Can adolescents with autism perceive emotional prosody?

Cristiane Hsu, Yi Xu

University College London, UK

cristiane.hsu.11@ucl.ac.uk, yi.xu@ucl.ac.uk

Abstract

Past findings on the perception of emotional prosody by individuals with Autism Spectrum Disorder (ASD) have been incongruent, and a main reason is the lack of clarity about the contributions of specific acoustic features to emotion perception. In this study we test the perception of emotional prosody by adolescents with ASD using a recently developed prosody control method based on a bio-informational dimensions (BID) theory of emotion expressions. We synthesized a Mandarin sentence with different voice qualities using an articulatory synthesizer, and then acoustically manipulated their formant dispersion, median pitch and pitch range. With these utterances we compared the ability to perceive body size, emotion and attitude by high-functioning adolescents with ASD, typically developing adolescents and young adults. Results showed that the three groups made similar perceptual judgements, but the sensitivity of adolescents with ASD to the acoustic manipulations was lower than their typically developing peers, who in turn exhibited less sensitivity than young adults. These findings show that individuals with ASD have a reduced rather than a total lack of ability to perceive emotional prosody, suggesting a delay in their developmental trajectory. The findings also demonstrate the effectiveness of the BID-based method in testing perception of emotional prosody.

Index Terms: ASD, perception, prosody, emotion, attitude

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental condition with two main characteristics: social interaction and communication deficits, and restricted, repetitive and stereotyped behaviours and interests [1]. Although not included as a criterion for ASD diagnosis, prosody has been widely investigated for ASD, and there are suggestions that prosodic ability may be a marker for language and social competence in ASD, since individuals with ASD with better performance in speech prosody also had better language skills and higher socialisation ratings [2], [3].

Instrumental acoustic analysis revealed the 'oddness' in speech prosody of high-functioning individuals with ASD, as they tend to speak with larger pitch range and pitch variation, and rely almost exclusively on pitch to convey emotional differences, while not sufficiently exploiting amplitude and duration cues [4]–[7]. Furthermore, recent empirical evidence suggests these deficits may extend beyond speech production, affecting also their competence in decoding prosody. Findings included difficulties in processing longer sentences, mistaking questions for statements, misidentifying stress/focus, and failing to appreciate emotions in prosody [8]–[10].

Recognition of emotion in prosody has been receiving special attention in research related to ASD, as individuals with ASD have been widely reported to show difficulty in understanding their own and other people's emotion due to deficits in Theory of Mind (ToM), i.e. the ability to infer other people's mental states, based on the understanding of one's own mental state [11]–[14]. However, the general findings are mixed, ranging from significantly poorer performance to compatible competence [15]–[19].

A possible reason for the inconsistent findings is that studies in this field rarely incorporate any specific theory of emotion to motivate the research questions and hypotheses [20], but instead interpret results only through the optics of ToM. Yet, being a theory in psychology, ToM does not take into consideration the basic facts about emotion and speech: both are biological phenomena, and should thus be appreciated from a biological perspective.

In 1977 Morton proposed that many animals use acoustic properties in their vocal calls to influence other animals [21]. When being aggressive, they use low-pitched and rough voice to project a large body size to dominate the opponents; when being associable, they use high-pitched and pure tone-like voice to project a small body size to attract the hearer [22]. Ohala extended Morton's theory by proposing that varying the length of the vocal tract achieves the same body-size projection effect as pitch and voice quality, and that humans also use similar strategies to express dominance and associability [23].

In a series of studies this body-size projection theory was tested in terms of its relevance to the human perception of emotions [22], [24]–[26]. These studies used synthesised Thai vowels, English digits, and English sentences with synthetically manipulated pitch, vocal tract length (VTL) and voice quality, and found that stimuli synthesised with longer VTL (smaller formant dispersion) and lower pitch were perceived as spoken by a larger person and sounding angry, while the ones with shorter VLT (wider formant dispersion) and higher pitch were perceived as from a smaller person and sounding happy [25], [26]. To explain additional acoustic properties in emotional prosody, Xu and colleagues proposed a set of bio-informational dimensions (BIDs), which include body-size projection, dynamicity, audibility and association [26]. Initial evidence for the relevance of these dimensions has been shown by a number of studies [22], [24], [26].

