Journal of Health, Wellness and Community Research
ISSN: 3007-0570

OPEN ACCESS

JHWCR

Type: Original Article
Published: 24 September 2025
Volume: III, **Issue:** XIII
DOI: https://doi.org/10.61919/ean25f93

Correspondence
✉ nawazsaffa@gmail.com

Received 25-08-25 **Accepted** 22-09-2025

Authors' Contributions
Concept: MSS, SN; Design: SN, SA; Data Collection: MSS, SA, AA; Analysis: SN, MSS; Drafting: SN, MSS, SA, AA

Declarations
No funding was received for this study. The authors declare no conflict of interest. The study received ethical approval. All participants provided informed consent.

"Click to Cite"

Comparison of Acoustic and Perceptual Analysis of Voice in Autistic and Normal Children

Maheen Sohail Sadiq[1], Saffa Nawaz[12], Saba Abrar[3], Ayesha Amin[3]

1 Speech and Language Pathologist, The University of Lahore, Lahore, Pakistan
2 Lecturer, Department of Rehabilitation Sciences, Faculty of Allied Health Sciences, The University of Lahore, Lahore, Pakistan
3 Speech and Language Pathologist, Lahore, Pakistan

ABSTRACT

Background: *Autism Spectrum Disorder (ASD) is associated with persistent deficits in communication and social reciprocity, with prosodic and vocal atypicalities frequently reported as clinically relevant features. Prior research has shown inconsistent findings across acoustic and perceptual domains, and there remains a need to clarify the contribution of resonance-based acoustic markers and perceptual voice characteristics in children with ASD.* ***Objective:*** *This study aimed to compare acoustic parameters and perceptual voice features between children with ASD and typically developing peers, with a specific focus on formant frequencies and perceptual ratings as potential discriminators of atypical voice.* ***Methods:*** *An analytical cross-sectional design was employed, enrolling 24 children aged 5–13 years, equally divided between ASD and typically developing groups. Data were collected using Speech Analyzer software and the GRBAS scale. Acoustic variables included duration, intensity, fundamental frequency, and formant frequencies (F1–F3). Perceptual ratings assessed grade, roughness, breathiness, asthenia, and strain, as well as supplementary speech features. Statistical analysis was performed using independent-samples t-tests with effect sizes.* ***Results:*** *No significant group differences were observed in duration, intensity, or fundamental frequency. However, F1 ($p = 0.040$, $d = 0.89$) and F2 ($p = 0.005$, $d = 1.27$) were significantly lower in the ASD group, with F3 trending toward significance ($p = 0.062$). Perceptual analysis revealed higher strain in ASD children ($p = 0.005$, $d = 0.89$). Additional deficits were observed in clarity, naturalness, pitch modulation, and fluency (all $p < 0.001$, $d > 1.2$), alongside increased speech pauses ($p = 0.009$).* ***Conclusion:*** *Children with ASD exhibit preserved global acoustic features but distinct resonance-based and perceptual voice abnormalities, particularly reduced formant frequencies and elevated strain, which may serve as sensitive clinical markers. Integration of acoustic–perceptual assessments in healthcare can enhance early detection and guide tailored interventions to improve communication outcomes.*

Keywords
Autism Spectrum Disorder, Voice Analysis, Acoustic Features, Perceptual Voice Assessment, Formant Frequencies, Speech Fluency

INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by persistent difficulties in social communication, restricted behaviors, and atypical interactional patterns, with global prevalence estimates ranging between 0.7% and 1.5% of children depending on diagnostic criteria and study setting (1). Although the behavioral and cognitive manifestations of ASD are well described, communication impairments, particularly in prosody and vocal expression, remain among the most salient and socially disabling features. Children with ASD frequently exhibit abnormalities in pitch, intonation, rhythm, and loudness that may lead to flat, monotonous, or exaggerated speech and limit intelligibility, naturalness, and pragmatic effectiveness (2). Such deviations can interfere with emotional expression, turn-taking, and conversational reciprocity, thereby amplifying the barriers to social integration commonly observed in this population (3).

