Article

Not peer-reviewed version

Acoustic Analysis of Vowel Production in Adults with Autism Spectrum Disorder: Beyond the Notion of a Unitary Autistic Voice

Georgios P. Georgiou * and Maria Paphiti

Posted Date: 7 April 2026

doi: 10.20944/preprints202604.0377.v1

Keywords: vowel; autism spectrum disorder; acoustic analysis; adults; Cypriot Greek

doi:10.20944/preprints202604.0377.v1

Article

Acoustic Analysis of Vowel Production in Adults with Autism Spectrum Disorder: Beyond the Notion of a Unitary Autistic Voice

Georgios P. Georgiou [1,*] and Maria Paphiti [2]

[1] Phonetic Lab, Department of Languages and Literature, University of Nicosia, Nicosia, Cyprus
[2] Department of Health Sciences, European University Cyprus, Nicosia, Cyprus
* **Correspondence**: georgiou.georg@unic.ac.cy

Abstract

Speech in autism spectrum disorder (ASD) carries distinctive acoustic signatures that can offer valuable insight into the nature of autistic communication and its identification. Among these, vowel production remains insufficiently understood, despite its central role in speech intelligibility. This study investigates whether ASD is associated with systematic differences in vowel production. Eighteen Cypriot Greek adults with ASD and 18 peers with neurotypical development (ND), comparable in age, gender, education, nonverbal IQ, and verbal fluency, completed a controlled reading task. Participants produced disyllabic pseudowords embedding the five Greek vowels across four stress-syllable contexts. Acoustic analyses measured vowel-space organization, static spectral properties (F0, F1, F2, F3), dynamic trajectories (ΔF0, ΔF1, ΔF2, ΔF3), duration, and voice-quality indices (jitter, shimmer, harmonic-to-noise ratio [HNR], intensity). Bayesian models were used to evaluate group and vowel-specific differences. The results revealed a larger vowel-space area and greater vowel-space dispersion in the ASD group relative to the ND group, indicating a more expanded and dispersed acoustic vowel system. Group differences in individual acoustic measures were mostly selective rather than global: the clearest effects emerged in vowel-specific patterns of pitch, formants, and some dynamic formant measures. By contrast, duration, jitter, shimmer, and intensity did not show robust vowel-specific group differences. Among voice-quality measures, HNR showed the most consistent group difference, with ASD speakers showing higher HNR across all vowels. These findings challenge the notion of a single, uniform autistic voice, instead demonstrating that autism-related speech differences are multidimensional, vowel-specific, and language-sensitive. They therefore underscore the critical importance of segment-focused, cross-linguistically grounded approaches for advancing theory, assessment, and future speech-based identification in autism research.

Keywords: vowel; autism spectrum disorder; acoustic analysis; adults; Cypriot Greek

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by persistent differences in social communication and social interaction, together with restricted and repetitive patterns of behavior, interests, or activities (Centers for Disease Control and Prevention, 2025; Kamp-Becker, 2024). Current diagnostic frameworks further emphasize that ASD is highly heterogeneous in its clinical presentation, developmental trajectory, and functional impact across individuals (Lord et al., 2018; Masi et al., 2017). Because no single biological marker is sufficient for diagnosis, behavioral manifestations remain central to clinical identification and characterization (Cortese et al., 2025). Among these, speech is particularly important, as it is one of the most immediate and socially consequential channels through which communicative differences are perceived in both clinical and everyday settings (Diehl et al., 2008; Trayvick et al., 2024). Accordingly, vocal production has become

 doi:10.20944/preprints202604.0377.v1

a key domain in research on autism, especially in efforts to understand how autism-related differences are expressed in observable communicative behavior.

Clinical descriptions have long suggested that some autistic individuals produce speech that is perceived as acoustically distinctive, often in terms of atypical prosody, unusual intonation, altered rhythm, or differences in voice quality (Godel et al., 2023; McCann & Peppé, 2003). Acoustic phonetics offers a useful framework for investigating such impressions because it allows speech to be analyzed through measurable properties of the acoustic signal. In autism research, this has motivated growing interest in whether ASD is associated with reliable differences in core dimensions of speech production, including prosodic control, phonation, articulation, and temporal organization (Briend et al., 2023; Lau et al., 2022). These dimensions can be examined through measures such as fundamental frequency (F0), vowel formants, within-vowel spectral change, duration, and voice-quality indices, including jitter, shimmer, harmonic-to-noise ratio (HNR), and intensity (Kent & Vorperian, 2018; Lee et al., 2023; Van Santen et al., 2009). Because these measures can be extracted from short and controlled speech samples, they provide a means of studying speech production while minimizing the lexical, pragmatic, and interactional demands that characterize spontaneous conversation (Georgiou & Kaskampa, 2024). In this way, acoustic analysis may contribute to the more precise characterization of autism-related speech patterns and, potentially, to the development of speech-based tools for screening-oriented identification.

1.1. Differences Between Individuals with ASD and ND in Various Acoustic Measures

Vowel production in ASD has frequently been characterized as atypical on perceptual grounds, but the acoustic-phonetic literature has increasingly specified the signal dimensions on which ASD and ND speakers converge or diverge.

F0-related differences have been examined less often at the vowel level than at the broader prosodic level, in which ASD individuals appear to have higher mean F0, wider F0 range, and greater F0 variability on average than ND controls (Asghari et al., 2021). At the level of vowel F0, the evidence does not currently support a robust, consistent ASD–ND difference. The clearest direct comparison comes from Arciuli and Bailey (2019), who examined lexical-stress production in autistic and neurotypical Italian-speaking children and measured F0 on the first two vowels of target trisyllabic words produced in a picture-naming task. They found that the two groups showed a similar magnitude of stress contrastivity for vowel F0, indicating a null group difference in this vowel-level pitch measure. In contrast, Mohanta and Mittal (2022) using a controlled word production task found that children with ASD showed higher mean F0 and greater F0 variability than ND children across the five English vowels. There is also related evidence from tone-language research. Xu et al. (2022) compared Mandarin-speaking autistic and neurotypical children and found atypical patterns in the height and shape of lexical tones. Similarly, Guo et al. (2022) reported that F0 in Mandarin tone production was higher in children with ASD than in ND peers across prosodic positions and lexical tones.

With respect to formants and vowel-space measures, the literature suggests that ASD–ND differences can emerge, although the direction of effects is not consistent across studies. Lyakso et al. (2016) reported larger vowel formant triangles in autistic children than in neurotypical peers, consistent with a more expanded vowel-space pattern. Mohanta and Mittal (2022) reported higher F1, F2, F3, and F5 values across all five target English vowels, as well as higher F4 values for three vowels, in autistic children than in neurotypical children, interpreting these differences as evidence of atypical vowel production. In contrast, Sadiq et al. (2025) reported significantly lower F1 and F2 in the ASD group; however, they used a sustained vowel elicitation method. In another line of evidence, Kissine et al. (2021) found that autistic adults showed reduced dispersion in F1–F3 space during native vowel production and less variable imitation of unfamiliar vowel-like targets, a pattern interpreted as greater articulatory stability but reduced phonetic flexibility. Finally, Hong et al. (2023) found no clear ASD–ND difference in vowel-formant entrainment, suggesting that some aspects of vowel adaptation may remain intact under controlled interactional conditions. Overall, these

 doi:10.20944/preprints202604.0377.v1

findings suggest that ASD–ND differences in vowel production can be reflected in formant values and vowel-space organization, but that the sign of the effect varies across studies, ranging from higher/lower formants or expanded vowel space to reduced dispersion, with some studies also reporting no reliable group difference in interactive adaptation.

Evidence on duration at the vowel level is limited and tends toward null findings rather than a clearly consistent atypical pattern. In the lexical-stress study by Arciuli and Bailey (2019), vowel duration was measured directly, and the ASD and neurotypical groups showed a similar magnitude of stress contrastivity. In addition, Sadiq et al. (2025) found no significant group difference in duration. Together, these findings suggest that vowel timing may often be comparable to neurotypical production under structured elicitation conditions, although broader speech-timing studies have reported occasional duration differences in other task contexts (Maffei et al., 2023). For jitter, shimmer, HNR, and intensity, the vowel-focused literature is smaller. For example, Kissine et al. (2021) reported lower jitter and lower shimmer in autistic adults, consistent with greater stability of voicing during vowel production. However, this pattern is not universal, and broader reviews note that shimmer differences are often absent or inconsistent (Bone et al., 2014), and that perturbation-related measures are especially sensitive to vowel type, elicitation conditions, and recording quality (MacCallum et al., 2011; Yousef et al., 2024). Evidence for HNR and intensity at the vowel level is more limited and less consistent. Some studies (e.g., Guo et al., 2022) report that HNR does not appear as a vowel-level marker of ASD, although Briend et al. (2023) found higher HNR in children with ASD during a repetition task using nonwords composed of the vowels [i a u]. Intensity findings are likewise limited: Arciuli and Bailey (2019) found largely similar group performance between ASD and ND children overall, and, similarly, Sadiq et al. (2025) found no significant global group intensity difference between ASD and ND individuals in this measure.

Overall, the vowel-production literature does not support a single, uniform acoustic profile distinguishing autistic from neurotypical speakers. Instead, some studies report differences in formants, vowel-space geometry, and phonatory stability, whereas others find no reliable group differences in vowel F0, duration, or broader vowel-based stress marking. The most balanced conclusion is therefore that ASD–ND differences in vowel production are measure-specific, task-sensitive, and frequently accompanied by substantial overlap and null findings.

