



## OPEN ACCESS

## EDITED BY

Cristina Costescu,  
Babeş-Bolyai University, Romania

## REVIEWED BY

Mohammed Al Jaffal,  
King Saud University, Saudi Arabia  
Mrinmoy Chakrabarty,  
Indraprastha Institute of Information  
Technology Delhi, India

## \*CORRESPONDENCE

Hirokazu Kumazaki  
✉ kumazaki@tiara.ocn.ne.jp

RECEIVED 06 June 2024

ACCEPTED 15 July 2025

PUBLISHED 28 July 2025

## CITATION

Okuizumi D, Terada K, Ishii A, Ohmoto Y,  
Shimizu H, Imamura A, Iwanaga R and  
Kumazaki H (2025) Acoustic features of  
emotional expression in 5-year-old children  
with autism spectrum disorder.  
*Front. Psychiatry* 16:1444675.  
doi: 10.3389/fpsy.2025.1444675

## COPYRIGHT

© 2025 Okuizumi, Terada, Ishii, Ohmoto,  
Shimizu, Imamura, Iwanaga and Kumazaki. This  
is an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Acoustic features of emotional expression in 5-year-old children with autism spectrum disorder

Daichi Okuizumi<sup>1,2</sup>, Kazunori Terada<sup>3</sup>, Azusa Ishii<sup>1</sup>,  
Yoshimasa Ohmoto<sup>4</sup>, Hitomi Shimizu<sup>1</sup>, Akira Imamura<sup>5</sup>,  
Ryoichiro Iwanaga<sup>5</sup> and Hirokazu Kumazaki<sup>1,2\*</sup>

<sup>1</sup>Department of Neuropsychiatry, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan, <sup>2</sup>Hokusuikai Kinen Hospital, Ibaraki, Japan, <sup>3</sup>Faculty of Engineering, Department of Electrical, Electronic, and Computer Engineering, Gifu University, Gifu, Japan, <sup>4</sup>Faculty of Informatics, Department of Behavior Informatics, Shizuoka University, Shizuoka, Japan, <sup>5</sup>Unit of Medical Science, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

**Background:** Children with autism spectrum disorders (ASD) exhibit poor prosodic performance, which is associated with their poor language and social skills. Prosody serves important communicative functions not only at grammatical and pragmatic levels but also at the emotional level. This study investigates the acoustic features of emotional expression in children with ASD compared to typically developing (TD) children, within a narrowly defined age cohort restricted to 5-year-old participants.

**Methods:** Nineteen children with ASD and 19 TD children, aged 5 years, participated in this study. We investigated the differences in the fundamental frequency (f0) ranges in three emotional expression settings (i.e., neutral, liking, and disliking).

**Results:** The f0 range in the neutral setting was greater in children with ASD than in TD children ( $p = 0.04$ ). There were no significant differences in the f0 range between the three settings in the ASD group ( $p = 0.61$ ). There were significant differences between the neutral and liking settings ( $p < 0.01$ ) and the liking and disliking settings ( $p < 0.01$ ) in the TD group. In the ASD group, a negative correlation was observed between the f0 range in the liking setting and the Social Responsiveness Scale, Second Edition T-score ( $p < 0.01$ ).

**Discussion:** By focusing on the relationship between acoustic features and emotional expression setting and by restricting the age of participants, our results demonstrate the trend of acoustic features in children with ASD. To deepen the understanding of the relationship between f0 and emotion, future studies investigating prosody in a range of emotional expression settings are needed.

## KEYWORDS

autism spectrum disorder, prosody, acoustic feature, F0, emotion

# 1 Introduction

Autism spectrum disorder (ASD) is a developmental disability that causes significant social, communication, and behavioral challenges. The Centers for Disease Control and Prevention (CDC) in the US estimates that one in 36 children has ASD (1). Children with ASD may have difficulty developing language skills and understanding what others say (2), which can limit their opportunities in higher education and employment, resulting in an overall negative impact on their quality of life (3). Additionally, they often struggle to communicate nonverbally through hand gestures, eye contact, facial expressions, and prosody (4).

Prosody is concerned with the suprasegmental features of speech and refers to speech rhythm as well as affective, pragmatic, and syntactic communicative functions (5, 6). Prosody operates at various levels, enabling speakers to construct their speech using expressive language. Children with ASD have some prosodic differences such as atypical intonation (a monotone intonation and robot-like voice), incorrect word stress, speech rhythm differences (too slow or too alert), difficulty using a high or low pitch and controlling intensity, poor resonance (nasalization and pharyngeal resonance) and voice quality (7, 8). Poor prosodic performance may lead to poor language skills in children with ASD (9). Prosodic deficits represent some of the most significant barriers to social integration and acceptance (8), leading to impaired social functioning.