Given their demonstrated effectiveness, the BIDs, especially body-size projection, can be used to assess an individual's sensitivity to emotional information in speech with greater precision than previously used methods. The present study investigated the perception of emotional prosody of Taiwan Mandarin (TM)-speaking high-functioning adolescents with ASD by comparing with typically developing adolescents and average young adults. Specifically, the following research questions were examined or addressed:

- Do high-functioning adolescents with ASD perceive the BID-based acoustic features in the same way as controls when interpreting body size, emotion and attitude?
- Are the three groups equally sensitive to all the acoustic features when perceiving body size, emotion and attitude?

2. METHOD

2.1. Participants

Ten adolescents with high-functioning autism or Asperger syndrome (AA group) were recruited as experimental group, and ten typically developing adolescents and ten average young adults served as two different control groups. All the respondents were native speakers of TM, currently living and studying or working in Taipei or New Taipei City. Written consents were obtained from all the participants and from the legal representatives of those who were under the age of 20.

The adolescents in AA group (8 males and 2 females, age mean = 15;11 (years;months), age range = 13;04–18;11; SD = 2;04) had been formally diagnosed according to DSM-IV-TR criteria by clinicians from teaching hospitals in Taipei, Taiwan, who held the Physically and Mentally Disabled Manual (a primary document issued by Taiwanese Ministry of Interior to receive social welfares, and/or special education) and were registered as having 'mild autism', and had no other learning difficulties or medical conditions unrelated to ASD.

The typically developing adolescents (TA group) (2 females and 8 males; age mean = 15;05; age range = 13;08-18;06; SD = 2;01) and the young adults (5 females and 5 males; age mean = 26;07; age range = 23;08-29;08; SD = 1;09) had no self-reported learning difficulties or communication disorders. All respondents had normal hearing and normal or corrected-to-normal vision.

2.2. Speech material

The stimuli were based on a Mandarin Chinese utterance "wo yu a yi you yue" ("I have (an) appointment with aunty") spoken by a male native speaker of TM, aged 37, in an emotionally 'neutral' voice, then used as a model to create three synthetic version of the same sentence using the articulatory speech synthesiser VocalTractLab [27], in modal, breathy and pressed voice. Similar to the manipulations in [22], the synthesised utterances were further modified by Praat [28] in terms of formant shift ratio, pitch shift and pitch range using a Praat script to generate the auditory stimuli, as summarised in Table 1. The total number of stimuli was 3 voice qualities x 3 formant shift ratios x 3 pitch shifts x 3 pitch ranges = 81.

Table 1. *Acoustic feature manipulations. Formant shift ratio changes formant dispersion, with larger ratio corresponding to greater dispersion. Pitch shift is median pitch relative to the original.*

Voice quality	Formant shift ratio	Pitch shift (st)	Pitch range
Breathy	1.1	2	2
Modal	1	0	1
Pressed	0.9	-2	0.5

2.3. Procedure

The experiment was run by the ExperimentMFC module of Praat software on a laptop computer (HP Pavilion dm3 Notebook PC). All stimuli were delivered from the computer via headphones (Sennheiser HD 265). The respondents could adjust the playback volume to a comfortable level.