1.1. Challenges and gaps in the Literature

Although many studies report acoustic differences between ASD and ND speakers, the evidence base remains difficult to consolidate into a stable profile. Reviews and meta-analyses suggest that ASD-related differences can be detected in vocal production, especially in prosodic dimensions, but they also show substantial variation across studies in both effect size and direction (Asghari et al., 2021; Fusaroli et al., 2017; Ma et al., 2024). As a result, the field has not yet established which acoustic measures can be treated as robust, generalizable markers of ASD rather than task- or sample-specific correlates. A major reason for this uncertainty is methodological variability. Measures often interpreted as indices of phonatory stability or voice quality are highly sensitive to recording conditions, segmentation practices, analysis settings, and signal quality (Carson et al., 2003; Dejonckere et al., 2012; Manfredi et al., 2011; Titze, 1995; Vogel et al., 2009). Even under typical clinical conditions, these measures can vary systematically with vowel identity, fundamental frequency, and vocal intensity (Brockmann et al., 2011). In addition, acoustic differences observed in conversational or narrative speech are often difficult to interpret because such tasks engage not only speech production but also lexical retrieval, discourse organization, pragmatic adaptation, affective stance, and interactional coordination (Ma et al., 2024; Patel et al., 2020). Consequently, apparent group differences may sometimes reflect elicitation, task demands, or processing decisions rather than stable differences in speech production itself. This has motivated the use of more structured elicitation paradigms, which can reduce some sources of task-related variability and allow more controlled assessment of speech output (Diehl & Paul, 2009, 2013; McCann & Peppé, 2003). Such designs are especially useful for isolating segmental and phonatory properties more precisely.

 doi:10.20944/preprints202604.0377.v1

Finally, the existing literature is heavily weighted toward child samples, with comparatively less evidence from adolescents and adults. This makes it difficult to determine whether reported acoustic characteristics reflect stable ASD-related patterns or developmental differences that may change with age.

A second limitation concerns language coverage. Much of the available evidence comes from English-speaking samples, even though the acoustic realization of prosody and segmental structure is shaped by language-specific phonological systems and culturally patterned speech norms (Fusaroli et al., 2022; Gussenhoven, 2002). Cross-linguistic work has already shown that not all potentially relevant acoustic dimensions behave similarly across languages. For example, Lau et al. (2022) found that some rhythm-related features generalized more readily across English and Cantonese, whereas intonation-related features were more language-dependent. This suggests that acoustic correlates of ASD cannot simply be assumed to transfer unchanged from one language to another. These concerns are especially relevant for Greek, and particularly Cypriot Greek, which remains largely absent from the autism acoustics literature despite having a well-described vowel system (Georgiou, 2023a; Georgiou & Savva, 2025). Existing Greek and Cypriot Greek ASD research (e.g., Charalambous Darden et al., 2016; Christopoulou et al., 2021; Zarokanellou et al., 2025) has focused more on broader language abilities than on acoustic-phonetic analyses of speech production. To date, the only phonetics-oriented study in this area appears to be Georgiou (2023b), which examined vowel identification in normal versus whispered speech and therefore addressed speech perception rather than production. As a result, there is still very limited evidence on how ASD and ND speakers of Cypriot Greek, or Greek in general, differ in vowel production, voice quality, or vowel-space organization.

The present study addresses these gaps by focusing on controlled vowel production in Cypriot Greek using phonotactically legal pseudoword frames. This design minimizes lexical and pragmatic confounds, allows consistent segmentation of vowel intervals, and provides a clearer basis for comparing ASD and ND adult speakers across multiple acoustic dimensions. In doing so, the study contributes new language-specific evidence to an underrepresented area and responds directly to the broader need for more controlled, cross-linguistically informed work in adult autism acoustics.

1.1. This Study

The present study examines whether speakers with ASD and ND peers differ in vowel production across a broad set of acoustic dimensions, including static spectral measures (F0, F1, F2, F3), dynamic spectral measures (ΔF0, ΔF1, ΔF2, ΔF3), temporal structure (duration), and voice-quality measures (jitter, shimmer, HNR, and intensity). The study is original in providing a controlled, segment-focused acoustic analysis of Cypriot Greek, a linguistic variety that remains largely absent from the autism acoustics literature despite the field's increasing recognition that acoustic markers may not generalize straightforwardly across languages. By focusing on a well-defined five-vowel system in a tightly controlled elicitation paradigm, the study addresses an important empirical gap while extending autism speech research beyond the predominantly English-based evidence base.

The contribution of the study is therefore twofold. First, it provides language-specific baseline evidence for Cypriot Greek, where production-based acoustic work on ASD is currently very limited. Second, it offers a more fine-grained account of ASD–ND differences by examining not only overall group effects, but also vowel-specific contrasts within a Bayesian modeling framework. In this way, the study moves beyond broad claims of "atypical prosody" or "atypical voice" and instead asks which specific acoustic dimensions and which specific vowels differentiate the two groups when communicative demands are minimized and measurement conditions are tightly controlled. This is an important step toward clarifying the mixed findings in the literature and toward evaluating which acoustic features may prove most robust for future descriptive, stratificational, and potentially screening-oriented work in ASD.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0377.v1

2. Methodology

2.1. Participants

Thirty-six individuals with Cypriot Greek as their native language participated in the study. The experimental group included 18 participants (15 males/3 females) with a diagnosis of ASD documented in official clinical records (*M diagnosis age* = 9.88; *SD* = 10.59). Diagnoses had been made by experienced clinicians, neurologists, and psychiatrists, in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria. Participants in the ASD group were aged 18–40 years and were recruited from specialized autism centers in Cyprus and, where eligible, from adult students in upper secondary schools. Exclusion criteria included a history of intellectual disability and co-occurring language disorder. The control group comprised 18 participants (15 males/3 females) with ND, aged 18–42 years, recruited across Cyprus. These participants had no reported history of neurodevelopmental, language, cognitive, or behavioral disorders.

To characterize the sample, participants completed brief background measures of empathy, autistic traits, verbal fluency, and nonverbal reasoning. Empathy was assessed with the Greek adaptation of the Empathy Quotient (EQ; Baron-Cohen & Wheelwright, 2004; Pehlivanidis et al., 2021), a self-report measure in which higher scores indicate greater empathy. Autistic traits were assessed with the Greek version of the Autism-Spectrum Quotient (AQ; Baron-Cohen et al., 2001), with higher scores indicating greater autistic trait expression. Verbal fluency was measured using the Greek adaptation of the fluency test (Kosmidis et al., 2004), including both semantic and phonemic tasks, from which total word production, mean cluster size, and number of switches were calculated. Nonverbal reasoning was assessed with Raven's Progressive Matrices, Second Edition (Raven et al., 2018), using Sets B–E, with total correct responses used as the score.

Groups were comparable for linguistic, cognitive, social, and other variables. More specifically, the ASD and ND groups did not differ significantly in terms of chronological age ($t = –0.17$, $df = 34$, $p = 0.86$), nonverbal IQ ($t = –1.97$, $df = 34$, $p = 0.06$), education ($W = 149$, $p = 0.68$), and gender distribution ($\chi^2 = 0$, $df = 1$, $p = 1$). They also did not differ in any measures of the fluency test; semantics test: total word production (WP) ($t = –1.65$, $df = 34$, $p = 0.11$), (b) mean cluster size (SC) ($t = 0.46$, $df = 34$, $p = 0.65$), reflecting semantic or phonemic grouping of responses, and (c) number of switches (SW) ($t = –1.72$, $df = 34$, $p = 0.09$); phonemic test: WP ($t = –1.97$, $df = 34$, $p = 0.06$), SC ($t = –0.83$, $df = 34$, $p = 0.41$), SW ($t = –0.83$, $df = 34$, $p = 0.08$). Furthermore, as expected, the two groups differed significantly in Empathy Quotient (EQ) scores ($t(34) = –3.55$, $p = 0.001$) and Autism–Spectrum Quotient (AQ) scores ($t(34) = 7.11$, $p < 0.001$), with the ASD group showing lower EQ scores and higher AQ scores, confirming autistic traits. All participants reported normal or corrected–to–normal vision and hearing.

Before participation, all individuals (and guardians where applicable) received written information about the study procedures and their rights, and provided informed consent in accordance with the Declaration of Helsinki. Participant demographics and test scores are summarized in Table 1.

Table 1. Participants' demographics and scores in the tests.

Variable	ASD (n = 18)	ND (n = 18)
Sex, male/female, n	15/3	15/3
Age, years (SD)	24.1 (6.43)	24.5 (7.11)
EQ score (SD)	29.7 (12.4)	43.1 (10.2)
AQ score (SD)	29.6 (6.49)	14.8 (5.96)
Nonverbal IQ(SD)	26.6 (3.91)	28.9 (3.35)
Highest educational attainment (UNESCO Institute for Statistics, 2012), median (IQR)	3	3

 doi:10.20944/preprints202604.0377.v1

Fluency: Semantics test – WP score (SD)	49.56 (7.79)	53.33 (5.84)
Fluency: Semantics test – SC score (SD)	7.44 (3.57)	6.94 (2.90)
Fluency: Semantics test – SW score (SD)	34.61 (7.70)	40.44 (12.18)
Fluency: Phonemic test – WP score (SD)	30.67 (6.24)	36.11 (9.92)
Fluency: Phonemic test – SC score (SD)	1.89 (1.37)	2.44 (2.48)
Fluency: Phonemic test – SW score (SD)	26.44 (3.75)	29.56 (6.17)

2.2. Materials and Tools

The materials consisted of disyllabic pseudowords presented in four contexts: /ˈsVsa/, /sVˈsa/, /ˈVsa/, and /Vˈsa/ (V = vowel). Each context targeted the five Greek vowels /i e a o u/. The target vowels were always those found in the first syllable. All stimuli were constructed to conform to Greek phonotactic constraints and to be word-like while remaining semantically meaningless. The four contexts were selected to allow the target vowels to be elicited in a range of controlled phonetic environments (the presence vs absence of an onset consonant and stressed vs. unstressed vowel). The fricative /s/ was chosen as a stable consonantal frame with relatively minimal coarticulatory influence on adjacent vowels, enabling clearer assessment of vowel realization. Comparing /sVsa/ with /Vsa/ contexts makes it possible to control for the effect of onset consonants and stress on vowel production. All contexts are phonotactically well-formed in Greek and word-like without carrying lexical meaning, ensuring naturalistic articulation while avoiding lexical or semantic confounds.