Prosody serves important communicative functions not only at grammatical and pragmatic levels but also at the emotional level (10, 11). The fundamental frequency ( $f_0$ ) is defined as the lowest frequency of the periodic waveform.  $F_0$  measures have often been used to identify specific acoustic markers of prosody that differentiate basic emotions (12–15). ASD is associated with impairments in processing one's own and others' emotions (16). Grossman et al. (17) reported that the  $f_0$  range of individuals with ASD was wider than that of typically developing (TD) individuals when expressing emotions related to gladness, fear, anger, and surprise, in a sample of participants aged 8 to 19 years (17). Hubbard and Trauner (18) reported that the  $f_0$  range of individuals with ASD was not significantly different from that of TD individuals in the context of happiness, sadness, and anger, in participants aged 6 to 18 years (18). Hubbard et al. (19) found that the  $f_0$  range of individuals with ASD was wider than that of TD individuals in emotional expression contexts involving happiness, sadness, and anger, in participants aged 18 to 50 years (19). However, they reported that the  $f_0$  range of individuals with ASD was not significantly different from that of TD individuals for neutral topics. Therefore, the results of previous studies investigating the difference in the  $f_0$  range between individuals with ASD and their TD peers according to their emotion are inconsistent. One plausible source of these inconsistencies is the ambiguity of emotion-category labels. The emotion concepts in preschool children are still coarse: between 2 to 5 years of age they tend to group facial expressions mainly by valence and only gradually learn to single out specific categories such as sadness or fear (20). Moreover, constructionist accounts propose that these categories are dynamically assembled from context rather

than fixed universals (21, 22). Mapping affect onto the continuous core-affect (valence–arousal) axes may therefore offer a more stable basis for comparing prosody across studies.

The correlation between prosodic features and age is complex, and interactions between them should be considered. It is well known that prosodic features are significantly correlated with speaker age (23). The prosodic features of school-aged children change with age due to factors such as acquiring accents (8). To deepen understanding of the prosodic features, research targeting children before entering elementary school is needed. In our preliminary study (unpublished), which involved children under 4 years of age, understanding the experiment's explanation proved too difficult, and many participants dropped out. This confounding factor should be minimized by using participants within a narrow age range, restricted to 5 years.

In our preliminary study, we confirmed that 5-year-old children could express emotions of liking and disliking. However, role-playing tasks (i.e., story replay and demonstration tasks) seemed to be difficult for them to complete. By conducting a task that involves showing pictures that provoke the emotions of liking and disliking, it is possible to explore whether emotions can be conveyed through prosody.

McAlpine et al. (24) reported no significant differences in the production of rate, loudness, or pitch between children with ASD and those with TD aged between 24 and 68 months. However, the ASD group exhibited atypical stress patterns significantly more often, such as misplaced stress in multisyllabic words and reduced stress. Yoshimatsu and Umino (25) found that children with ASD scored lower on both prosody comprehension and prosody expression tests compared to typically developing 5-year-old controls. These findings suggest that, among children with ASD around the age of five, prosodic variability and atypicalities are frequently observed.

In this study, we investigated the  $f_0$  of children with ASD compared to their TD peers, restricting the sample to 5-year-olds, across different emotional expression settings (i.e., neutral, liking, and disliking). We predicted that our results would reflect a primary difference in the acoustic features of emotional expression in children with ASD.

## 2 Materials and methods

### 2.1 Participants

This study was approved by Hokuuikai Kinen Hospital Institutional Review Board (No. 081). The legal guardians of the participants provided written informed consent, and the participants provided assent to participate in this study. All procedures involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. After receiving a complete explanation of the study, all participants and their guardians agreed to participate. All participants and their

guardians provided written informed consent. The inclusion criteria for the ASD group were as follows: they had a diagnosis of ASD based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) by a supervising study psychiatrist (26); they were 5 years of age; and their IQ scores were 70 or higher. During enrollment, the diagnoses of all participants were confirmed by a psychiatrist with more than 15 years of experience in ASD using standardized criteria derived from the Diagnostic Interview for Social and Communication Disorders (DISCO), which has demonstrated good psychometric properties (27, 28).

Children with TD were recruited from a public offering. The inclusion criteria for the TD group were: children had to be 5 years of age, have a Social Responsiveness Scale, Second Edition (SRS-2) T-score of 59 or lower, and attend a mainstream preschool with no evidence of intellectual impairment. A total of 19 children with ASD and 19 TD participants were included in this study.

Parents of children in both groups completed the SRS-2 (1) to screen for clinically significant autistic symptoms. Higher scores on the SRS-2 indicate a higher degree of autistic traits. Raw SRS-2 scores were converted to T-scores (with a mean of 50 and a standard deviation of 10) for each sex. We classified the data as TD based on a cutoff value “59” according to the previous study (29).

The participants also completed the Social Communication Questionnaire (SCQ) (30). The SCQ is frequently used as a screening tool in ASD research. It was designed as a questionnaire version of the Autism Diagnostic Interview-Revised (ADI-R; (31)), the gold-standard developmental history measure widely used in research and clinical practice. In this study, we did not set a cutoff based on the SCQ score and only used the SRS-2 to gather children with TD.

Full-scale IQ scores were obtained using the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III).

There were significant differences in the SCQ total scores ( $p < 0.01$ ) and SRS-2 T-scores ( $p < 0.01$ ) between children with ASD and their TD peers. Details of the demographic data are presented in Table 1.