Acoustic research has consistently highlighted atypicalities in fundamental frequency, intensity, and prosodic variability among children with ASD. Studies indicate that over half of affected individuals demonstrate measurable deviations in vocal features such as elevated mean pitch, restricted modulation, and inconsistent loudness control, reflecting disrupted neuromotor coordination between respiratory, laryngeal, and articulatory subsystems (4,5). Perceptual studies using clinical voice rating tools such as the GRBAS scale complement these findings by documenting impressions of strained, rough, or breathy vocal quality in ASD speakers compared to typically developing peers (6). Importantly, while parent- or clinician-reported assessments capture overt abnormalities, they may fail to detect subtle but systematic differences in resonance, formant distribution, or articulatory precision that are quantifiable only through acoustic analysis (7).

Despite the clinical and functional importance of prosodic and voice characteristics, prior investigations reveal mixed outcomes. Some studies report significantly higher pitch and increased jitter in autistic speech (8), whereas others find minimal or no differences in global acoustic parameters but identify consistent abnormalities in spectral measures such as formant frequencies (9). This inconsistency may reflect heterogeneity in age ranges, ASD severity, comorbidities, and methodological tools across studies. Moreover, most available data are derived from Western

populations and clinical recruitment pathways, with limited evidence from diverse educational and cultural contexts, creating gaps in generalizability (10).

The need to integrate acoustic and perceptual approaches is particularly critical, as each provides complementary insights. Acoustic analysis offers objective quantification of pitch, intensity, duration, and resonance, while perceptual analysis reflects the lived communicative experience, capturing how speech is received and judged by listeners in real contexts (11). By jointly applying these methods, researchers can better identify distinctive markers of autistic voice and establish whether certain acoustic anomalies translate into perceptible communicative deficits.

Given these considerations, the present study aimed to compare the acoustic and perceptual voice characteristics of children with ASD and typically developing children within the Pakistani context. Specifically, it sought to determine whether differences exist in fundamental frequency, intensity, duration, and formant frequencies, alongside perceptual ratings of vocal quality using the GRBAS scale. The objective was to clarify whether children with ASD present with distinctive acoustic deviations that align with or diverge from listener-based perceptual impressions, thereby addressing the gap in integrated voice analysis in non-Western pediatric ASD populations. Objective: To identify and compare acoustic parameters and perceptual voice features between children with Autism Spectrum Disorder and typically developing children, with the hypothesis that children with ASD would demonstrate significant differences in resonance-related formant frequencies and perceptual strain relative to their peers.

MATERIAL AND METHODS

This study employed an analytical cross-sectional design to evaluate and compare acoustic and perceptual voice features in children with Autism Spectrum Disorder and their typically developing peers. The design was chosen to allow simultaneous assessment of exposure and outcome variables in both groups, thereby enabling direct group comparisons while minimizing temporal and recall biases inherent in retrospective approaches (12). The research was conducted in Lahore, Pakistan, between January and June 2025, in collaboration with specialized educational centers for children with developmental disorders and mainstream schools in the same region. This dual recruitment ensured representation of both populations within the same sociolinguistic environment, reducing contextual confounding related to language and cultural background.

Participants were eligible if they were between 5 and 13 years of age and demonstrated either a formal diagnosis of mild Autism Spectrum Disorder according to DSM-5 criteria or were typically developing children without neurological, psychiatric, or speech-language disorders. Children with comorbidities such as cerebral palsy, hearing impairment, intellectual disability, or genetic syndromes were excluded to avoid confounding effects of additional impairments on speech production. Recruitment was conducted through purposive sampling in schools, with collaboration from teachers and clinicians who identified potential participants. Parents were approached, and written informed consent was obtained prior to enrollment, while verbal assent was secured from children when age appropriate.