2.3. Procedure

2.3.1. Production Test

Each participant completed the production task individually in a quiet room. Participants were seated at a desk facing a PC monitor at an approximate distance of 50–70 cm. Stimuli were presented in Standard Modern Greek orthography, and participants were instructed to read each word aloud at a comfortable speaking rate and volume. Recordings were made using a Zoom H5 digital recorder (44.1 kHz, 16-bit; WAV format). The recorder placement and mouth-to-recorder distance were kept constant across participants, and input levels were adjusted prior to testing to avoid clipping. Each participant produced 80 phrases (5 vowels × 4 contexts × 4 repetitions), yielding 2880 phrases overall. Items were randomized within each repetition block. A brief familiarization phase preceded the experiment to ensure accurate reading; practice items were not included in the analysis. Productions containing misreadings, disfluencies, or excessive noise were re-recorded when possible or excluded. Participants were allowed short breaks of up to five minutes between repetition blocks.

2.3.2. Feature Extraction

The target words produced by the speakers were extracted and processed in Praat (Boersma & Weenink, 2026). Spectrograms and waveforms were visually inspected to identify key acoustic landmarks, enabling the measurement of vowel boundaries and the extraction of F0, F1–F3, duration, jitter, shimmer, HNR, and intensity (for example, see Figure 1). The analyses were carried out using the following settings: a 0.025 s window length, 50 Hz pre-emphasis, and a spectrogram display range of up to 5500 Hz. Acoustic measurements were obtained over the vowel interval for each token. In the /ˈsVsa/ and /sVˈsa/ contexts, the measurement window began at the offset of frication noise of the initial /s/ (i.e., vowel onset) and ended at the offset of vocalic periodicity, immediately before the onset of the following consonant. In the /ˈVsa/ and /Vˈsa/ contexts, which lacked an initial fricative, the window began at the onset of vocalic periodicity and ended at the offset of vocalic periodicity preceding the onset of /s/.

For the measurement of F0, the pitch range was set between 75 and 300 Hz. F0, F1–F3, and intensity were measured at the midpoint of the vowel segment to minimize coarticulatory effects

 doi:10.20944/preprints202604.0377.v1

from adjacent sounds. Vowel duration corresponded to the total length of the vowel segment. For the computation of jitter, shimmer, and HNR, we used Praat's jitter (local), shimmer (local), and harmonicity (cc) functions. Dynamic spectral change was quantified using ΔF0, ΔF1, ΔF2, and ΔF3, computed as the difference between acoustic values at 75% and 25% of the vowel interval (75% − 25%). These measures captured the within-vowel change in F0 and Fx while minimizing the influence of segment boundaries. Vowel durations were manually labeled by the second author through identification of the onset and offset boundaries of each vowel token (see also Georgiou & Kaskampa, 2024). Polygonal vowel space area (VSA) was computed from the F1–F2 coordinates of the five vowels to quantify the overall size of the vowel space, whereas vowel space dispersion (VSD) was calculated as the mean distance of individual vowel tokens from the group vowel-space centroid, indexing the degree of acoustic spread within the vowel system.

To reduce between-speaker anatomical variation, F1, F2, and F3 were normalized using the Lobanov *z*-score method (Lobanov, 1971), whereas F0 was normalized using a semitone transformation relative to each speaker's median F0. The remaining acoustic measures were also standardized using *z*-scores to enhance cross-speaker comparability and support more robust statistical analysis.

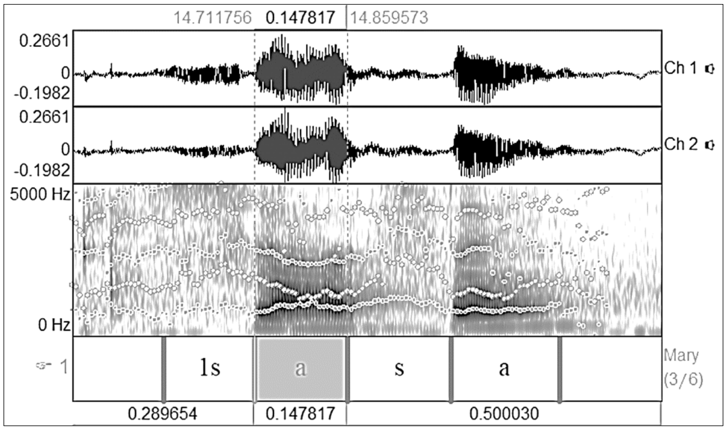

Figure 1. Example from the segmentation process.

2.4. Statistical Analyses

Group differences in VSA and VSD were analyzed using Bayesian regression models implemented in the brms package in R (R Core Team, 2026). Two separate models were estimated, one with VSA as the outcome and one with VSD as the outcome, with group entered as the fixed-effect predictor. For both models, weakly informative priors were specified for all regression parameters, including the intercept and residual variance, using a Student-t distribution with 3 degrees of freedom, mean 0, and scale 1 [student_t(3, 0, 1)]. These priors were chosen to provide regularization without imposing overly restrictive assumptions on the parameter estimates (see Georgiou, 2024). Inference regarding group differences in VSA and VSD was based on the posterior estimates and their 95% credible intervals (CrI).

To examine group differences in acoustic measures at the vowel level, we fitted a series of Bayesian mixed-effects regression models using the brms package in R. Separate models were estimated for each dependent variable: F0, F1, F2, F3, ΔF0, ΔF1, ΔF2, and ΔF3, duration, jitter, shimmer, HNR, and intensity. In each model, vowel, group, the vowel × group interaction, context, and stress were included as fixed effects, and participant was included as a random intercept to account for repeated observations within speakers. Weakly informative Student-t priors, student_t(3, 0, 1), were specified for the regression coefficients, intercept, participant-level standard deviation, and residual standard deviation. Inference was based on posterior estimates and their 95% CrIs. To evaluate group differences within individual vowels, we conducted directed Bayesian hypothesis

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0377.v1

tests on linear combinations of model coefficients. These tests assessed whether the effect of group (ND relative to the reference group, namely ASD) was negative for each vowel category, based on the model parameterization and the relevant group and vowel × group terms. These tests were interpreted using the evidence ratio (ER) and posterior probability (PP), with ER ≥ 10 taken to indicate strong evidence for a group difference in the predicted direction and ER ≤ 0.1 taken to indicate strong evidence against the predicted direction. Posterior predictive checks were conducted for all models and indicated good model fit.

3. Results

3.1. Vowel Space Measures

The two vowel space measures focused on the VSA and VSD. The Bayesian regression models indicated a credible effect of group on VSA. Compared to the ASD group (reference level), the ND group showed a lower VSA ($\beta = -0.56$, $SE = 0.03$, 95% CrI [−0.62, −0.50]). In addition, there was a credible effect of group on VSD, with the ND group showing lower VSD ($\beta = -0.08$, $SE = 0.01$, 95% CrI [−0.09, −0.07]) compared to the ASD group. Figure 2 shows the F1 × F2 vowel space of the ASD and ND groups.

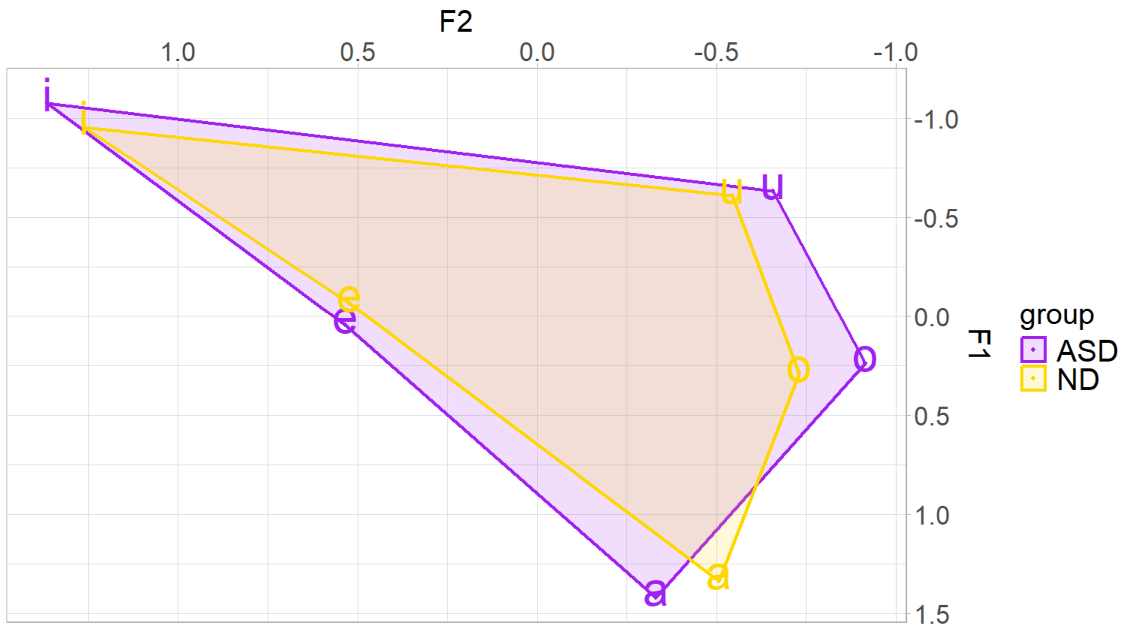

Figure 2. Two-dimensional vowel space (F1–F2) comparing productions of vowels for ASD (purple) and ND (yellow) groups. Vowel categories are connected to form polygonal spaces.