The stimuli were presented in random order. The respondents were allowed to listen to each stimulus only once. After listening to a stimulus, the respondent performed a three-choice task. They were instructed to make judgements instinctively without thinking too hard. The same 81 stimuli were used for the three tests: body size, emotion and attitude. The respondents were asked to make a choice according to the test type. For body size, the choices were small, medium and large; for emotion, happy, neutral and angry; and for attitude, friendly, neutral and serious. The responses were computed by the ExperimentMFC module in table format, and coded as:

- Small as 0% large, medium as 50% large, and large as 100% large.
- Happy as 0% angry, neutral as 50% angry, and angry as 100% angry.
- Friendly as 0% serious, neutral as 50% serious, and serious as 100% serious.

3. RESULTS

The coded responses of each test were analysed with 4-way mixed design ANOVA by using SPSS version 22.0 [29].

3.1. Size projection

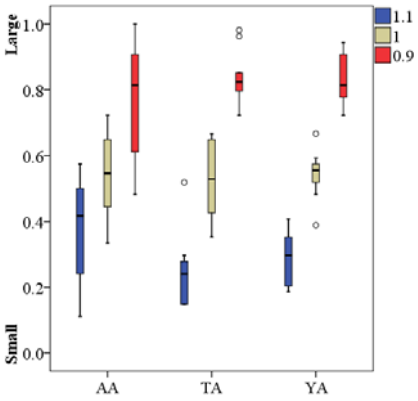

Figure 1: *Formant shift ratio in size projection*

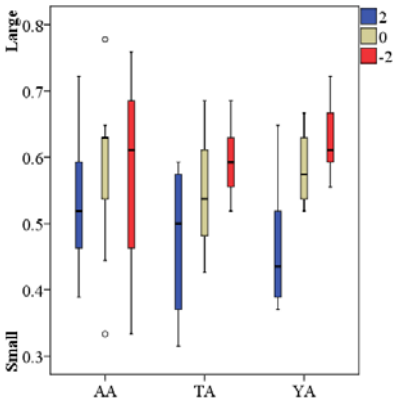

Figure 2: *Pitch shift in size projection*

Figure 1 displays the body size judgements as a function of formant shift ratio by all three groups. As can be seen, all groups rated utterances with narrower formant shift ratio as having a larger body size, while wider formant shift ratio as having a smaller size. However, the AA group showed greater overlap than the other two groups between the three body size judgements. ANOVA results showed that formant shift ratio had significant effect for all three groups, but the significance

level of the AA group ($F(2, 18) = 16.96$, $p < 0.001$) is smaller than both the TA group ($F(2, 18) = 79.05$, $p < 0.001$) and the YA group ($F(2, 18) = 140.09$, $p < 0.001$).

The body size judgement through pitch shift is shown in Figure 2. Although all the three groups rated downward pitch shift as from larger speakers, and upward pitch shift as from smaller ones, pitch shift had significant effect only for the TA group ($F(2, 18) = 11.10$, $p = 0.001$) and the YA group ($F(2, 18) = 24.08$, $p < 0.001$).

Voice quality had a significant effect ($F(2, 18) = 14.47$, $p < 0.001$) only for the YA group when judging body size. However, they perceived breathy voice as from a larger speaker, and pressed voice as from a smaller speaker, and this tendency contradicts findings in [22], [25].

3.2. Emotion

Figure 3 depicts the emotion judgement according to formant shift ratio by the three groups. All of them rated smaller formant shift ratio as sounding angry and greater formant shift ratio as happy. The overlap between the three emotion judgements is small across all groups, and ANOVA results show that formant shift ratio had significant effect for every group, with significance level smaller for the TA group ($F(2, 18) = 17.11$, $p < 0.001$) than the TA group ($F(2, 18) = 35.79$, $p < 0.001$), and YA group $F(2, 18) = 25.93$, $p < 0.001$).

Figure 4 shows the effect of pitch shift on emotion judgement. In general, downward pitch shift is perceived as angry, and upward pitch shift as happy. Once again, ANOVA results show different significance levels in the effect of pitch shift across the three groups: smaller for the AA group ($F(2, 18) = 11.10$, $p = 0.001$) than for the TA group ($F(2, 18) = 19.57$, $p < 0.001$) and the YA group ($F(2, 18) = 24.13$, $p < 0.001$).