Data collection followed standardized procedures to ensure consistency and reproducibility. A structured demographic proforma captured age, gender, language spoken at home, history of speech therapy, and parental concerns about speech and voice. Acoustic parameters were measured using Speech Analyzer software, which provided quantitative measures of duration, intensity, fundamental frequency, and formant frequencies (F1, F2, F3) across sustained vowel samples. Perceptual analysis was performed using the GRBAS scale (Grade, Roughness, Breathiness, Asthenia, Strain), rated independently by two trained speech-language pathologists blinded to group assignment, with consensus scores used for final analysis. All recordings were collected in a quiet clinical environment using the same calibrated microphone to minimize background noise and equipment variability.

Operational definitions were applied consistently: fundamental frequency represented average pitch across the sustained vowel, intensity reflected mean sound pressure level, and formant frequencies were defined as spectral energy peaks corresponding to articulatory resonance. GRBAS ratings were treated as ordinal variables but analyzed for group mean differences. To address potential bias, evaluators underwent calibration sessions to establish inter-rater reliability, and blinding was maintained to reduce observer bias. Confounding by age and gender was assessed through descriptive stratification, while recruitment from the same sociolinguistic setting minimized environmental confounding.

The sample size of 24 participants (12 ASD, 12 typically developing) was pragmatically determined based on feasibility and availability within the recruitment window. A priori power analysis indicated that with alpha set at 0.05, the study would have 80% power to detect medium-to-large effect sizes (Cohen's $d \geq 0.8$) in acoustic outcomes, though smaller effects may remain undetected. Statistical analysis was performed using SPSS version 25.0. Continuous variables were summarized using means and standard deviations, and independent samples t-tests were applied for group comparisons under the assumption of normally distributed data, supported by Shapiro-Wilk tests. Levene's test guided equal-variance assumptions. For variables violating normality assumptions, non-parametric sensitivity analyses were conducted using Mann–Whitney U tests. Missing data were minimal; when present, pairwise deletion was used. Exploratory subgroup analyses were performed by gender to assess potential differential effects, though sample sizes limited statistical power.

Ethical approval was granted by the Institutional Review Board of The University of Lahore, and all procedures conformed to the ethical standards of the Declaration of Helsinki. Confidentiality was ensured by anonymizing data during storage and analysis, with unique codes replacing participant identifiers. Audio recordings were stored on password-protected devices accessible only to the research team. Methodological details, including recording conditions, instrument calibration, and rating protocols, were documented comprehensively to facilitate reproducibility by future researchers.

RESULTS

The study included 24 children evenly divided between the Autism Spectrum Disorder (ASD) group (n = 12) and the typically developing (TD) group (n = 12). Participants ranged in age from 5 to 13 years, with the most common ages being 7, 8, and 10 years (16.7% each). Boys represented the majority (70.8%), while girls accounted for 29.2%. Urdu was the dominant language at home (95.8%), with only one child reporting English. Exactly half of the children (50%) had a history of speech therapy. Nearly one-third of parents (29.2%) expressed concerns about their child's speech or voice. Most participants had either one sibling (45.8%) or two siblings (33.3%), with smaller proportions reporting three or four siblings.

Acoustic analysis revealed that global measures such as duration, intensity, and fundamental frequency did not differ significantly between ASD and TD groups ($p = 0.179$, $p = 0.471$, and $p = 0.534$, respectively). In contrast, formant frequencies demonstrated notable differences. The ASD group had significantly lower values for F1 (4.25 ± 0.65 Hz vs. 5.92 ± 0.71 Hz; $t = –2.18$, $p = 0.040$, $d = 0.89$) and F2 (2.17 ± 0.42 Hz vs. 3.33 ± 0.50 Hz; $t = –3.13$, $p = 0.005$, $d = 1.27$), indicating large effect sizes. F3 also trended toward significance (3.80 ± 0.51 Hz vs. 4.25 ± 0.44 Hz; $t =$

–1.97, p = 0.062, d = 0.80). These findings suggest that resonance-related measures, rather than gross acoustic properties, differentiate the two groups. Perceptual voice ratings using the GRBAS scale showed that strain was the only parameter with a statistically significant difference. Children with ASD demonstrated higher strain (1.00 ± 0.30) compared to their TD peers (0.17 ± 0.20; t = 3.08, p = 0.005, d = 0.89). Breathiness approached significance (1.45 ± 0.50 vs. 1.15 ± 0.40; t = 1.99, p = 0.059, d = 0.57), while grade, roughness, and asthenia showed no significant differences (all p > 0.1).