## 2.2 Materials

In this study, we categorized the participants’ speech into three emotional expression settings: neutral, liking, and disliking. These

settings were selected because they occupied distinct regions along the valence-arousal and valence dimensions (32). Utterance in the liking setting evoked high arousal and positive valence, whereas utterance in the disliking setting evoked low arousal and negative valence.

Teaching materials were created using vocabulary acquired before 3 years of age, based on the MacArthur-Bates Communicative Development Inventories (MB-CDIs) (33), which evaluate language development based on parent reports. The teaching materials included actions and pictures of the categories. The pictures were displayed on a tablet. We used 10 action-picture cards (throwing a ball, eating rice, swimming in the pool, walking on the road, waking up in the morning, cutting a tree, sitting on a chair, kicking the ball, drinking water, and putting on clothes). Examples of action pictures are illustrated in Figure 1. We also used 14 categories of picture cards (animals, sea creatures, 4-wheeled vehicles, vehicles, indoor toys, outdoor toys, food, vegetables, fruits, teenagers, colors, outings, occupations, and characters). In each picture card category, six nouns corresponded to one category. An example category image is depicted in Figure 2.

## 2.3 Procedure

Speech recordings in the three emotional expression settings were conducted in the following order: neutral, liking, and then disliking. Examiners of emotional expressions were blinded to the diagnosis group.

In the neutral setting, 10 action-picture cards were displayed one after the other, and participants were instructed to verbally

TABLE 1 Demographic data of participants.

	ASD ( $n = 19$ )	TD ( $n = 19$ )	
	Mean (SD)	Mean (SD)	$p$
Age (months)	64.79 (3.81)	66.95 (2.50)	0.06
Sex ratio (M: F)	13:6	13:6	
SCQ total score	10.16 (4.60)	2.21 (1.51)	<0.01**
SRS-2 T-scores	71.79 (11.21)	44.42 (5.22)	<0.01**
IQ	83.79 (9.10)	N/A	N/A

SD, standard deviation; ASD, autism spectrum disorder; TD, typical development; SCQ, Social Communication Questionnaire; SRS-2, Social Responsiveness Scale-Second edition.

\*\* $p < 0.01$ .



FIGURE 1  
Example of action pictures.

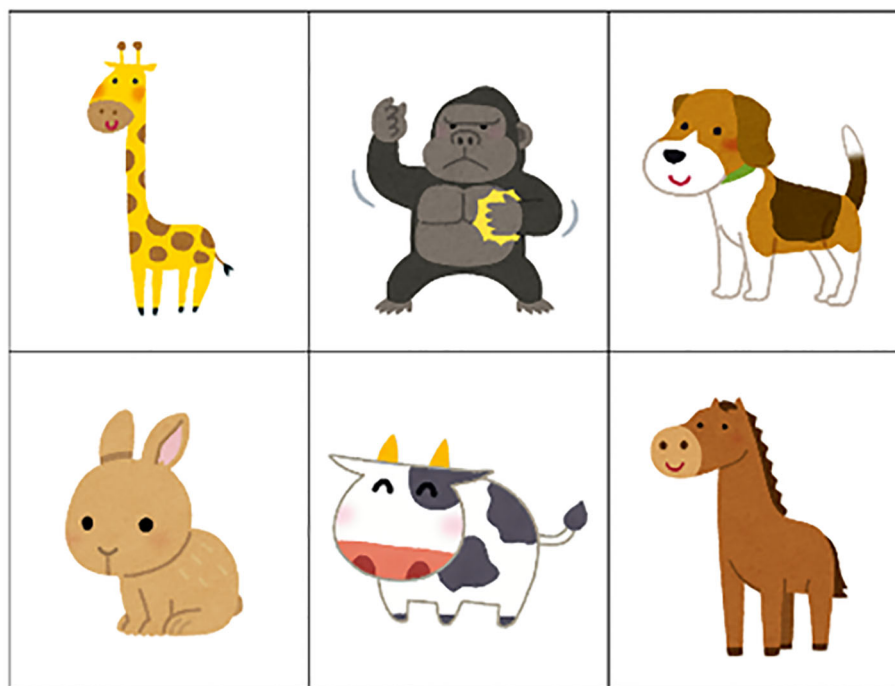


FIGURE 2  
Example of category pictures.

name each action (e.g., “throw a ball”). After displaying all 10 action-picture cards, participants took a 1-min break before repeating the exercise. Each picture card was displayed, and the examiners asked the participants to name the object shown. Since the examiner did not utter the name of the object, the participants were unable to imitate their utterances directly. If the participants gave an incorrect response or did not answer, the data were removed from the analysis, and the examiners displayed the next action-picture card.

In the liking and disliking settings, 14 category picture cards were displayed individually. Participants were encouraged to select a card and express their liking as “I like giraffes.” Subsequently, they were encouraged to select another card and express their dislike as “I don’t like gorillas.” If the participants provided an incorrect response or did not answer, the examiners displayed the next picture card. The participants checked their expressed emotions using a 5-point Likert scale after completing all the sentence tests. The recording equipment consisted of a microphone (the Shure MV7) and an audio capture system (the TASCAM PortaCapture X8). The microphone was maintained at a constant distance from the participants during recording. The recording parameters included a 48-kHz sampling frequency and 24-bit quantization.