Figure 5 shows that for the effect of pitch range the AA group had greater overlap than the other two groups among the three emotion types. ANOVA results show that pitch range has significant effect only for the TA group ($F(2, 18) = 19.55$, $p < 0.001$) and the YA group ($F(2, 18) = 8.60$, $p = 0.002$). Both groups judged smaller pitch ranges as angry, while larger pitch ranges as happy.

The effect of voice quality reached statistical significance for both the TA group ($F(2, 18) = 3.72$, $p = 0.045$) and the YA group ($F(2, 18) = 8.89$, $p = 0.002$). However, these groups rated pressed voice as angry and breathy as happy, which once again went against the findings of [22], [25].

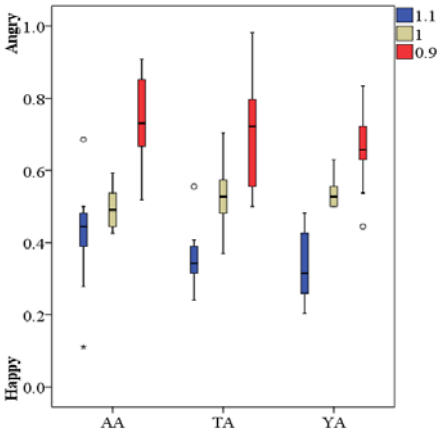

Figure 3: *Formant shift ratio in emotion*

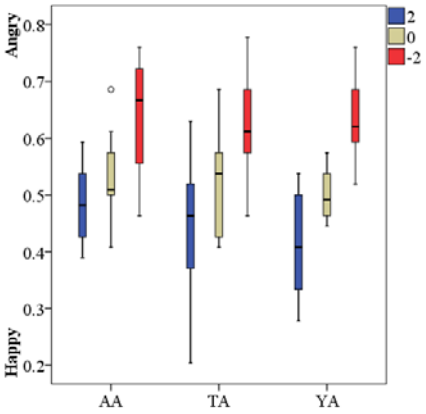

Figure 4: *Pitch shift in emotion*

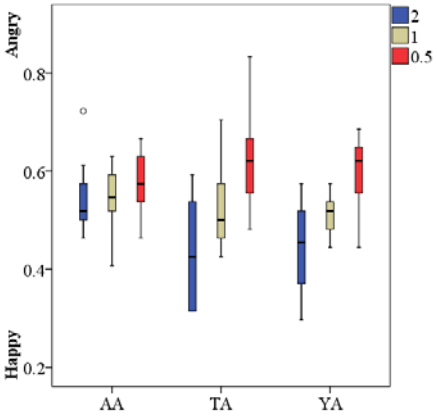

Figure 5: *Pitch range in emotion*

3.3. Attitude

Figure 6 illustrates the attitude judgement based on formant shift ratio for the three groups. All the groups rated smaller formant shift ratio as from a serious speaker, and wider formant shift ratio as from a friendly speaker. There is a small overlap between the three attitudes for all three groups. ANOVA results show that formant shift ratio had significant effect for every group, but smaller for the AA group ($F(2, 18) = 8.70$, $p = 0.002$) than for the TA group ($F(2, 18) = 22.85$, $p < 0.001$) and the YA group ($F(2, 18) = 15.16$, $p < 0.001$).

The effect of pitch shift on the attitude judgement is illustrated in Figure 7. Overall, the three groups perceived downward pitch shift as more serious sounding, while upward pitch shift as friendlier. ANOVA results show that for the AA group ($F(2, 18) = 11.33$, $p = 0.001$) and the TA group ($F(2, 18) = 8.37$, $p = 0.003$), the effect of pitch shift is smaller than for the YA group ($F(2, 18) = 31.42$, $p < 0.001$).