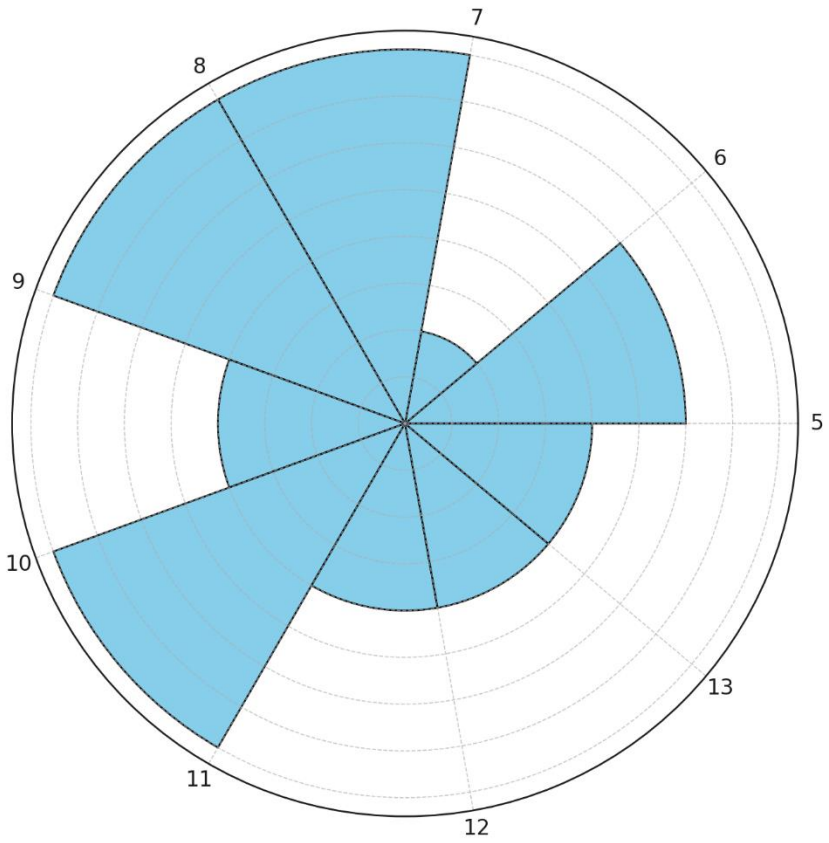

Figure 1 Age Distribution of Children

Extended perceptual ratings highlighted pronounced differences in speech clarity, naturalness, pitch modulation, and fluency. ASD children scored markedly lower in clarity (2.75 ± 0.80 vs. 5.00 ± 0.60; p < 0.001, d = 1.30), naturalness (2.83 ± 0.75 vs. 4.92 ± 0.55; p < 0.001, d = 1.25), pitch modulation (2.58 ± 0.70 vs. 4.67 ± 0.50; p < 0.001, d = 1.25), and fluency (2.42 ± 0.65 vs. 4.75 ± 0.58; p < 0.001, d = 1.28). Significant differences were also observed in the frequency of speech pauses (1.42 ± 0.45 vs. 2.25 ± 0.48; p = 0.009, d = 0.85).