## 2.4 Criteria for utterance selection

To obtain relatively comparable speech samples from participants’ utterances, only complete utterances containing both a noun and a verb were included in the analysis, provided they did

not meet any of the exclusion criteria. Utterances were excluded if they were questions, fillers, repetitive speech, consisted only of a noun or a verb, were interrupted by the examiner, were unintelligible, were directed toward someone else in the room and unrelated to the task, or were abandoned. Only complete utterances containing more than two words were considered. In the neutral setting, action utterances (e.g., “throw a ball”) were selected, whereas in the liking and disliking settings, preference utterances (e.g., “I like giraffes”) were chosen. These exclusion criteria were applied to ensure consistency in utterance length and type across participants.

## 2.5 Measurement

The audio data for analysis consisted of utterances in three settings. Using sound editing software (free software, Praat version 6.2.14), the sampling frequency of the recorded data was converted to 24 kHz, and the audio data were isolated for each sentence. We checked all audio data for artifacts that might interfere with accurate pitch analysis (e.g., background noise, coughs, or artifacts from glottal stops) and removed them. For each sentence, we extracted  $f_0$  (in Hz) with a timestep of 0.01 seconds. The  $f_0$  range, which measures the extent to which an individual’s pitch varies during speech, was calculated by subtracting the minimum  $f_0$  value from the maximum  $f_0$  value obtained from each sentence. First, the  $f_0$  range of each sentence was calculated for each emotional expression setting. The  $f_0$  range in each setting was subsequently averaged across all sentences.



## 2.6 Data analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 24.0 (IBM Corp., Armonk, NY, USA). Differences in the number of speeches between children with ASD and TD were analyzed using Mann–Whitney *U* tests. Kruskal–Wallis tests were used to determine if differences in the f0 range existed in each emotional expression setting (i.e., neutral, liking, disliking) between children with ASD and TD. The Friedman test was used to determine the f0 range in each setting in children with ASD and TD. Spearman's rank correlation coefficients were used to explore the relationships between the f0 range and the SRS-2 T-score and between the f0 range and their IQ in each setting in children with ASD.

## 3 Results

No significant difference was observed between the number of sentences in children with ASD and TD (neutral setting,  $U = 121.50$ ,  $p = 0.09$ ; liking setting,  $U = 131.00$ ,  $p = 0.15$ ; disliking setting,  $U = 142.00$ ,  $p = 0.27$ ). The details are presented in Table 2.

The f0 range in the neutral setting was greater in children with ASD than TD children ( $H = 4.24$ ,  $df = 1$ ,  $p = 0.04$ ). There were no significant differences in the f0 range across groups for the liking setting ( $H = 0.79$ ,  $df = 1$ ,  $p = 0.37$ ) or disliking setting ( $H = 1.26$ ,  $df = 1$ ,  $p = 0.26$ ) between children with ASD and TD children.

There were no significant differences in the f0 range between the three settings in the ASD group ( $\chi^2(2) = 1.00$ ,  $p = 0.61$ ). Bonferroni-corrected *post-hoc* tests revealed no significant differences between the neutral and liking settings ( $p = 0.31$ ), the neutral and disliking settings ( $p = 0.62$ ), or the liking and disliking settings ( $p = 0.62$ ) in the ASD group. There were significant differences in the f0 range between the three settings in the TD group ( $\chi^2(2) = 13.37$ ,  $p < 0.01$ ). Bonferroni-corrected *post-hoc* tests revealed that the liking setting had a greater f0 range than the neutral context ( $p < 0.01$ ) and the disliking setting ( $p = 0.01$ ) in the TD group. There were no significant differences between the neutral and disliking settings in the TD group ( $p = 0.87$ ). The details are illustrated in Figure 3.

In children with ASD, a negative correlation was observed between the f0 range in the liking setting and the SRS-2 T-score ( $r = -0.60$ ,  $p < 0.01$ ). There were no significant differences in the

correlations between the SRS-2 T-score and f0 range in the neutral ( $r = -0.17$ ,  $p = 0.49$ ) and disliking settings ( $r = -0.38$ ,  $p = 0.12$ ). The details are presented in Table 3. There were no significant differences between IQ and f0 in the neutral ( $r = 0.11$ ,  $p = 0.67$ ), liking ( $r = 0.18$ ,  $p = 0.46$ ), or disliking ( $r = 0.40$ ,  $p = 0.09$ ) settings. The details are presented in Table 3.

## 4 Discussion

This study investigated the f0 of children with ASD compared to their TD peers, restricting the age range of participants to 5 years, across different emotional expression settings (i.e., neutral, liking, and disliking) within a narrow age range restricted to 5 years. The results revealed that the f0 range in the neutral setting was greater in children with ASD than in TD children. This suggests that in emotionally neutral situations, children with ASD may show greater pitch variation compared to TD children. There were no significant differences in the f0 range between the three settings (i.e., neutral, liking, disliking) in the ASD group, whereas significant differences in the f0 range were observed between the three settings in the TD group. These findings suggest that children with ASD have difficulty varying pitch according to their emotions, whereas children with TD emphasize positive emotions in their speech. In children with ASD, a negative correlation was observed between the f0 range in the liking setting and the SRS-2 T-score, suggesting that increased severity of autistic traits is associated with reduced pitch variation. Our results indicate a trend in the acoustic features of children with ASD.