3.2. Acoustic Measures

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0377.v1

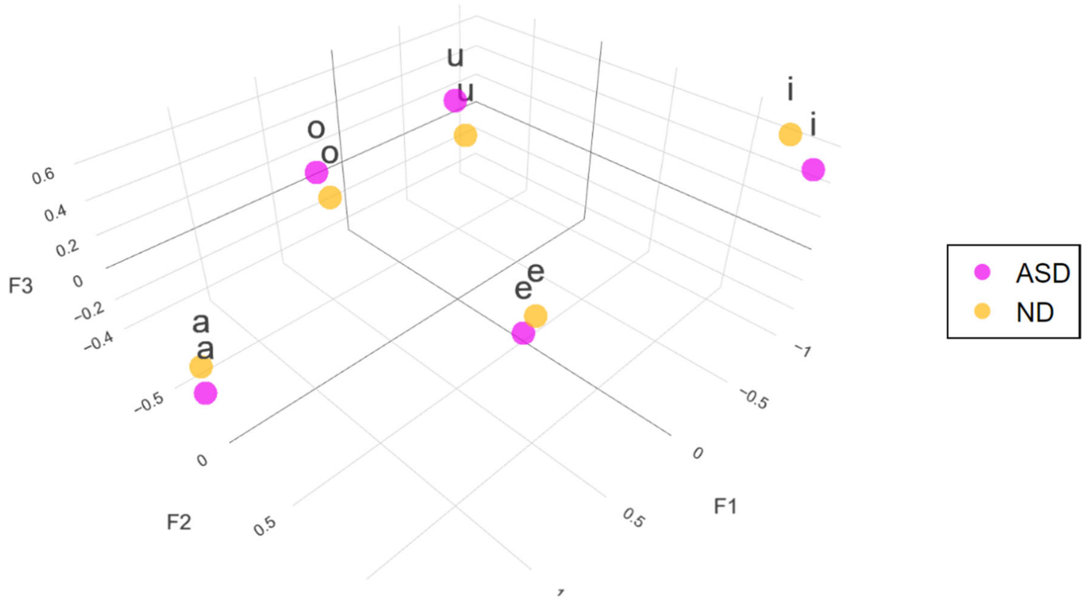

Figure 3. 3D acoustic vowel space (F1–F2–F3) comparing productions of the five vowels between ASD (purple) and ND (yellow) groups.

The Bayesian models indicated that group effects were generally small and inconsistent across acoustic measures (see Table 2). For F0, there was no credible difference between groups. Similarly, no reliable group differences were observed for F1 or F3. In contrast, the ND group showed a small but credible decrease in F2 relative to the ASD group ($\beta = -0.17$, $SE = 0.02$, 95% CrI [–0.27, –0.08]). For dynamic measures, no credible group differences were observed for ΔF0, ΔF2, or ΔF3. However, there was weak evidence of a positive group effect for ΔF1 ($\beta = 0.11$, $SE = 0.05$, 95% CrI [0.00, 0.22]), suggesting slightly greater F1 movement in the ND group. Figure 3 illustrates the vowel acoustic space for F1, F2, and F3.

Table 2. The results of the Bayesian models regarding F0, formants, and dynamic trajectories. The intercept terms are vowela and groupASD.

	Estimate	Est.Error	l –95% CI	u–95% CI	Rhat Bulk	ESS Tail	ESS
F0							
Intercept	0.79	0.14	0.50	1.08	1.01	950	1466
vowele	0.26	0.13	0.01	0.50	1	1893	2770
voweli	0.70	0.13	0.45	0.94	1	1482	2326
vowelo	0.35	0.13	0.10	0.59	1	2047	2548
vowelu	0.74	0.12	0.49	0.98	1	1816	2683
groupND	–0.32	0.20	–0.71	0.06	1.01	736	1154
contextmedial	–0.01	0.06	–0.12	0.11	1	6033	3357
stressunstressed	–1.92	0.06	–2.03	–1.80	1	5833	3033
vowele:groupND	0.07	0.18	–0.28	0.43	1	1802	2511
voweli:groupND	0.04	0.18	–0.32	0.38	1	1512	2573
vowelo:groupND	–0.09	0.18	–0.45	0.26	1	1985	2636
vowelu:groupND	0.20	0.18	–0.14	0.55	1	1802	2664

F1							
Intercept	1.50	0.04	1.43	1.57	1	2071	2584
vowele	–1.37	0.05	–1.46	–1.28	1	2273	2507
voweli	–2.47	0.05	–2.57	–2.38	1	2324	2917
vowelo	–1.17	0.05	–1.26	–1.09	1	2372	2565
vowelu	–2.05	0.05	–2.14	–1.96	1	2207	2632
groupND	–0.07	0.05	–0.16	0.02	1	1563	2444
contextmedial	–0.04	0.02	–0.08	0.00	1	8062	2895
stressunstressed	–0.12	0.02	–0.16	–0.08	1	8989	2845
vowele:groupND	–0.03	0.06	–0.17	0.09	1	2047	2648
voweli:groupND	0.19	0.07	0.06	0.32	1	2111	2559
vowelo:groupND	0.13	0.07	0.00	0.25	1	2180	2592
vowelu:groupND	0.10	0.06	–0.03	0.23	1	1906	2467
F2							
Intercept	–0.41	0.04	–0.48	–0.34	1	2729	3144
vowele	0.87	0.05	0.77	0.96	1	2819	3497
voweli	1.68	0.05	1.59	1.78	1	2918	3075
vowelo	–0.58	0.05	–0.68	–0.49	1	3054	3104
vowelu	–0.32	0.05	–0.42	–0.23	1	2749	3198
groupND	–0.17	0.05	–0.27	–0.08	1	2029	2974
contextmedial	0.07	0.02	0.03	0.11	1	8613	2703
stressunstressed	0.09	0.02	0.05	0.13	1	10696	2737
vowele:groupND	0.16	0.07	0.03	0.29	1	2726	3079
voweli:groupND	0.07	0.07	–0.06	0.21	1	2412	2606
vowelo:groupND	0.36	0.07	0.23	0.49	1	2698	2984
vowelu:groupND	0.28	0.07	0.15	0.41	1	2549	3160
F3							
Intercept	–0.42	0.06	–0.53	–0.30	1	2506	3141
vowele	0.07	0.08	–0.08	0.21	1	2594	3140
voweli	0.93	0.08	0.78	1.08	1	3059	3369
vowelo	0.51	0.08	0.36	0.66	1	2978	2798
vowelu	0.75	0.08	0.60	0.90	1	2739	2994
groupND	0.00	0.07	–0.15	0.15	1	1876	2448
contextmedial	–0.04	0.03	–0.11	0.02	1	9074	2607
stressunstressed	–0.02	0.03	–0.09	0.05	1	8375	2703
vowele:groupND	0.05	0.10	–0.16	0.25	1	2247	2262
voweli:groupND	0.20	0.11	–0.01	0.41	1	2751	3116
vowelo:groupND	–0.06	0.10	–0.26	0.15	1	2620	3076
vowelu:groupND	–0.18	0.11	–0.39	0.04	1	2494	3023
ΔF0							
Intercept	0.16	0.15	–0.12	0.45	1	525	1137
vowele	–0.11	0.11	–0.33	0.11	1	1689	2367

voweli	–0.20	0.11	–0.42	0.01	1	1683	2404
vowelo	–0.07	0.11	–0.28	0.14	1	1603	1504
vowelu	–0.20	0.11	–0.42	0.01	1	1589	2345
groupND	0.03	0.20	–0.37	0.41	1	570	1059
contextmedial	–0.29	0.05	–0.39	–0.18	1	4455	3209
stressunstressed	–0.83	0.05	–0.93	–0.72	1	5648	3038
vowele:groupND	–0.04	0.16	–0.35	0.28	1	1575	2480
voweli:groupND	–0.10	0.16	–0.41	0.22	1	1554	2587
vowelo:groupND	0.17	0.15	–0.13	0.47	1	1442	2194
vowelu:groupND	0.05	0.16	–0.26	0.37	1	1397	2156
ΔF1							
Intercept	–0.22	0.04	–0.30	–0.13	1	2794	3178
vowele	0.11	0.05	0.00	0.22	1	2825	3325
voweli	0.28	0.05	0.17	0.38	1	2893	3263
vowelo	0.23	0.05	0.13	0.34	1	2822	3051
vowelu	0.38	0.06	0.27	0.49	1	2998	3320
groupND	0.11	0.05	0.00	0.22	1	2049	2743
contextmedial	0.08	0.03	0.04	0.13	1	9814	2762
stressunstressed	–0.05	0.03	–0.09	0.01	1	10385	2766
vowele:groupND	–0.02	0.08	–0.18	0.13	1	2517	2868
voweli:groupND	–0.12	0.08	–0.27	0.03	1	2640	2963
vowelo:groupND	–0.18	0.08	–0.32	–0.03	1	2715	3297
vowelu:groupND	–0.25	0.08	–0.40	–0.09	1	2882	3096
ΔF2							
Intercept	–0.02	0.04	–0.10	0.07	1	2450	2475
vowele	–0.01	0.05	–0.12	0.10	1	2768	3248
voweli	0.01	0.06	–0.10	0.12	1	2565	2571
vowelo	–0.02	0.06	–0.13	0.09	1	2895	2965
vowelu	0.09	0.06	–0.02	0.20	1	2663	2600
groupND	–0.01	0.05	–0.12	0.09	1	1728	2561
contextmedial	–0.08	0.02	–0.12	–0.03	1	10696	2668
stressunstressed	0.08	0.02	0.03	0.13	1	8475	2742
vowele:groupND	0.04	0.08	–0.11	0.19	1	2602	3014
voweli:groupND	0.15	0.08	0.00	0.30	1	2393	3005
vowelo:groupND	–0.01	0.08	–0.16	0.14	1	2298	2944
vowelu:groupND	–0.12	0.08	–0.27	0.03	1	2496	2774
ΔF3							
Intercept	–0.08	0.06	–0.20	0.04	1	2587	2635
vowele	0.01	0.08	–0.15	0.16	1	2737	2481
voweli	–0.14	0.08	–0.29	0.01	1	2874	3220
vowelo	0.02	0.08	–0.14	0.18	1	3070	2933
vowelu	0.09	0.08	–0.07	0.25	1	2979	2862

doi:10.20944/preprints202604.0377.v1

groupND	0.06	0.08	–0.10	0.22	1	2021	2742
contextmedial	0.05	0.04	–0.02	0.12	1	11521	2780
stressunstressed	0.12	0.04	0.05	0.20	1	9274	2019
vowele:groupND	–0.06	0.11	–0.28	0.16	1	2638	3229
voweli:groupND	–0.03	0.11	–0.25	0.20	1	2510	3016
vowelo:groupND	–0.10	0.11	–0.32	0.12	1	2575	2827
vowelu:groupND	–0.11	0.11	–0.33	0.12	1	2801	2854

For the remaining acoustic measures, no credible group differences were observed. The ND group did not differ reliably from the ASD group in duration, jitter, shimmer, or intensity. For HNR, there was a negative trend for the ND group relative to the ASD group, but the credible interval included zero (see Table 3).