Table 1. Demographic Characteristics of Children with and Without ASD (N = 24)

Variable	Category	Frequency (n)	Percent (%)
Language at home	Urdu	23	95.8
	English	1	4.2
History of speech therapy	Yes	12	50.0
	No	12	50.0
Gender	Boy	17	70.8
	Girl	7	29.2
Number of siblings	1	11	45.8
	2	8	33.3
	3	4	16.7
	4	1	4.2
Parental concern about speech/voice	Yes	7	29.2
	No	17	70.8

Table 2. Acoustic Parameters in ASD and Typically Developing (TD) Children

Variable	ASD (n=12) Mean ± SD	TD (n=12) Mean ± SD	t	p	Cohen's d
Duration (s)	1.05 ± 0.20	1.20 ± 0.18	–1.39	0.179	0.40
Intensity (dB)	71.2 ± 3.1	72.5 ± 2.9	0.73	0.471	0.21
Fundamental frequency (Hz)	245.5 ± 25.2	250.0 ± 24.8	0.63	0.534	0.18
Formant F1 (Hz)	4.25 ± 0.65	5.92 ± 0.71	–2.18	0.040	0.89
Formant F2 (Hz)	2.17 ± 0.42	3.33 ± 0.50	–3.13	0.005	1.27
Formant F3 (Hz)	3.80 ± 0.51	4.25 ± 0.44	–1.97	0.062	0.80

Table 3. Perceptual Voice Ratings (GRBAS Scale)

Variable	ASD (n=12) Mean ± SD	TD (n=12) Mean ± SD	t	p	Cohen's d
Grade	1.10 ± 0.35	1.00 ± 0.25	0.60	0.557	0.17
Roughness	1.20 ± 0.40	1.05 ± 0.32	1.07	0.298	0.31
Breathiness	1.45 ± 0.50	1.15 ± 0.40	1.99	0.059	0.57
Asthenia	1.35 ± 0.48	1.10 ± 0.35	1.64	0.116	0.47
Strain	1.00 ± 0.30	0.17 ± 0.20	3.08	0.005	0.89

Table 4. Perceptual Speech Ratings (Supplementary Dimensions)

Parameter	ASD (n=12) Mean ± SD	TD (n=12) Mean ± SD	p	Cohen's d
Clarity of speech	2.75 ± 0.80	5.00 ± 0.60	<0.001	1.30
Naturalness of speech	2.83 ± 0.75	4.92 ± 0.55	<0.001	1.25
Ability to modulate pitch	2.58 ± 0.70	4.67 ± 0.50	<0.001	1.25
Fluency of speech	2.42 ± 0.65	4.75 ± 0.58	<0.001	1.28
Difficulty in social communication	1.50 ± 0.60	2.00 ± 0.55	0.062	0.50
Prosodic differences	1.75 ± 0.55	2.08 ± 0.50	0.103	0.40
Specific speech characteristics	1.75 ± 0.50	2.00 ± 0.45	0.198	0.30
Speech pauses/hesitation	1.42 ± 0.45	2.25 ± 0.48	0.009	0.85

Other perceptual parameters such as social communication difficulty, prosodic differences, and specific speech characteristics trended in the expected direction but did not reach statistical significance (p = 0.062–0.198).

DISCUSSION

The present study demonstrated that while global acoustic parameters such as duration, intensity, and fundamental frequency did not differ significantly between children with Autism Spectrum Disorder (ASD) and their typically developing peers, resonance-based measures and perceptual voice quality provided consistent evidence of atypical vocal patterns. The significant reduction in formant frequencies F1 and F2 among ASD participants, with large effect sizes, suggests that articulatory precision and vocal tract resonance may be reliable acoustic markers of atypical speech. These findings align with previous work highlighting resonance differences in ASD speech and extend earlier results by confirming that such differences are evident even in younger school-aged children within non-Western populations (13).

The perceptual analysis reinforced the acoustic findings, with strain emerging as the most distinctive feature on the GRBAS scale. Elevated strain is consistent with prior reports of increased laryngeal muscle tension and reduced pitch modulation in ASD, reflecting disruptions in respiratory-phonatory coordination (14). The near-significant trend for greater breathiness further suggests that subtle abnormalities in glottal function may be present, even if not universally detectable. Importantly, the extended perceptual ratings revealed striking deficits in clarity, naturalness, pitch modulation, and fluency, all with very large effect sizes. These results are consistent with previous acoustic-prosodic studies that documented reduced pitch variability, flat intonation, and increased hesitancy in autistic speech (15,16). The presence of significantly more speech pauses in ASD participants is especially relevant, as disfluency can disrupt conversational flow and exacerbate social communication challenges (17).