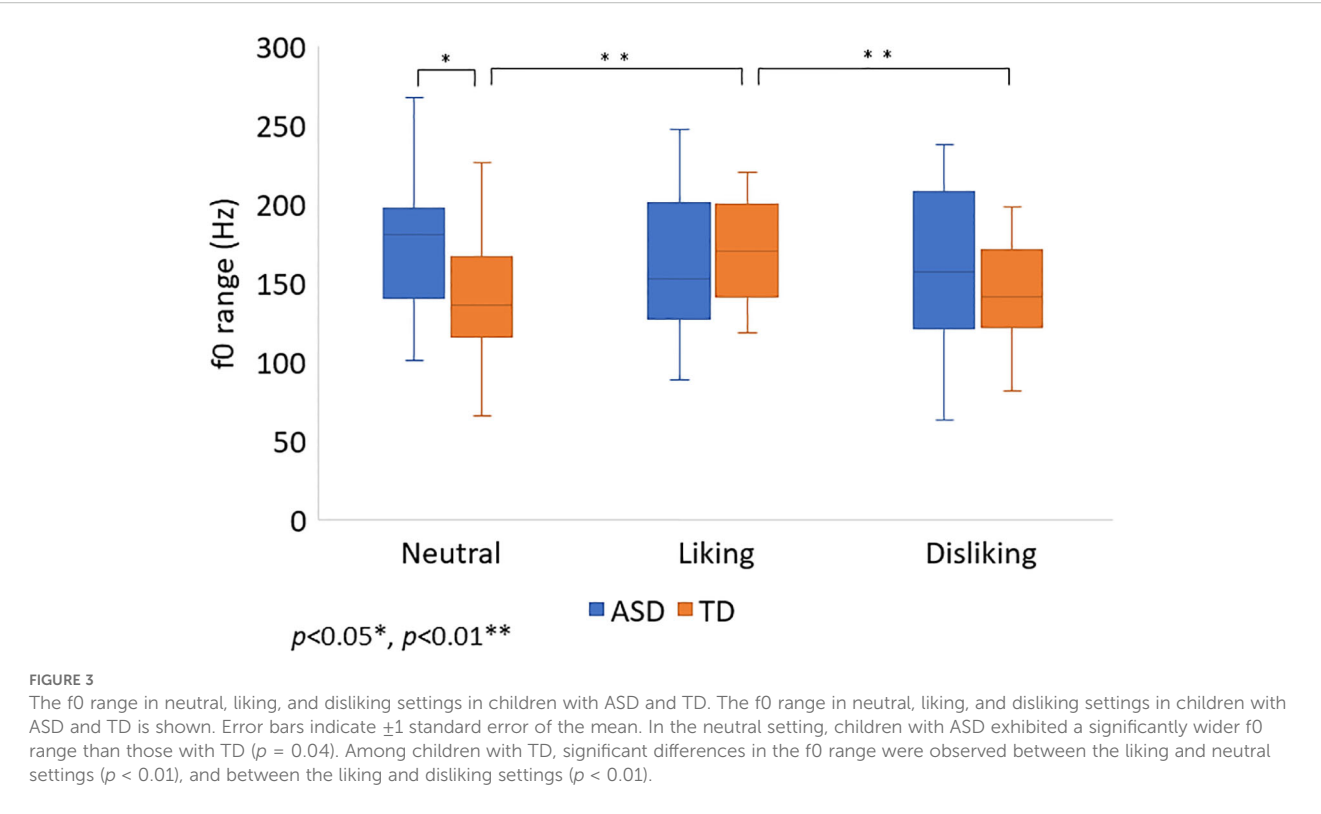
Previous studies (34–37) have reported that the f0 range in general conversations in individuals with ASD was higher than that in TD individuals, which is consistent with the results of this study, that the f0 range in the neutral setting in children with ASD was greater than that in TD children. No significant differences were observed in the f0 range between the ASD and TD groups in the liking and disliking settings. This finding is consistent with those of Hubbard and Trauner (18), who reported no significant differences in the f0 range between ASD and TD groups aged 6 to 21 years during emotional expression. Conversely, Hubbard et al. (19) reported significant differences in the f0 range between ASD and TD groups aged 21 to 41 years during emotional expression. According to a previous study by Lee et al. (38), f0 variability usually decreases with age, beginning around 10 years of age. A previous meta-analysis (7) suggested that pitch difference between individuals with ASD and those with TD were significant during adulthood compared to other age groups, which may explain the discrepancies between the findings of Hubbard and Trauner (18) and those reported by Hubbard et al. (19). Given these factors, the finding of the present study—that there were no significant differences in the f0 range between the ASD and TD groups in both the liking and disliking settings—is understandable.

Contrary to their TD peers, children with ASD are unable to change pitch in areas where it is usually emphasized and are unable to adjust pitch depending on the communication situation (39).

TABLE 2 The number of sentences in children with ASD and TD children in Neutral, Liking, and Disliking settings.