Table 3. Results of the Bayesian models regarding duration and voice quality measures. The intercept terms are vowela and groupASD.

	Estimate	Est.Error	l –95% CI	u–95% CI	Rhat	ESS Tail Bulk	ESS
Duration							
Intercept	0.40	0.10	0.20	0.60	1.01	672	1303
vowele	–0.07	0.08	–0.22	0.07	1	1593	2476
voweli	–0.12	0.08	–0.27	0.04	1	1683	2695
vowelo	–0.11	0.08	–0.25	0.05	1	1601	2072
vowelu	–0.18	0.08	–0.33	–0.03	1	1560	2261
groupND	0.04	0.14	–0.22	0.31	1.01	633	1440
contextmedial	–0.12	0.03	–0.19	–0.06	1	5745	2957
stressunstressed	–0.43	0.03	–0.49	–0.36	1	5516	2843
vowele:groupND	–0.09	0.11	–0.30	0.12	1	1685	2557
voweli:groupND	–0.16	0.11	–0.37	0.05	1	1657	2781
vowelo:groupND	–0.05	0.11	–0.27	0.16	1	1566	2546
vowelu:groupND	–0.05	0.11	–0.27	0.16	1	1637	2448
Jitter							
Intercept	–0.45	0.12	–0.69	–0.20	1.01	402	676
vowele	0.14	0.07	0.01	0.27	1	1255	1803
voweli	0.32	0.07	0.18	0.46	1.01	1254	2052
vowelo	0.03	0.07	–0.10	0.17	1	1308	1838
vowelu	0.12	0.07	–0.01	0.26	1	1331	2039
groupND	0.09	0.17	–0.25	0.43	1	453	862
contextmedial	–0.12	0.03	–0.18	–0.05	1	3126	2654
stressunstressed	0.68	0.03	0.62	0.74	1	3424	2873
vowele:groupND	–0.06	0.10	–0.25	0.13	1	1177	2169
voweli:groupND	–0.07	0.10	–0.27	0.12	1.01	1189	1955
vowelo:groupND	0.09	0.10	–0.10	0.28	1	1318	2092
vowelu:groupND	0.00	0.10	–0.18	0.19	1	1359	2092

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0377.v1

Shimmer							
Intercept	–0.18	0.12	–0.42	0.05	1.01	429	758
vowele	0.05	0.07	–0.10	0.19	1	1142	2074
voweli	0.05	0.07	–0.09	0.19	1	1149	1778
vowelo	–0.04	0.07	–0.18	0.10	1.01	1130	2452
vowelu	0.01	0.07	–0.12	0.16	1.01	995	2302
groupND	0.15	0.17	–0.17	0.48	1.01	478	890
contextmedial	–0.21	0.03	–0.28	–0.15	1	3731	2825
stressunstressed	0.45	0.03	0.39	0.52	1	3571	2919
vowele:groupND	–0.06	0.10	–0.26	0.14	1	1162	2071
voweli:groupND	0.11	0.10	–0.10	0.32	1	1248	2240
vowelo:groupND	–0.01	0.10	–0.20	0.19	1	1149	2259
vowelu:groupND	–0.10	0.10	–0.30	0.10	1.01	1149	2035
HNR							
Intercept	–0.02	0.15	–0.31	0.27	1.01	404	788
vowele	0.31	0.06	0.20	0.42	1	1418	1924
voweli	0.85	0.06	0.75	0.96	1	1466	2065
vowelo	0.40	0.05	0.30	0.51	1	1463	2388
vowelu	0.94	0.06	0.83	1.05	1	1267	1675
groupND	–0.36	0.20	–0.74	0.02	1.01	487	926
contextmedial	0.08	0.03	0.03	0.13	1	2311	2383
stressunstressed	–0.67	0.03	–0.72	–0.62	1	2750	2173
vowele:groupND	–0.01	0.08	–0.17	0.14	1	1349	1514
voweli:groupND	0.02	0.08	–0.14	0.18	1	1443	1417
vowelo:groupND	–0.03	0.08	–0.18	0.13	1	1538	1863
vowelu:groupND	–0.05	0.08	–0.20	0.11	1	1334	1562
Intensity							
Intercept	0.48	0.21	0.07	0.89	1.01	936	1383
vowele	–0.15	0.03	–0.22	–0.09	1	2566	2758
voweli	–0.28	0.03	–0.35	–0.21	1	2548	2679
vowelo	–0.03	0.03	–0.10	0.03	1	2658	2585
vowelu	–0.21	0.03	–0.27	–0.14	1	2672	2913
groupND	–0.26	0.29	–0.82	0.31	1.01	725	1263
contextmedial	0.17	0.02	0.15	0.20	1	4262	2603
stressunstressed	–0.59	0.01	–0.62	–0.56	1	4340	2462
vowele:groupND	0.07	0.05	–0.02	0.16	1	2560	2746
voweli:groupND	0.02	0.05	–0.07	0.11	1	2389	2862
vowelo:groupND	–0.02	0.05	–0.11	0.07	1	2697	2752
vowelu:groupND	0.03	0.05	–0.07	0.12	1	2641	2641

Evidence ratios indicated strong support for several vowel-specific group differences. For F0, there was strong evidence that the ND group had lower values than the ASD group for /i/ (ER = 11.62, PP = 0.92), /a/ (ER = 19.30, PP = 0.95), and /o/ (ER = 55.34, PP = 0.98). For F1, strong evidence indicated lower values in the ND group for /e/ (ER = 87.89, PP = 0.99) and /a/ (ER = 17.26, PP = 0.95), while there was strong evidence in the opposite direction for /i/ (ER = 0.01, PP. = 0.01). For F2, strong evidence showed lower values for the ND group in /i/ (ER = 56.14, PP= 0.98) and /a/ (ER = ∞, PP = 1.00), and higher values in the ND group for /o/ (ER = 0.00, PP = 0.00) and /u/ (ER = 0.01, PP = 0.01). For F3, strong evidence indicated lower values in the ND group for /u/ (ER = 77.43, PP = 0.99) and higher values for /i/ (ER = 0.00, PP = 0.00). No strong evidence was observed for ΔF0. For ΔF1, strong evidence indicated lower values in the ND group for /u/ (ER = 110.11, PP = 0.99) and higher values for /a/ (ER = 0.02, PP = 0.02) and /e/ (ER = 0.06, PP = 0.02). For ΔF2, strong evidence showed higher values in the ND group for /i/ (ER = 0.01, PP = 0.01) and lower values for /u/ (ER = 136.93, PP = 0.99). No strong evidence was observed for ΔF3. These results are presented in Table 4. Figure 4 illustrates the differences between the ASD and ND groups in terms of F0, formants, and dynamic trajectories.

Table 4. Results of hypothesis testing for F0, formants, and dynamic trajectories.

ND < ASD	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post Prob.
F0						
i	–0.28	0.20	–0.62	0.05	11.62	0.92
e	–0.24	0.20	–0.56	0.07	8.30	0.89
a	–0.32	0.20	–0.64	0.00	19.30	0.95
o	–0.41	0.20	–0.75	–0.08	55.34	0.98
u	–0.11	0.20	–0.44	0.21	2.46	0.71
F1						
i	0.12	0.05	0.04	0.20	0.01	0.01
e	–0.11	0.05	–0.19	–0.03	87.89	0.99
a	–0.07	0.05	–0.15	0.00	17.26	0.95
o	0.05	0.05	–0.03	0.13	0.17	0.14
u	0.03	0.05	–0.05	0.10	0.38	0.28
F2						
i	–0.10	0.05	–0.18	–0.02	56.14	0.98
e	–0.01	0.05	–0.09	0.07	1.60	0.62
a	–0.17	0.05	–0.25	–0.09	Inf	1.00
o	0.18	0.05	0.10	0.26	0.00	0.00
u	0.11	0.05	0.03	0.18	0.01	0.01
F3						
i	0.20	0.08	0.07	0.33	0.00	0.00
e	0.04	0.07	–0.08	0.16	0.38	0.27
a	0.00	0.07	–0.13	0.12	1.05	0.51
o	–0.06	0.07	–0.18	0.06	4.01	0.80
u	–0.18	0.08	–0.31	–0.05	77.43	0.99
ΔF0						
i	–0.07	0.20	–0.39	0.26	1.79	0.64
e	–0.02	0.20	–0.35	0.32	1.15	0.54
a	0.03	0.20	–0.30	0.35	0.79	0.44
o	0.20	0.20	–0.14	0.52	0.19	0.16
u	0.07	0.20	–0.26	0.40	0.54	0.35

ΔF1						
i	–0.01	0.05	–0.10	0.08	1.17	0.54
e	0.09	0.06	0.00	0.18	0.06	0.06
a	0.11	0.05	0.02	0.20	0.02	0.02
o	–0.07	0.05	–0.15	0.03	7.85	0.89
u	–0.13	0.06	–0.23	–0.04	110.11	0.99
ΔF2						
i	0.14	0.06	0.05	0.23	0.01	0.01
e	0.02	0.06	–0.07	0.12	0.48	0.32
a	–0.01	0.05	–0.10	0.08	1.36	0.58
o	–0.02	0.06	–0.11	0.07	1.80	0.64
u	–0.13	0.06	–0.22	–0.04	136.93	0.99
ΔF3						
i	0.03	0.08	–0.10	0.16	0.57	0.36
e	0.00	0.08	–0.14	0.13	1.08	0.52
a	0.06	0.08	–0.07	0.19	0.29	0.22
o	–0.04	0.08	–0.17	0.09	2.33	0.70
u	–0.05	0.08	–0.19	0.09	2.90	0.74