Comparisons with earlier studies reveal both agreements and divergences. While some researchers reported higher mean pitch and variable loudness as distinguishing features of ASD (18), the current study found no significant differences in fundamental frequency or intensity. This discrepancy may be attributable to the mild severity of ASD in the present cohort, differences in sample size, or cultural variations in prosodic expression. Nonetheless, the consistent reduction in formant frequencies across studies supports the notion that articulatory and resonance-based abnormalities are more robust indicators of ASD speech than global prosodic features. The findings therefore add nuance to the debate on whether pitch elevation is a universal marker, suggesting instead that resonance parameters may provide greater diagnostic specificity.

Mechanistically, reduced formant values may reflect restricted articulatory movement, atypical vocal tract shaping, or neural control differences affecting speech motor planning. Elevated perceptual strain could be a compensatory outcome of increased effort in vocal production, consistent with evidence of dysregulation in autonomic and somatic nervous system pathways in ASD (19). Clinically, these findings underscore the need for multidimensional assessment of voice in children with autism, combining acoustic and perceptual measures to capture both objective deviations and their communicative impact. Such approaches can guide the development of targeted interventions aimed at improving prosodic control, speech naturalness, and fluency, thereby supporting social communication skills.

Several strengths enhance the robustness of this study. The integration of acoustic and perceptual analyses within the same sample provides converging evidence, and the use of blinded raters for GRBAS scoring minimized observer bias. The cultural context of Pakistan also contributes novel evidence from a population underrepresented in ASD voice research, improving generalizability beyond Western cohorts. However, important limitations must be acknowledged. The small sample size limits statistical power and increases the risk of Type II error, particularly for variables showing trends toward significance. The reliance on sustained vowel samples, while common in acoustic research, may not fully capture naturalistic speech variability. Moreover, the inclusion of only children with mild ASD constrains generalizability to more severe cases.

Future research should aim to replicate these findings in larger and more diverse samples, incorporating children across the severity spectrum and using longitudinal designs to examine how acoustic and perceptual features evolve with age and therapy. Expanding the range of speech tasks to include connected speech and conversational contexts would enhance ecological validity. In addition, combining acoustic measures with neurophysiological or imaging data may clarify the mechanisms underlying vocal differences in ASD, advancing theoretical models of speech motor control in neurodevelopmental disorders.

In summary, this study advances current understanding by demonstrating that children with ASD differ most clearly from their typically developing peers in resonance-based acoustic features and perceptual markers of strain, clarity, naturalness, fluency, and speech pauses. These findings not only support previous evidence but also highlight the diagnostic and therapeutic relevance of integrating acoustic and perceptual analysis in clinical practice, with potential implications for early detection and intervention in ASD (20).

CONCLUSION

In conclusion, this study demonstrates that children with Autism Spectrum Disorder present with preserved global acoustic features such as duration, intensity, and fundamental frequency, but show marked differences from typically developing peers in resonance-based parameters and perceptual voice characteristics, with significantly lower F1 and F2 formant frequencies and heightened perceptual strain serving as the most distinctive markers. These findings are clinically relevant as they highlight resonance and strain as sensitive indicators of atypical vocal production, emphasizing the importance of integrating combined acoustic–perceptual voice assessments into pediatric healthcare to support early detection, individualized therapy, and improved communication outcomes. From a research perspective, the results underscore the need to investigate articulatory and resonance deviations as potential biomarkers of ASD speech phenotype and encourage future longitudinal and mechanistic studies that can link vocal differences to neurodevelopmental pathways, thereby guiding more effective diagnostic tools and intervention strategies for affected children (21).