	ASD ( $n = 19$ )	TD ( $n = 19$ )	
	Mean (SD)	Mean (SD)	$p$
Neutral setting	7.68 (1.80)	8.63 (1.42)	0.09
Liking setting	11.79 (2.52)	13.05 (1.08)	0.15
Disliking setting	7.32 (4.10)	8.84 (3.50)	0.27

ASD, autism spectrum disorder; TD, typical development; SD, standard deviation.



These results may explain why children with TD emphasize positive emotions in their speech, whereas children with ASD have difficulty varying pitch according to their emotions.

Nakai et al. (40) reported that pitch variation in word utterances was negatively correlated with the severity of autistic traits (40). A previous study also found that pitch measures extracted from the ADOS-2 conversational task were significantly negatively correlated with the SRS-2 T-score (41).

By focusing on the relationship between acoustic features and emotional expression settings, and by restricting the age range of participants, our study found a negative correlation between the f0 range in the liking condition and autistic traits, consistent with previous findings (8, 40). Given that the f0 range is associated with emotional expression (12–15) and that emotional expression in favorable settings is linked to social functioning (42), the f0 range in emotional contexts may reflect social dysfunction in children with ASD.

When considered in the context of previous studies (8, 40), our findings support the notion that greater severity of autistic traits is

associated with reduced f0 range, even in positive emotional contexts. These results underscore the potential future applications of (i) developing voice-based biomarkers utilizing the f0 range, and (ii) implementing interventions targeting emotional prosody.

Our findings—that children with ASD have difficulty modulating pitch according to emotional context, particularly in the lower f0 range of the liking condition, which is associated with greater autistic traits—can be interpreted within the framework of neurodevelopmental theories of emotional processing and prosody.

Emotional prosody processing is typically associated with right-hemisphere brain regions, including the right superior temporal gyrus and inferior frontal gyrus (43, 44). Atypical development or reduced activation in these regions has been reported in individuals with ASD (45), which may underlie difficulties in modulating vocal pitch to match emotional valence.

Furthermore, theory of mind (ToM)—the ability to infer others’ mental and emotional states—is also implicated in prosodic expression (46). If children with ASD have reduced ToM abilities, they may not only struggle to interpret others’ emotional prosody but may also have difficulty expressing their own emotions vocally in socially appropriate ways.

These findings suggest that a reduced f0 range in children with ASD, even in positive emotional contexts, reflects underlying neurodevelopmental mechanisms affecting both emotional processing and social communication. This highlights the importance of incorporating such theoretical frameworks when developing voice-based biomarkers and interventions targeting emotional prosody.

TABLE 3 The correlations between SRS-2 T-scores and IQ, and f0 range of neutral, liking, and disliking settings.

Item	SRS-2 T-scores	IQ
Neutral setting	-0.17	-0.11
Liking setting	-0.60**	0.18
Disliking setting	-0.38	0.40

SRS-2, Social Responsiveness Scale.  
\*\**p* < 0.01.

## 4.1 Limitations

This study has several limitations. First, the sample size was relatively small, and most participants were male. Wehrle et al. (47) described the prosody of men with ASD as more exaggerated than that of women. Furthermore, parents have reported that boys with ASD are more likely to speak with an “unusual tone of voice” than girls (48). Therefore, gender differences in prosody may exist among children with ASD. Second, we did not conduct formal IQ testing in children with TD and instead relied only on the records of a normal preschool performance for those enrolled in this study. However, all the children attended mainstream preschools with no evidence of intellectual impairment. We confirmed that 5-year-old children with average intellectual and verbal competencies could complete the experimental process in our preliminary experiments. These results also demonstrated that IQ was not correlated with the  $f_0$  range in children with ASD. Previous longitudinal research (49) suggests that age-appropriate performance in preschool reliably predicts later cognitive functioning, including IQ. Therefore, assuming average IQ in TD children based on typical preschool performance may be justified, although we acknowledge the limitations of inferring cognitive level without formal testing. This limitation should be considered when interpreting between-group comparisons, as individual differences in cognitive ability within the TD group may not be fully accounted for in the present study. Third, since the participants were 5 years old, it was difficult to perform an experiment involving complex procedures. This study focused on the neutral, liking, and disliking settings. It is generally accepted that there are various emotions such as happiness, sadness, anger, disgust, fear, and surprise.

## 4.2 Future research directions

To address these limitations and promote advanced understanding, we highlight the following: First, studies with larger sample sizes, including numerous female participants are required to validate the results. Second, further research assessing IQ in TD children is needed. Third, further studies are needed to investigate  $f_0$  in a variety of emotional expression settings in order to advance our understanding of the relationship between  $f_0$  and emotion.

## 5 Conclusion

This study identified the acoustic features according to each emotional expression setting in children with ASD compared to TD children. Given the relationship between prosody, language skills, and social dysfunction, focusing on prosodic features screening children whose social function is poor, the significance of focusing on prosody is great. Further studies are needed to investigate  $f_0$  in a variety of emotional expression settings, in order to deepen our understanding of the relationship between  $f_0$  and emotion.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Hokusai Kinen Hospital Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

DO: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. KT: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. Als: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. YO: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. HS: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. AIm: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. RI: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. HK: Formal Analysis, Investigation, Methodology, Writing – original draft, Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported in part by JST, CREST (grant number JPMJCR21D4), Japan.

