paid attention to during tasks. When multiple signals are present in a stimulus, ample behavioural evidence have demonstrated that children with autism show atypical attention patterns to certain dimensions of the signal [31]. For example, in speech perception tasks by Ploog and colleagues [32], the TD children showed a strong preference for content so that they chose the test sentences that had the same lexicon but different prosody compared to the training sentences, while the children with autism showed no preference between content and prosody, despite intact language comprehension and pitch perception across groups. Similarly, in our study, the TD group only showed accommodation to pitch when explicitly instructed to copy the voice of the model talkers. When only asked to repeat the sentences, the TD group may have perceived the semantic content and lexicon in the sentences more functionally important or more task-relevant than the pitch of the model talkers, since pitch only varied as a function of individual talker vocal characteristics, and did not contribute to a change in meaning across conditions The TD group thus attended selectively to the content and only copied the word form used and the semantic information in the sentences with their own pitch. The ASD group, however, may have paid attention to all attributes of speech from the model talkers and accommodated to both details, the pitch and content. Nevertheless, the role of selective attention can only be inferred in our current paradigm. Future studies using more implicit electrophysiological or other brain imaging techniques to measure auditory attention will be needed to directly assess whether children engage more attention to meaning or to prosody during speech accommodation.

Differences in social biases towards the model talkers may also explain the group differences in pitch accommodation. Previous studies have found that TD adults were more likely to accommodate to model talkers that were deemed more attractive [33]. TD children at the age of 3 to 5 years were found to copy more novel word labels and actions from proficient and knowledgeable models than models who provide inconsistent or inaccurate information. TD children at age 4 showed more bias towards models with higher social status who were endorsed by bystanders than those that were not [34]. The effect of social biases on imitation develops and becomes more robust as children grow older [7]. In our study, the TD children in the follow-up experiment seemed to associate the voice of the model talkers with different social attributes ("chattier" vs. "more shy"). Some TD participants commented on the voice of the higher-pitch model talker being less typical (e.g., "less normal", "unreal voice") than the lower-pitch model talker. Given that 4-to-5-year-old children were already found to show social biases in imitation, the markedness of the higher-pitch model talker might explain the lack of pitch accommodation in 5-to-10-year-old TD children in this study if they considered the higher-pitch model talker as less likable or less proficient. Compared to TD individuals, autistic adults were found to show attenuated social biases and stereotypes in tasks where their tendency to associate a social group with particular attributes were tested implicitly [35]. If the attenuated social biases in autistic adults are an outcome of atypical development of social communication skills, then the pitch accommodation to the higher-pitch model talker in the autistic children may be explained by a delayed development of social biases in this group. Future studies examining accommodation in naturalistic and conversational settings would need to investigate whether autistic children form the same social attributes based on the voice of the model talkers and whether that influences their pitch accommodation.

5. REFERENCES

[1] J. S. Pardo, “Measuring phonetic convergence in speech production,” *Front. Psychol.*, vol. 4, no. AUG, p. 559, Aug. 2013.

[2] E. S. Ko, A. Seidl, A. Cristia, M. Reimchen, and M. Soderstrom, “Entrainment of prosody in the interaction of mothers with their young children,” *J. Child Lang.*, vol. 43, no. 2, pp. 284–309, Mar. 2016.

[3] K. Nielsen, “Phonetic Imitation by Young Children and Its Developmental Changes,” *J. Speech, Lang. Hear. Res.*, vol. 57, no. 6, pp. 2065–2075, Dec. 2014.

[4] F. Subiaul, K. Winters, K. Krumpak, and C. Core, “Vocal overimitation in preschool-age children,” *J. Exp. Child Psychol.*, vol. 141, pp. 145–160, 2016.

[5] S. S. Jones, “The development of imitation in infancy,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 364, no. 1528, pp. 2325–2335, Aug. 2009.

[6] T. L. Chartrand and J. L. Lakin, “The antecedents and consequences of human behavioral mimicry,” *Annu. Rev. Psychol.*, vol. 64, pp. 285–308, 2013.

[7] L. A. Wood, R. L. Kendal, and E. G. Flynn, “Whom do children copy? Model-based biases in social learning,” *Developmental Review*, vol. 33, no. 4. Academic Press, pp. 341–356, 01-Dec-2013.

[8] M. Nielsen, “Copying actions and copying outcomes: Social learning through the second year,” *Dev. Psychol.*, vol. 42, no. 3, pp. 555–565, 2006.

[9] K. D. Kinzler, K. H. Corriveau, and P. L. Harris, “Children’s selective trust in native-accented speakers,” *Dev. Sci.*, vol. 14, no. 1, p. 106–111, 2011.

[10] V. K. Jaswal and L. A. Neely, “Adults don’t always know best preschoolers use past reliability over age when learning new words,” *Psychol. Sci.*, vol. 17, no. 9, pp. 757–758, Sep. 2006.