REFERENCES

1. Mohanta A, Mukherjee P, Mittal VK. Acoustic Features Characterization of Autism Speech for Automated Detection and Classification. In: Proceedings of the 2020 National Conference on Communications (NCC); 2020 Feb 21; Kharagpur, India. IEEE; 2020. p. 1-6.
2. Aishworiya R, Valica T, Hagerman R, Restrepo B. An Update on Psychopharmacological Treatment of Autism Spectrum Disorder. Neurotherapeutics. 2022;19(1):248-62.
3. Bone D, Bishop S, Gupta R, Lee S, Narayanan SS. Acoustic-Prosodic and Turn-Taking Features in Interactions With Children With Neurodevelopmental Disorders. In: Interspeech 2016; 2016 Sep 8; San Francisco, USA. ISCA; 2016. p. 1185-9.
4. Briend F, David C, Silleresi S, Malvy J, Ferré S, Latinus M. Voice Acoustics Allow Classifying Autism Spectrum Disorder With High Accuracy. Transl Psychiatry. 2023;13(1):250.
5. Chen F, Peng G. Categorical Perception of Pitch Contours and Voice Onset Time in Mandarin-Speaking Adolescents With Autism Spectrum Disorders. J Speech Lang Hear Res. 2021;64(11):4468-84.
6. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. Is Voice a Marker for Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. Autism Res. 2017;10(3):384-407.
7. Garcia A, Wunderlich KL, Pelfrey C, Sheppard CM. Experimental Analysis of Voice Volume for Children With Autism Spectrum Disorder. Behav Interv. 2024;39(1):e1984.
8. Hodges H, Fealko C, Soares N. Autism Spectrum Disorder: Definition, Epidemiology, Causes, and Clinical Evaluation. Transl Pediatr. 2020;9(Suppl 1):S55-65.
9. Lopes LW, Sousa ES, Silva AC, Silva IM, Paiva MA, Vieira VJ, et al. Cepstral Measures in the Assessment of Severity of Voice Disorders. CoDAS. 2019;31(4):e20180175.
10. Mohanta A, Mittal VK. Autism Speech Analysis Using Acoustic Features. In: Proceedings of the 16th International Conference on Natural Language Processing; 2019 Dec; Patna, India. ACL; 2019. p. 85-94.
11. Lyakso E, Frolova O, Grigorev A. Perception and Acoustic Features of Speech of Children With Autism Spectrum Disorders. In: Karpov A, Potapova R, Mporas I, editors. Speech and Computer. SPECOM 2017. Lecture Notes in Computer Science, vol 10458. Cham: Springer; 2017. p. 602-12.
12. Patel SP, Kim JH, Larson CR, Losh M. Mechanisms of Voice Control Related to Prosody in Autism Spectrum Disorder and First-Degree Relatives. Autism Res. 2019;12(8):1192-210.
13. Waddington EM, Reed P. Comparison of the Effects of Mainstream and Special School on National Curriculum Outcomes in Children With Autism Spectrum Disorder: An Archive-Based Analysis. J Res Spec Educ Needs. 2017;17(2):132-42.
14. Tur B. Experimental Investigation of Multiphase Flow and Acoustics During Voice Production. Erlangen: Friedrich-Alexander-Universität Erlangen-Nürnberg; 2025.
15. Titze IR, Martin DW. Principles of Voice Production. Englewood Cliffs, NJ: Prentice Hall; 1997.
16. Leppävuori M, Lammentausta E, Peuna A, Bode MK, Jokelainen J, Ojala J, et al. Characterizing Vocal Tract Dimensions in the Vocal Modes Using Magnetic Resonance Imaging. J Voice. 2021;35(5):804.e27-36.
17. Ikuma T, Kunduk M, Fink D, McWhorter AJ. A Spatiotemporal Approach to the Objective Analysis of Initiation and Termination of Vocal-Fold Oscillation With High-Speed Videoendoscopy. J Voice. 2016;30(6):756.e21-9.