## Acknowledgments

We thank all the children and their families who participated in this study. Moreover, we thank all the team members of Nagasaki University for their incredible work on this project.

## Conflict of interest

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Cakir J, Frye RE, Walker SJ. The lifetime social cost of autism: 1990–2029. *Res Autism Spec Disord.* (2020) 72:101502. doi: 10.1016/j.rasd.2019.101502
- Vogindroukas I, Stankova M, Chelas EN, Proedrou A. Language and speech characteristics in autism. *Neuropsychiatr Dis Treat.* (2022) 18:2367–77. doi: 10.2147/NDT.S331987
- Al Jaffar M. Video-based interventions for adolescents and young adults with autism spectrum disorder: A systematic review. *Int J Ment Health Promot.* (2023) 25:881–90. doi: 10.32604/ijmhp.2023.028982
- Strickland DC, Coles CD, Southern LB. A transition to employment program for individuals with autism spectrum disorders. *J Autism Dev Disord.* (2013) 43:2472–83. doi: 10.1007/s10803-013-1800-4
- Belyk M, Brown S. Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Soc Cognit Affect Neurosci.* (2014) 9:1395–403. doi: 10.1093/scan/nst124
- McCann J, Peppé S, Gibbon FE, O'Hare A, Rutherford M. Prosody and its relationship to language in school-aged children with high-functioning autism. *Int J Lang Commun Disord.* (2007) 42:682–702. doi: 10.1080/13682820601170102
- Asghari SZ, Farashi S, Bashirian S, Jenabi E. Distinctive prosodic features of people with autism spectrum disorder: A systematic review and meta-analysis study. *Sci Rep.* (2021) 11:23093. doi: 10.1038/s41598-021-02487-6
- Eigsti IM, de Marchena AB, Schuh JM, Kelley E. Language acquisition in autism spectrum disorders: A developmental review. *Res Autism Spec Disord.* (2011) 5:681–91. doi: 10.1016/j.rasd.2010.09.001
- Peppé S, McCann J, Gibbon F, O'Hare A, Rutherford M. Receptive and expressive prosodic ability in children with high-functioning autism. *J Speech Lang Hear Res.* (2007) 50:1015–28. doi: 10.1044/1092-4388(2007)071
- Ma W, Thompson WF. Human emotions track changes in the acoustic environment. *Proc Natl Acad Sci U.S.A.* (2015) 112:14563–8. doi: 10.1073/pnas.1515087112
- Paul R, Augustyn A, Klin A, Volkmar FR. Perception and production of prosody by speakers with autism spectrum disorders. *J Autism Dev Disord.* (2005) 35:205–20. doi: 10.1007/s10803-004-1999-1
- Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol.* (1996) 70:614–36. doi: 10.1037//0022-3514.70.3.614
- Justin PN, Laukka P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol Bull.* (2003) 129:770–814. doi: 10.1037/0033-2909.129.5.770
- Pell MD, Paulmann S, Dara C, Alasseri A, Kotz SA. Factors in the recognition of vocally expressed emotions: A comparison of four languages. *J Phon.* (2009) 37:417–35. doi: 10.1016/j.wocn.2009.07.005
- Thompson WF, Balkwill LL. Decoding speech prosody in five languages. *Semiotica.* (2006) 2006:407–24. doi: 10.1515/SEM.2006.017
- Kanner L. Autistic disturbances of affective contact. *Nerv Child.* (1943) 2:217–50.
- Grossman RB, Edelson LR, Tager-Flusberg H. Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. *J Speech Lang Hear Res.* (2013) 56:1035–44. doi: 10.1044/1092-4388(2012)12-0067
- Hubbard K, Trauner DA. Intonation and emotion in autistic spectrum disorders. *J Psycholinguist Res.* (2007) 36:159–73. doi: 10.1007/s10936-006-9037-4
- Hubbard DJ, Faso DJ, Assmann PF, Sasson NJ. Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Res.* (2017) 10:1991–2001. doi: 10.1002/aur.1847
- Widen SC, Russell JA. Children acquire emotion categories gradually. *Cognit Dev.* (2008) 23:291–312. doi: 10.1016/j.cogdev.2008.01.002
- Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J Pers Soc Psychol.* (1999) 76:805–19. doi: 10.1037//0022-3514.76.5.805
- Barrett LF. Solving the emotion paradox: Categorization and the experience of emotion. *Pers Soc Psychol Rev.* (2006) 10:20–46. doi: 10.1207/s15327957pspr1001\_2
- Müller C. *Speaker Classification I: Fundamentals, Features, and Methods.* Berlin Heidelberg: Springer (2007). doi: 10.1007/978-3-540-74200-5
- McAlpine A, Plexico LW, Plumb AM, Cleary J. Prosody in young verbal children with autism spectrum disorder. *Contemp Issues Commun Sci Disord.* (2014) 41:120–32. doi: 10.1044/cicsd\_41\_S\_120
- Yoshimatsu Y, Umino A. Characteristics of the understanding and expression of emotional prosody among children with autism spectrum disorder. *Autism Open Access.* (2016) 6. doi: 10.4172/2165-7890.1000185