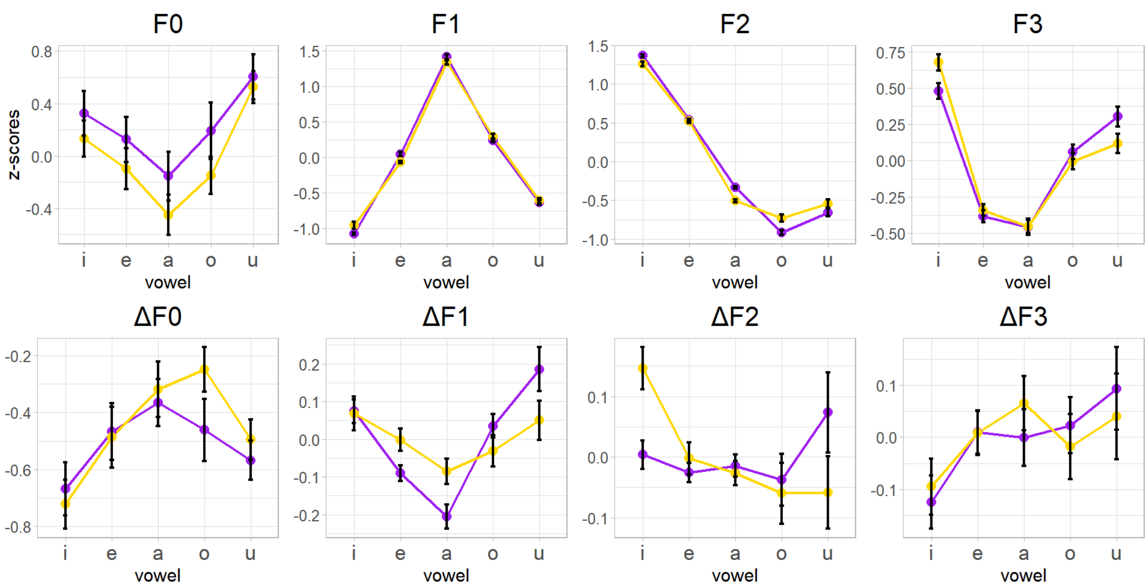

Figure 4. Vowel-specific acoustic measures (z-scores) for ASD (purple) and ND (yellow) groups across the following parameters: F0, F1, F2, F3, ΔF0, ΔF1, ΔF2, and ΔF3. Points represent mean values for each vowel, with error bars indicating variability (±SE).

Strong evidence of group differences was observed only for HNR. Specifically, the ND group showed consistently lower HNR than the ASD group across all vowels: /i/ (ER = 18.61, PP = 0.95), /e/ (ER = 27.17, PP = 0.96), /a/ (ER = 28.85, PP = 0.97), /o/ (ER = 36.04, PP = 0.97), and /u/ (ER = 46.62, PP = 0.98). No strong evidence of group differences was found for duration, jitter, shimmer (except /i/), or intensity. The results are shown in Table 5. Figure 5 illustrates the differences between the ASD and ND groups in terms of duration, voice quality measures, and intensity.

Table 5. Results of hypothesis testing for duration and voice quality measures.

ND < ASD	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post Prob.
Duration						
i	–0.12	0.14	–0.34	0.10	4.35	0.81
e	–0.05	0.13	–0.28	0.16	1.81	0.64
a	0.04	0.14	–0.19	0.26	0.63	0.39
o	–0.01	0.13	–0.23	0.20	1.06	0.52
u	–0.01	0.13	–0.23	0.20	1.09	0.52
Jitter						
i	0.03	0.17	–0.26	0.30	0.75	0.43
e	0.04	0.17	–0.26	0.32	0.68	0.40
a	0.09	0.17	–0.19	0.37	0.40	0.28
o	0.19	0.17	–0.09	0.46	0.16	0.14
u	0.10	0.17	–0.19	0.38	0.39	0.28
Shimmer						
i	0.26	0.17	–0.01	0.54	0.06	0.06
e	0.09	0.17	–0.18	0.37	0.38	0.28
a	0.15	0.17	–0.11	0.43	0.21	0.18
o	0.14	0.17	–0.13	0.42	0.25	0.20
u	0.05	0.17	–0.22	0.33	0.58	0.37
HNR						
i	–0.33	0.20	–0.67	0.00	18.61	0.95
e	–0.37	0.20	–0.70	–0.04	27.17	0.96
a	–0.36	0.20	–0.69	–0.03	28.85	0.97
o	–0.39	0.20	–0.72	–0.05	36.04	0.97
u	–0.41	0.20	–0.73	–0.08	46.62	0.98
Intensity						
i	–0.24	0.29	–0.71	0.23	4.12	0.80
e	–0.19	0.29	–0.65	0.27	2.96	0.75
a	–0.26	0.29	–0.72	0.20	4.57	0.82
o	–0.28	0.29	–0.75	0.18	5.36	0.84
u	–0.23	0.29	–0.70	0.23	3.83	0.79

 doi:10.20944/preprints202604.0377.v1

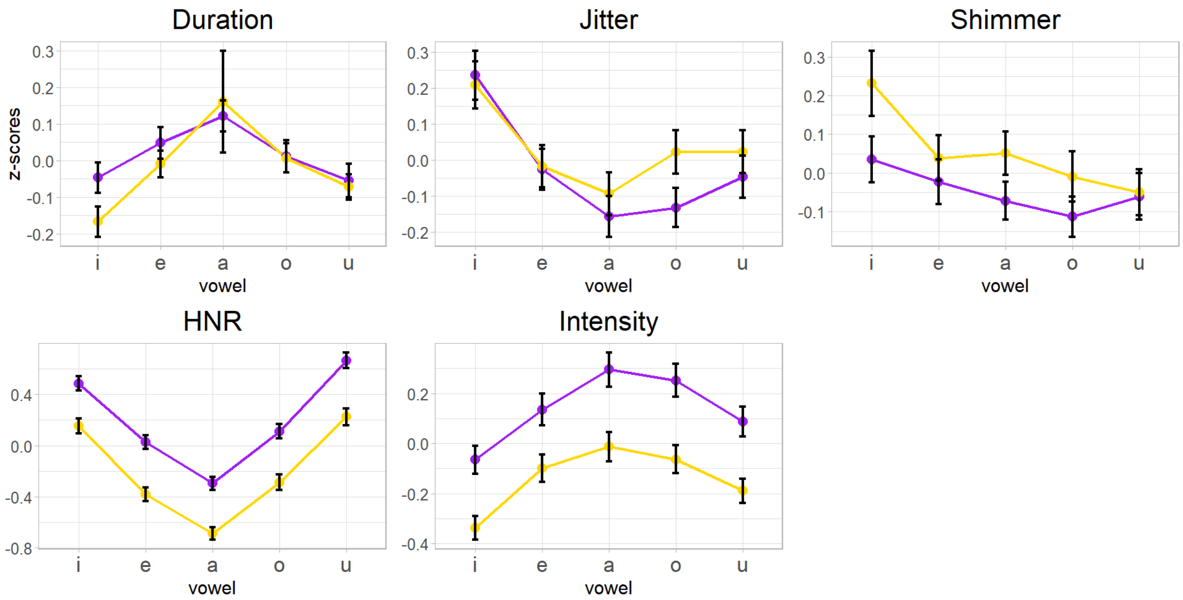

Figure 5. Vowel-specific acoustic measures (z-scores) for ASD (purple) and ND (yellow) groups across five parameters: duration, jitter, shimmer, HNR, and intensity. Points represent mean values for each vowel, with error bars indicating variability (±SE).

4. Discussion

The present study examined vowel production in Cypriot Greek speakers with ASD and ND peers using a controlled pseudoword-reading paradigm and a broad acoustic battery spanning static spectral measures, dynamic spectral change, temporal structure, vowel-space organization, and voice-quality indices. Three main findings emerged. First, the ASD group showed a larger vowel-space area and greater vowel-space dispersion than the ND group. Second, group differences at the level of individual acoustic parameters were mostly selective rather than global, with the clearest effects concentrated primarily in F0 and formant frequencies, and secondarily in the dynamic formant measures ΔF1 and ΔF2, and with many effects localized to particular vowels rather than extending uniformly across the system. Third, among the voice-quality measures, the most consistent result was that the ND group showed lower HNR than the ASD group across all vowels, whereas duration, jitter, shimmer, and intensity did not yield robust overall group differences. Collectively, these findings argue against a unitary notion of an "autistic voice" and instead support a more differentiated account in which autism-related differences in speech are distributed unevenly across acoustic domains, vowels, and task conditions (Asghari et al., 2021; Fusaroli et al., 2017, 2022; Ma et al., 2024; Van Santen et al., 2010). This study advances autism speech research by showing, in a tightly controlled framework and an underrepresented language variety, that autism-related speech differences are concentrated in vowel-space organization, selected spectral properties, and HNR rather than uniformly distributed across all acoustic domains.

A central contribution of the present study is the finding that the ASD group occupied a larger and more dispersed acoustic vowel space than the ND group. This pattern is noteworthy because previous research has identified ASD–ND differences in vowel-space organization, but has not converged on a single direction of effect. Lyakso et al. (2016), for example, reported larger vowel formant triangles in children with ASD, a pattern broadly compatible with the present findings. By contrast, Kissine et al. (2021) found the opposite tendency in autistic adults, who showed reduced dispersion and greater phonetic stability, interpreted as stronger attraction toward native vowel targets. The present results, therefore, align more closely with Lyakso et al. (2016) and suggest that, at least in Cypriot Greek and under the elicitation conditions used here, ASD speakers may realize vowel categories with greater acoustic spread and less centralized clustering than ND speakers. One

 doi:10.20944/preprints202604.0377.v1

possible interpretation is not simply that the task reduced lexical and pragmatic demands, but that it allowed segmental implementation itself to become more visible. Under these conditions, the larger vowel space may reflect a tendency toward more extreme realization of some vowel targets, greater trial-to-trial variability in articulatory positioning, or weaker compression of productions around category prototypes. On this view, the key difference may lie less in whether ASD speakers can produce the relevant vowel categories and more in how consistently and tightly those categories are organized in acoustic space. This interpretation also helps explain why the direction of effect may vary across studies. Vowel-space patterns are likely shaped by the interaction of language, age, task structure, and the specific balance between phonetic stability and flexibility that a given paradigm elicits (DiCanio et al., 2015; Pettinato et al., 2016).