Figure 8 shows greater overlap between the three attitude judgements based on pitch range for the AA group. ANOVA results show that the effect of this acoustic feature is significant only for the TA group ($F(2, 18) = 5.20$, $p = 0.016$) and the YA group ($F(2, 18) = 5.77$, $p = 0.012$).

Once again, only the YA group showed significant effect of voice quality in the attitude judgement. Yet, they perceived breathy voice as more serious and pressed voice as friendlier.

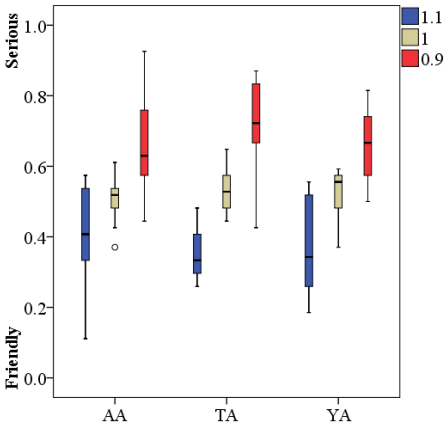

Figure 6: *Formant shift ratio in attitude*

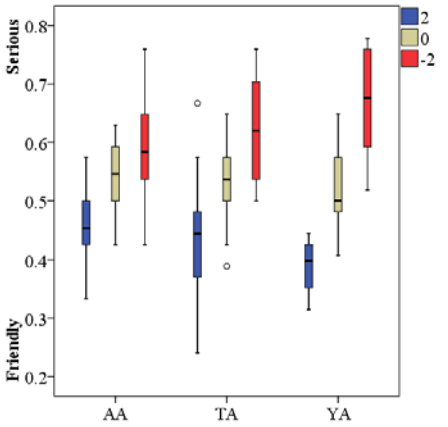

Figure 7: *Pitch shift in attitude*

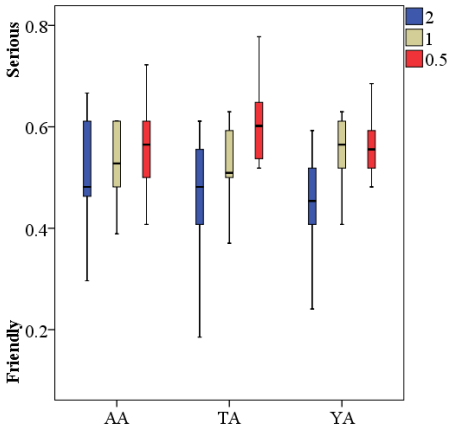

Figure 8: *Pitch range in attitude*

4. DISCUSSION

The present study investigated the perception of body size, emotion and attitude in prosody by high-functioning adolescents with ASD in comparison with typically developing adolescents and young adults. In general, the three groups made similar perceptual judgements, but the sensitivity of adolescents with ASD to the acoustic manipulations was lower than their typically developing peers, who in turn exhibited less sensitivity than young adults. More specifically, adolescents with ASD were less sensitive to manipulated voice quality. They seemed to rely mainly on formant shift ratio and pitch shift, and on pitch range to a less extent.

Interestingly, typically developing adolescents were less sensitive to the acoustic changes than young adults, suggesting that the ability to perceive emotional prosody is still developing in adolescence towards early adulthood, which is in line with the suggestions in [30]. Moreover, if perceptual sensitivity to acoustic changes is related to different stages in prosodic development, then less sensitivity of adolescents with ASD in this study could be interpreted as a delay rather than a deviation in the developmental trajectory. Further research on the perception of emotional prosody by children and adults with ASD may provide more insights.

Although young adults were the only ones who showed sensitivity to voice quality manipulations, their responses were the opposite from previous findings, where pressed voice associated with larger and angrier speakers, and breathy voice with smaller and happier speakers, they perceived breathy voice as from large, angry or serious speakers, and pressed as small, happy or friendly speakers. A possible explanation is that the synthetic utterances in this study were based on a human utterance with clear and careful articulation. Upon listening examination ourselves after the experiments, we noticed that the breathy version sounded effortful. It is possible that this effortfullness may be associated with angrier and/or more serious speakers, which is consistent with the basic BID hypothesis. So the association of careful articulation with anger and seriousness needs to be further explored in future research.