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. 5th ed.* United States: American Psychiatric Association Publishing (2013). doi: 10.1176/appi.books.9780890425596
- Leekam SR, Libby SJ, Wing L, Gould J, Taylor C. The Diagnostic Interview for Social and Communication Disorders: Algorithms for ICD-10 childhood autism and Wing and Gould autistic spectrum disorder. *J Child Psychol Psychiatry.* (2002) 43:327–42. doi: 10.1111/1469-7610.00024
- Wing L, Leekam SR, Libby SJ, Gould J, Larcombe M. The Diagnostic Interview for Social and Communication Disorders: Background, inter-rater reliability and clinical use. *J Child Psychol Psychiatry.* (2002) 43:307–25. doi: 10.1111/1469-7610.00023
- Constantino JN, Gruber CP. *Social Responsiveness Scale. 2nd ed.* Kamio Y, editor. Tokyo: Nihon Bunka Kagakusha (Original Work Published 2012), Trans (2017).
- Rutter M, Bailey A, Lord C. The social communication questionnaire. *Western psychol Serv.* (2003).
- Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord Rev version.* (1994) 24:659–85. doi: 10.1007/BF02172145
- Russell JA. Core affect and the psychological construction of emotion. *Psychol Rev.* (2003) 110:145–72. doi: 10.1037/0033-295X.110.1.145
- Fernald A, Morikawa H. Common themes and cultural variations in Japanese and American Mothers Speech to infants. *Child Dev.* (1993) 64:637–56. doi: 10.2307/1131208
- Eni M, Dinstein I, Ilan M, Menashe I, Meiri G, Zigel Y. Estimating autism severity in young children from speech signals using a deep neural network. *IEEE Access.* (2020) 8:139489–500. doi: 10.1109/ACCESS.2020.3012532
- Godel M, Robain F, Journal F, Kojovic N, Latrèche K, Dehaene-Lambertz G, et al. Prosodic signatures of ASD severity and developmental delay in preschoolers. *NPJ Digit Med.* (2023) 6:99. doi: 10.1038/s41746-023-00845-4
- Nadig A, Shaw H. Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *J Autism Dev Disord.* (2012) 42:499–511. doi: 10.1007/s10803-011-1264-3
- Sharda M, Subhadra TP, Sahay S, Nagaraja C, Singh L, Mishra R, et al. Sounds of melody-Pitch patterns of speech in autism. *Neurosci Lett.* (2010) 478:42–5. doi: 10.1016/j.neulet.2010.04.066
- Lee S, Potamianos A, Narayanan S. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J Acoust Soc Am.* (1999) 105:1455–68. doi: 10.1121/1.426686
- DePape A-MR, Chen A, Hall GBC, Trainor LJ. Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Front Psychol.* (2012) 3:72. doi: 10.3389/fpsyg.2012.00072
- Nakai Y, Takashima R, Takiguchi T, Takada S. Speech intonation in children with autism spectrum disorder. *Brain Dev.* (2014) 36:516–22. doi: 10.1016/j.braindev.2013.07.006
- Asgari M, Chen L, Fombonne E. Quantifying voice characteristics for detecting autism. *Front Psychol.* (2021) 12:665096. doi: 10.3389/fpsyg.2021.665096
- Van Kleef GA, Homan AC, Beersma B, Van Knippenberg D, Van Knippenberg B, Damen F. Searing sentiment or cold calculation? the effects of leader emotional displays on team performance depend on follower epistemic motivation. *Acad Manag J.* (2009) 52:562–80. doi: 10.5465/AMJ.2009.4131253
- Ross ED. The Aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Arch Neurol.* (1981) 38:561–9. doi: 10.1001/archneur.1981.00510090055006
- Kirk E, Sharma S. Mind-mindedness in mothers of children with autism spectrum disorder. *Res Autism Spec Disord.* (2017) 43–44:18–26. doi: 10.1016/j.rasd.2017.08.005
- Wang AT, Lee SS, Sigman M, Dapretto M. Reading affect in the face and voice: Neural correlates of interpreting communicative intent in children and adolescents with autism spectrum disorders. *Arch Gen Psychiatry.* (2007) 64:698–708. doi: 10.1001/archpsyc.64.6.698



46. Golan O, Baron-Cohen S, Hill JJ, Rutherford MD. 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.* (2007) 37:1096–106. doi: 10.1007/s10803-006-0252-5
47. Wehrle S, Cangemi F, Hanekamp H, Vogeley K, Grice M. Assessing the intonation style of speakers with autism spectrum disorder. *Speech Prosody.* (2020) 2020:809–13. doi: 10.21437/SpeechProsody.2020
48. de Giambattista C, Ventura P, Trerotoli P, Margari F, Margari L. Sex differences in autism spectrum disorder: Focus on high functioning children and adolescents. *Front Psychiatry.* (2021) 12:539835. doi: 10.3389/fpsy.2021.539835
49. Józsa K, Amukune S, Zentai G, Barrett KC. School readiness test and intelligence in preschool as predictors of middle school success: Result of an eight-year longitudinal study. *J Intell.* (2022) 10:66. doi: 10.3390/jintelligence10030066