The absence of a robust overall group effect for mean F0 also deserves comment. At first glance, this may seem surprising given the long-standing association between autism and atypical prosody (e.g., Asghari et al., 2021), but it is, in fact, compatible with current syntheses of the field. More specifically, the present data align with the results of Arciuli and Bailey (2019), who indicated no differences in F0 between autistic and neurotypical children in vowel acoustics. Yet, vowel-specific tests still showed evidence of F0 differences for three vowels (/i a o/), with the clearest effect in /o/. This suggests that autistic speech in the present data was not simply globally higher-pitched or lower-pitched. Rather, pitch differences, when present, appear to have been localized. Furthermore, the overall ASD–ND difference in F2 is in line with prior findings suggesting that formant structure may help distinguish autistic from neurotypical speech (Mohanta & Mittal, 2022; Sadiq et al., 2025). The particularly broad involvement of F2 suggests that group differences may be related in part to front–back acoustic differentiation. However, the statistical models did not indicate broad, uniform group effects across all formants, as vowel-specific hypothesis tests revealed a selective pattern of differences across the vowel system. This result is partly consistent with Mohanta and Mittal (2022), who also reported differences in some vowels between ASD and ND children, although these were observed in F4. In this study, evidence for group differences emerged in F1 for three vowels (/i e a/) and in F2 for four vowels (/i a o u/), with more limited effects in F3 (/i u/). This pattern is theoretically meaningful because several of the affected vowels are corner or peripheral vowels, which contribute disproportionately to vowel-space geometry and to the acoustic distinctiveness of the system as a whole. The dynamic results extend this picture, since although there was little evidence for systematic differences in ΔF0 or ΔF3, clearer effects emerged in ΔF1 for /e a u/ and in ΔF2 for /i u/; this indicates that within-vowel spectral movement also contributed to some extent to group differentiation. In sum, these findings suggest that the ASD group differed from the ND group not only in static vowel targets but also in aspects of within-vowel acoustic change over time. This interpretation is consistent with emerging work showing that autism-related speech differences are not well captured by a single dimension, but may instead be distributed across articulation, spectral structure, and dynamic acoustic patterning (Briend et al., 2023; Patel et al., 2020).

A further notable outcome was the lack of robust overall group differences in duration, jitter, shimmer, and intensity. In light of previous work, this pattern is not entirely surprising. For duration, the vowel-level literature remains limited and tends to favor null or near-null findings under structured elicitation conditions (Arciuli & Bailey, 2019). The present findings, therefore, add to a growing indication that vowel timing, at least in controlled tasks, may often be broadly comparable across ASD and ND speakers. The null findings for jitter and shimmer can also be interpreted cautiously in relation to the existing literature. Although Kissine et al. (2021) reported lower jitter and shimmer in autistic adults, broader evidence suggests that these measures are not consistently replicated and may be highly sensitive to methodological factors (MacCallum et al., 2011; Yousef et al., 2024). The absence of such differences in the present study – both global and vowel-specific – therefore supports the view that such measures may not generalize reliably across samples and paradigms, even when some individual studies report significant effects. A similar interpretation applies to intensity, for which no within-vowel differentiations between the two groups emerged (see also Arciuli & Bailey, 2019; Sadiq et al., 2025). The present null result is therefore consistent with the possibility that intensity is not a robust differentiating feature of vowel production in ASD under the current elicitation conditions.

Perhaps the most intriguing result concerns HNR, for which the ND group showed reliably lower values than the ASD group across all vowels. In many clinical voice contexts, higher HNR is interpreted as reflecting a more periodic, less noisy signal. On the surface, this may seem counterintuitive relative to work linking autism with hoarser or creakier voice qualities, or with reduced harmonicity in spontaneous speech. For example, Bone et al. (2014) reported that in spontaneous prosody, children's speech associated with ASD tended to show decreased HNR, while Fusaroli et al. (2022) identified increased hoarseness and creakiness as part of a minimal cross-linguistic acoustic profile. Our result is consistent with Briend et al. (2023), who also found higher HNR in children with ASD during a nonword repetition task. One possible explanation for our finding is that the higher HNR observed in the ASD group reflects a more controlled or constrained production style during the elicited word task. Because predictable task structure can facilitate performance consistency in autistic adults, it is plausible that the repetition-based format supported more stable and less noisy phonatory output, resulting in higher HNR (Lacroix et al., 2025). Because HNR is highly sensitive to recording quality, segmentation, and phonatory context, the present finding should be interpreted as a task-specific acoustic pattern rather than a direct clinical index of "better" or "worse" voice quality.

5. Conclusions

The present study provides new evidence that vowel production in Cypriot Greek differs between ASD and ND speakers in specific, non-uniform ways. The most robust differences were not global shifts in mean pitch, duration, or perturbation, but rather expanded vowel-space structure, selective pitch and formant contrasts, dynamic spectral differences, and consistently higher HNR across vowels in the ASD group. These findings strengthen the case for moving beyond broad labels and toward a more fine-grained account of autistic speech, one that recognizes the joint roles of articulation, phonation, vowel-specific properties, task structure, and language-specific phonetic organization. In this way, the study contributes both to the cross-linguistic literature on autism acoustics and to the broader effort to identify speech features that are sufficiently robust, interpretable, and context-sensitive to support future descriptive and potentially screening-oriented research. It also broadens the empirical basis for theories of autism-related speech by suggesting that acoustic distinctions may be shaped partly by the structure of the ambient vowel system.

The findings should be interpreted in light of several limitations. The sample was necessarily modest, and although the Bayesian framework is well-suited to smaller datasets, the present results still require replication in larger cohorts. In addition, the task was intentionally constrained; accordingly, the findings cannot be assumed to generalize directly to spontaneous speech, affective prosody, or conversation. This point is especially relevant given evidence that interactional synchrony and discourse context can reveal differences not visible in decontextualized speech (Ochi et al., 2019), and that listener impressions of autistic speech may depend on combinations of cues not well captured by single acoustic measures (Patel et al., 2020). A further limitation is that the study centered on vowel acoustics and did not model consonantal timing, speech rhythm, cepstral measures, or articulatory variability across repetitions. These domains are promising targets for future work, especially in light of recent research suggesting that speech consistency, coordination, and repetition-based measures may uncover additional dimensions of motor-acoustic organization in autism (Maffei et al., 2025).

Future research should therefore proceed in three directions. First, the present vowel-focused findings should be replicated in larger and clinically more stratified samples, including groups differing in language ability, support needs, and co-occurring conditions, to determine whether the observed acoustic profile characterizes autism broadly or particular phenotypic subgroups. Second, it will be important to test whether the vowel-space and HNR effects observed here persist across task types, especially when comparing nonword reading, sentence reading, narration, and dialogue

 doi:10.20944/preprints202604.0377.v1

within the same speakers. Third, because the current study identified localized vowel-specific effects, future analyses should explicitly model individual variability rather than relying only on group means; recent work increasingly argues that heterogeneity is not noise to be averaged away but a core property of autism that should itself be modeled (Liu et al., 2025). In practical terms, this means combining fine-grained acoustic phonetics with multilevel and multivariate approaches that can distinguish language-general tendencies from person-specific and language-specific patterns.

Funding: This study forms part of the research project CULTURE/AWARD-YR/0523, awarded to the first author, Georgios P. Georgiou, by the Cyprus Research and Innovation Foundation.

Ethics approval: The study was approved by the Cyprus National Bioethics Committee (approval no. ΕΕΒΚ ΕΠ 2024.01.57).

Informed consent : All participants were informed about the aims of the study prior to participation, and written informed consent was obtained from the participants themselves and/or their legal guardians, in accordance with the Declaration of Helsinki.

Conflict of interest: The authors declare no conflicts of interest.

Author contributions: GG: Conceptualization, GG: Data curation, GG: Formal analysis, GG: Funding acquisition, GG: Investigation, GG and MP: Methodology, GG: Project administration, GG: Resources, GG: Software, GG: Supervision, GG: Validation, GG: Visualization, GG: Writing – original draft, GG and MP: Writing – review & editing.

Data availability statement: Data are available on request from the first author.

Acknowledgments: We sincerely thank all participants for taking part in this study. We also thank the Day Care Centre B, Pancyprian Association for Autism, for its assistance with participant recruitment and the Department of Secondary General Education of the Cyprus Ministry of Education, Sport, and Youth for granting permission (no. 07.30.002.002) to access upper secondary schools for data collection. This work was also supported by the Phonetic Lab at the University of Nicosia.

References

Arciuli, J., and Bailey, B. (2019). An acoustic study of lexical stress contrastivity in children with and without autism spectrum disorders. *J. Child Lang.* 46, 142–152.

Asghari, S. Z., Farashi, S., Bashirian, S., and Jenabi, E. (2021). Distinctive prosodic features of people with autism spectrum disorder: a systematic review and meta-analysis study. *Sci. Rep.* 11:23093.

Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17.

Boersma, P., and Weenink, D. (2026). *Praat: Doing Phonetics by Computer* [Computer program]. Available at: http://www.fon.hum.uva.nl/praat/

Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S. H., Levitt, P., et al. (2014). The psychologist as an interlocutor in autism spectrum disorder assessment: insights from a study of spontaneous prosody. *J. Speech Lang. Hear. Res.* 57, 1162–1177.

Briend, F., David, C., Silleresi, S., Malvy, J., Ferré, S., and Latinus, M. (2023). Voice acoustics allow classifying autism spectrum disorder with high accuracy. *Transl. Psychiatry* 13:250.