Finally, the findings of this study provide support for the BID-based approach to emotional prosody, as they resonate with previous findings [22], [25]. All the respondents perceived smaller formant dispersion, lower pitch and narrower pitch range as from a large, angry or serious speaker, whereas larger formant dispersion, higher pitch shift and wider pitch range were perceived as from a small, happy or friendly speaker. The overall perception sensitivity of the respondents to these bio-informational dimensions in the directions predicted by the BID theory is a further indication of its effectiveness as a research paradigm.

5. CONCLUSION

High-functioning adolescents with ASD use fewer cues from acoustic features to perceive size, emotion and attitude of speakers through prosody, compared to typically developing peers and young adults. However, young adults can appreciate cues in all acoustic features. These findings lead us to two conclusions: the ability to process emotional prosody may be still in development from adolescence to early adulthood, and prosodic deficits in perception observed in individuals with ASD may be associated to delay rather than deviance in developmental trajectory.

6. ACKNOWLEDGEMENTS

We would like to thank Autism Society Taiwan, R.O.C. and Autism Parent's Association for the help in divulging the research advert, and all the respondents and parents who kindly participated in the experiment. The present study has been approved by University College London Research Ethics Committee and National Taiwan University Hospital Research Ethics Committee.

7. REFERENCES

[1] American Psychiatric Association, "DSM-5 Autism Spectrum Disorder Fact Sheet," Arlington, 2013.

[2] J. McCann, S. Peppé, F. E. Gibbon, A. O'Hare, and M. Rutherford, "Prosody and its relationship to language in school-aged children with high-functioning autism," *Int. J. Lang. Commun. Disord.*, vol. 42, no. 6, pp. 682–702, 2007.

[3] R. Paul, L. D. Shriberg, J. McSweeny, D. Cicchetti, A. Klin, and F. Volkmar, "Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders," *J. Autism Dev. Disord.*, vol. 35, no. 6, pp. 861–869, 2005.

[4] Y. S. Bonneh, Y. Levanon, O. Dean-Pardo, L. Lossos, and Y. Adini, "Abnormal speech spectrum and increased pitch variability in young autistic children," *Front. Hum. Neurosci.*, vol. 4, p. 237, Jan. 2011.

[5] J. J. Diehl, D. Watson, L. Bennetto, J. Mcdonough, and C. Gunlogson, "An acoustic analysis of prosody in high-functioning autism," *Appl. Psycholinguist.*, vol. 30, no. 03, pp. 385–404, May 2009.

[6] K. Hubbard and D. A. Trauner, "Intonation and emotion in autistic spectrum disorders," *J. Psycholinguist. Res.*, vol. 36, no. 2, pp. 159–73, Mar. 2007.

[7] 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.

[8] A. Järvinen-Pasley, S. Peppé, G. King-Smith, and P. Heaton, "The relationship between form and function level receptive prosodic abilities in autism," *J. Autism Dev. Disord.*, vol. 38, no. 7, pp. 1328–40, Aug. 2008.

[9] J. L. Lindner and L. A. Rosén, "Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome," *J. Autism Dev. Disord.*, vol. 36, no. 6, pp. 769–77, Aug. 2006.

[10] R. Paul, A. Augustyn, A. Klin, and F. R. Volkmar, "Perception and production of prosody by speakers with autism spectrum disorders," *J. Autism Dev. Disord.*, vol. 35, no. 2, pp. 205–220, Apr. 2005.

[11] U. Frith and F. Happé, "Theory of mind and self-consciousness: What is it like to be autistic?," *Mind Lang.*, vol. 14, no. 1, pp. 1–22, 1999.