[11] J. McCann and S. Peppé, “Prosody in autism spectrum disorders: A critical review,” vol. 38, no. 4. *Int J Lang Commun Disord*, pp. 325–350, 2003.

[12] M. D. Kogan *et al.*, “The prevalence of parent-reported autism spectrum disorder among US children,” *Pediatrics*, vol. 142, no. 6, 2018.

[13] R. Fusaroli, A. Lambrechts, D. Bang, D. M. Bowler, and S. B. Gaigg, “‘Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis,’” *Autism Res.*, vol. 10, no. 3, 2017.

[14] J. J. Diehl and R. Paul, “Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders,” *Res. Autism Spectr. Disord.*, vol. 6, no. 1, pp. 123–134, 2012.

[15] H. Green and Y. Tobin, “Prosodic analysis is difficult ... but worth it: A study in high functioning autism,” *International Journal of Speech-Language Pathology*, vol. 11, no. 4 pp. 308–315, 2009.

[16] M. Sharda *et al.*, “Sounds of melody—Pitch patterns of speech in autism,” *Neurosci. Lett.*, vol. 478, no. 1, pp. 42–45, Jun. 2010.

[17]A. Nadig and H. Shaw, “Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners,” *J. Autism Dev. Disord.*, vol. 42, no. 4, pp. 499–511, Apr. 2012.

[18] H. Lehnert-LeHouillier, S. Terrazas, and S. Sandoval, “Prosodic Entrainment in Conversations of Verbal Children and Teens on the Autism Spectrum,” *Front. Psychol.*, vol. 0, p. 2718, Oct. 2020.

[19] M. Losh and L. Capps, “Narrative ability in high-functioning children with autism or Asperger’s Syndrome,” *J. Autism Dev. Disord.*, vol. 33, no. 3, pp. 239–251, Jun. 2003.

[20] L. Wang, P. Q. Pfordresher, C. Jiang, and F. Liu, “Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation,” *Autism Res.*, vol. 14, no. 11, pp. 2355–2372, 2021.

[21] P. Feliciano *et al.*, “SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research,” *Neuron*, vol. 97, no. 3. Cell Press, pp. 488–493, 07-Feb-2018.

[22] E. Fombonne, L. Coppola, S. Mastel, and B. J. O’Roak, “Validation of Autism Diagnosis and Clinical Data in the SPARK Cohort,” *J. Autism Dev. Disord.*, vol. 52, no. 8, pp. 3383–3398, Aug. 2022.

[23] M. Rutter, A. Bailey, and C. Lord, *The social communication questionnaire: Manual*. 2003.

[24] J. Bodfish, F. Symons, and M. Lewis, *The Repetitive Behavior Scale*. Western Carolina Center Research Reports, 1999.

[25] S. M. Redmond, “Differentiating SLI from ADHD using children’s sentence recall and production of past tense morphology,” *Clin. Linguist. Phonetics*, vol. 19, no. 2, pp. 109–127, 2005.

[26] E. Moulines and F. Charpentier, “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,” *Speech Commun.*, vol. 9, no. 5–6, pp. 453–467, Dec. 1990.

[27] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glot Int.*, vol. 5, no. 9, pp. 341–345, 2001.

[28] Y. Xu, “ProsodyPro—A tool for large-scale systematic prosody analysis,” 2013.

[29] M. Babel and D. Bulatov, “The Role of Fundamental Frequency in Phonetic Accommodation,” *Lang. Speech*, vol. 55, no. 2, pp. 231–248, Jun. 2012.

[30] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, “Fitting linear mixed-effects models using lme4,” *J. Stat. Softw.*, vol. 67, no. 1, 2015.

[31] B. O. Ploog, “Stimulus Overselectivity Four Decades Later: A Review of the Literature and Its Implications for Current Research in Autism Spectrum Disorder,” *J. Autism Dev. Disord. 2010 4011*, vol. 40, no. 11, pp. 1332–1349, Mar. 2010.

[32] B. O. Ploog, S. Banerjee, and P. J. Brooks, “Attention to prosody (intonation) and content in children with autism and in typical children using spoken sentences in a computer game,” *Res. Autism Spectr. Disord.*, vol. 3, no. 3, pp. 743–758, Jul. 2009.

[33] M. Babel, “Evidence for phonetic and social selectivity in spontaneous phonetic imitation,” *J. Phon.*, vol. 40, no. 1, pp. 177–189, Feb. 2012.

[34] M. Fusaro and P. L. Harris, “Children assess informant reliability using bystanders’ non-verbal cues,” *Dev. Sci.*, vol. 11, no. 5, pp. 771–777, 2008.

[35] E. Birmingham, D. Stanley, R. Nair, and R. Adolphs, “Implicit Social Biases in People With Autism,” *Psychol. Sci.*, vol. 26, no. 11, pp. 1693–1705, Nov. 2015.