Brockmann, M., Drinnan, M. J., Storck, C., and Carding, P. N. (2011). Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *J. Voice* 25, 44–53.

 doi:10.20944/preprints202604.0377.v1

Carson, C. P., Ingrisano, D. R. S., and Eggleston, K. D. (2003). The effect of noise on computer-aided measures of voice: a comparison of CSpeechSP and the Multi-Dimensional Voice Program software using the CSL 4300B module and Multi-Speech for Windows. *J. Voice* 17, 12–20.

Centers for Disease Control and Prevention. (2025). Clinical testing and diagnosis for autism spectrum disorder. U.S. Department of Health and Human Services. Available at: https://www.cdc.gov/ncbddd/autism/hcp-diagnosis.html

Charalambous Darden, N., Mavrou, K., and Armostis, S. (2016). "Dialect and standard language in ASD: the case of Cypriot Greek dialect and standard modern Greek," in *XI Autism-Europe International Congress*.

Christopoulou, M., Voniati, L., Drosos, K., and Armostis, S. (2021). Colorful semantics in Cypriot-Greek-speaking children with autism spectrum disorder. *Folia Phoniatr. Logop.* 73, 185–194.

Cortese, S., Bellato, A., Gabellone, A., Marzulli, L., Matera, E., Parlatini, V., et al. (2025). Latest clinical frontiers related to autism diagnostic strategies. *Cell Rep. Med.* 6.

Dejonckere, P. H., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., and Manfredi, C. (2012). To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition. *Biomed. Signal Process. Control* 7, 37–42.

DiCanio, C., Nam, H., Amith, J. D., García, R. C., and Whalen, D. H. (2015). Vowel variability in elicited versus spontaneous speech: evidence from Mixtec. *J. Phonetics* 48, 45–59.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., and McDonough, J. (2008). Resolving ambiguity: a psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain Lang.* 106, 144–152.

Diehl, J. J., and Paul, R. (2009). The assessment and treatment of prosodic disorders and neurological theories of prosody. *Int. J. Speech Lang. Pathol.* 11, 287–292.

Diehl, J. J., and Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Appl. Psycholinguist.* 34, 135–161.

Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J. R. M., and Weed, E. (2022). Toward a cumulative science of vocal markers of autism: a cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Res.* 15, 653–664.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., and Gaigg, S. B. (2017). Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Res.* 10, 384–407.

Georgiou, G. P. (2023a). Speakers of different L1 dialects with acoustically proximal vowel systems present with similar nonnative speech perception abilities: data from Greek listeners of Dutch. *Speech Commun.* 150, 32–40.

Georgiou, G. P. (2023b). Identification of native vowels in normal and whispered speech by individuals with autism spectrum disorder. *J. Autism Dev. Disord.* 53, 858–862.

Georgiou, G. P. (2024). Bayesian models are more sensitive than frequentist models in identifying differences in small datasets comprising phonetic data. *Stats* 7, 1483–1495.

Georgiou, G. P., and Kaskampa, A. (2024). Differences in voice quality measures among monolingual and bilingual speakers. *Ampersand* 12:100175.

Georgiou, G. P., and Savva, E. (2025). Exploring acoustic overlap in second language vowel productions. *Int. J. Appl. Linguist.* doi: 10.1111/ijal.12859

Godel, M., Robain, F., Journal, F., Kojovic, N., Latrèche, K., Dehaene-Lambertz, G., et al. (2023). Prosodic signatures of ASD severity and developmental delay in preschoolers. *npj Digit. Med.* 6:99.

Guo, C., Chen, F., Chang, Y., and Yan, J. (2022). Applying random forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production. *Biomed. Signal Process. Control* 77:103811.

Gussenhoven, C. (2002). "Intonation and interpretation: phonetics and phonology," in *Proceedings of Speech Prosody* (Aix-en-Provence: University de Provence), 47–57.

Hong, Y., Chen, S., Zhou, F., Chan, A., and Tang, T. (2023). Phonetic entrainment in L2 human-robot interaction: an investigation of children with and without autism spectrum disorder. *Front. Psychol.* 14:1128976.

 doi:10.20944/preprints202604.0377.v1

Kamp-Becker, I. (2024). Autism spectrum disorder in ICD-11—a critical reflection of its possible impact on clinical practice and research. *Mol. Psychiatry* 29, 633–638.

Kent, R. D., and Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: a review. *J. Commun. Disord.* 74, 74–97.

Kissine, M., Geelhand, P., Philippart De Foy, M., Harmegnies, B., and Deliens, G. (2021). Phonetic inflexibility in autistic adults. *Autism Res.* 14, 1186–1196.

Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., and Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: normative data, and clustering and switching strategies. *J. Int. Neuropsychol. Soc.* 10, 164–172.

Lacroix, A., Torija, E., Logemann, A., Baciu, M., Cserjesi, R., Dutheil, F., et al. (2025). Cognitive flexibility in autism: how task predictability and sex influence performances. *Autism Res.* 18, 281–294.

Lau, J. C., Patel, S., Kang, X., Nayar, K., Martin, G. E., Choy, J., et al. (2022). Cross-linguistic patterns of speech prosodic differences in autism: a machine learning study. *PLoS One* 17:e0269637.

Lee, S., Yeo, E. J., Kim, S., and Chung, M. (2023). Knowledge-driven speech features for detection of Korean-speaking children with autism spectrum disorder. *Phonetics Speech Sci.* 15, 53–59.

Liu, W., Pluta, A., Charpentier, C. J., and Rosenblau, G. (2025). A computational cognitive neuroscience approach for characterizing individual differences in autism: introduction to special issue. *Pers. Neurosci.* 8:e2.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49, 606–608.

Lord, C., Elsabbagh, M., Baird, G., and Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *Lancet* 392, 508–520.

Lyakso, E., Frolova, O., and Grigorev, A. (2016). "A comparison of acoustic features of speech of typically developing children and children with autism spectrum disorders," in *International Conference on Speech and Computer* (Springer International Publishing), 43–50.

Ma, W., Xu, L., Zhang, H., and Zhang, S. (2024). Can natural speech prosody distinguish autism spectrum disorders? A meta-analysis. *Behav. Sci.* 14:90.

MacCallum, J. K., Zhang, Y., and Jiang, J. J. (2011). Vowel selection and its effects on perturbation and nonlinear dynamic measures. *Folia Phoniatr. Logop.* 63, 88–97.

Maffei, M. F., Chenausky, K. V., Gill, S. V., Tager-Flusberg, H., and Green, J. R. (2023). Oromotor skills in autism spectrum disorder: a scoping review. *Autism Res.* 16, 879–917.

Maffei, M. F., Chenausky, K. V., Tager-Flusberg, H., and Green, J. R. (2025). An acoustic analysis of speech motor performance in autistic children. *J. Speech Lang. Hear. Res.* 68, 2802–2824.

Manfredi, C., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., and Dejonckere, P. (2011). Validity of jitter measures in non-quasi-periodic voices. Part II: the effect of noise. *Logoped. Phoniatr. Vocol.* 36, 78–89.

Masi, A., DeMayo, M. M., Glozier, N., and Guastella, A. J. (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neurosci. Bull.* 33, 183–193.

McCann, J., and Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *Int. J. Lang. Commun. Disord.* 38, 325–350.

Mohanta, A., and Mittal, V. K. (2022). Analysis and classification of speech sounds of children with autism spectrum disorder using acoustic features. *Comput. Speech Lang.* 72:101287.

Ochi, K., Ono, N., Owada, K., Kojima, M., Kuroda, M., Sagayama, S., et al. (2019). Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PLoS One* 14:e0225377.

Patel, S. P., Nayar, K., Martin, G. E., Franich, K., Crawford, S., Diehl, J. J., et al. (2020). An acoustic characterization of prosodic differences in autism spectrum disorder and first-degree relatives. *J. Autism Dev. Disord.* 50, 3032–3045.

Pehlivanidis, A., Tasios, K., Papanikolaou, K., Douzenis, A., and Michopoulos, I. (2021). Validation of the empathy quotient (EQ)—Greek version. *Psychiatriki* 32, 43–50.

Pettinato, M., Tuomainen, O., Granlund, S., and Hazan, V. (2016). Vowel space area in later childhood and adolescence: effects of age, sex and ease of communication. *J. Phonetics* 54, 1–14.

R Core Team. (2026). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Raven, J., Rust, J., Chan, F., and Zhou, X. (2018). *Raven's 2 Progressive Matrices, Clinical Edition (Raven's 2)*. Pearson.

 doi:10.20944/preprints202604.0377.v1

Sadiq, M. S., Nawaz, S., Abrar, S., and Amin, A. (2025). Comparison of acoustic and perceptual analysis of voice in autistic and normal children. *J. Health Wellness Community Res.* e791.

Titze, I. R. (1995). *Workshop on Acoustic Voice Analysis: Summary Statement*. National Center for Voice and Speech.

Trayvick, J., Barkley, S. B., McGowan, A., Srivastava, A., Peters, A. W., Cecchi, G. A., et al. (2024). Speech and language patterns in autism: towards natural language processing as a research and clinical tool. *Psychiatry Res.* 340:116109.

UNESCO Institute for Statistics. (2012). *International Standard Classification of Education: ISCED 2011*. UNESCO Institute for Statistics.

Van Santen, J. P., Prud'hommeaux, E. T., and Black, L. M. (2009). Automated assessment of prosody production. *Speech Commun.* 51, 1082–1097.

Van Santen, J. P., Prud'Hommeaux, E. T., Black, L. M., and Mitchell, M. (2010). Computational prosodic markers for autism. *Autism* 14, 215–236.

Vogel, A. P., Maruff, P., Snyder, P. J., and Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behav. Res. Methods* 41, 318–324.

Xu, K., Yan, J., Ma, C., Chang, X., and Chien, Y.-F. (2022). Atypical patterns of tone production in tone-language-speaking children with autism. *Front. Psychol.* 13:1023205.

Yousef, A. M., and Hunter, E. J. (2024). Sensitivity of acoustic voice quality measures in simulated reverberation conditions. *Bioengineering* 11:1253.

Zarokanellou, V., Gryparis, A., and Papanikolaou, K. (2025). Exploring narrative ability in Greek-speaking children with high-functioning ASD: associations with memory and attention. *Brain Sci.* 15:73.