[12] I. A. Apperly, "What is 'theory of mind'? Concepts, cognitive processes and individual differences," *Quartely J. Exp. Psychol.*, vol. 65, no. 5, pp. 825–839, 2012.

[13] S. Holroyd and S. Baron-Cohen, "Brief reports: How far can people with autism go in developing a theory of mind?," *J. Autism Dev. Disord.*, vol. 23, no. 2, pp. 379–385, 1993.

[14] A. A. Spek, E. M. Scholte, and I. A. Van Berckelaer-Onnes, "Theory of mind in adults with HFA and Asperger syndrome," *J. Autism Dev. Disord.*, vol. 40, no. 3, pp. 280–9, Mar. 2010.

[15] O. Golan, S. Baron-Cohen, J. J. Hill, and M. D. Rutherford, "The 'Reading the Mind in the Voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions," *J. Autism Dev. Disord.*, vol. 37, no. 6, pp. 1096–106, Jul. 2007.

[16] M. D. Rutherford, S. Baron-Cohen, and S. Wheelwright, "Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism," *J. Autism Dev. Disord.*, vol. 32, no. 3, pp. 189–94, Jun. 2002.

[17] R. Brennand, A. Schepman, and P. Rodway, "Vocal emotion perception in pseudo-sentences by secondary-school children with Autism Spectrum Disorder," *Res. Autism Spectr. Disord.*, vol. 5, no. 4, pp. 1567–1573, Oct. 2011.

[18] S. Le Sourn-Bissaoui, M. Aguert, P. Girard, C. Chevreuil, and V. Laval, "Emotional speech comprehension in children and adolescents with autism spectrum disorders," *J. Commun. Disord.*, vol. 46, no. 4, pp. 309–20, 2013.

[19] R. B. Grossman, R. H. Bemis, D. P. Skwerer, and H. Tager-Flusberg, "Lexical and affective prosody in children with high-functioning autism," *J. Speech, Lang. Hear. Res.*, vol. 53, pp. 778–793, 2010.

[20] K. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, no. 1–2, pp. 227–256, Apr. 2003.

[21] E. S. Morton, "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *Am. Nat.*, vol. 111, no. 981, pp. 855–69, 1977.

[22] Y. Xu, A. Lee, W.-L. Wu, X. Liu, and P. Birkholz, "Human vocal attractiveness as signaled by body size projection.," *PLoS One*, vol. 8, no. 4, p. e62397, Jan. 2013.

[23] J. J. Ohala, "The frequency code underlies the sound symbolic use of voice pitch," in *Sound Symbolism*, L. Hinton, J. Nichols, and J. J. Ohala, Eds. Cambridge: Cambridge University Press, 1994, pp. 325–347.

[24] L. Noble and Y. Xu, "Friendly speech and happy speech – Are they the same?," in *ICPhS XVII*, 2011, no. August, pp. 1502–1505.

[25] S. Chuenwattanapranithi, Y. Xu, B. Thipakorn, and S. Maneewongvatana, "Encoding emotions in speech with the size code - A perceptual investigation," *Phonetica*, vol. 65, pp. 210–230, 2008.

[26] Y. Xu, A. Kelly, and C. Smillie, "Emotional expressions as communicative signals," in *Prosody and Iconicity*, S. Hancil and D. Hirst, Eds. John Benjamins Publishing Company, 2013, pp. 33–60.

[27] P. Birkholz, B. J. Kröger, and C. Neuschaefer-Rube, "Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis," in *INTERSPEECH-2011*, 2011, pp. 2681–2684.

[28] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer." 2014.

[29] IBM Corporation, "IBM SPSS Statistics for Windows." IBM Corp., Armonk, NY, 2013.

[30] B. Wells, S. Peppé, and N. Goulandris, "Intonation development from five to thirteen," *J. Child Lang.*, vol. 31, pp. 749–778, 2004.