

Investigating the Interplay Between Affective, Phonatory and Motoric Subsystems in Autism Spectrum Disorder Using a Multimodal Dialogue Agent

Hardik Kothare^{1,2}, Vikram Ramanarayanan^{1,2}, Oliver Roesler¹, Michael Neumann¹, Jackson Liscombe¹, William Burke¹, Andrew Cornish¹, Doug Habberstad¹, Alaa Sakallah^{2,3}, Sara Markuson^{2,3}, Seemran Kansara^{2,4}, Afik Faerman^{2,3}, Yasmine Bensidi-Slimane^{2,3}, Laura Fry^{2,3}, Saige Portera^{2,3}, David Suendermann-Oeft¹, David Pautler¹, Carly Demopoulos²

¹ Modality.AI, Inc.

² University of California, San Francisco

³ Palo Alto University

⁴ University of California, Berkeley

hardik.kothare, vikram.ramanarayanan@modality.ai

Abstract

We explore the utility of an on-demand multimodal conversational platform in extracting speech and facial metrics in children with Autism Spectrum Disorder (ASD). We investigate the extent to which these metrics correlate with objective clinical measures, particularly as they pertain to the interplay between the affective, phonatory and motoric subsystems. 22 participants diagnosed with ASD engaged with a virtual agent in conversational affect production tasks designed to elicit facial and vocal affect. We found significant correlations between vocal pitch and loudness extracted by our platform during these tasks and accuracy in recognition of facial and vocal affect, assessed via the Diagnostic Analysis of Nonverbal Accuracy-2 (DANVA-2) neuropsychological task. We also found significant correlations between jaw kinematic metrics extracted using our platform and motor speed of the dominant hand assessed via a standardised neuropsychological finger tapping task. These findings offer preliminary evidence for the usefulness of these audiovisual analytic metrics and could help us better model the interplay between different physiological subsystems in individuals with ASD.

Index Terms: multimodal conversational agent, autism spectrum disorder, affect production, motor coordination, audiovisual analytics.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that manifests as deficits in social communication and interaction [1]. Epidemiological surveys estimate the prevalence of ASD in the United States at 18.5 per 1,000 children [2]. Autism diagnosis rates have increased due to an increase in awareness and access to resources. In the state of California, the diagnosed autism incidence rate increased by 612% in the past two decades [3]; however, sociodemographic barriers such as race, economic challenges and stigma still hamper early detection and intervention in children with autism [4].

Individuals with ASD are known to have impaired processing of emotions (affect recognition) and impaired production of both vocal and non-verbal emotional expressions (affect production) during communication [5, 6, 7]. Acoustic characteristics like prosody and fundamental frequency of speech in children with ASD are often perceived as atypical when compared to speech in typically-developing children [8, 9]. Individuals

with ASD produce emotional phrases that are louder, longer, more variable in pitch, and sound less natural [10]. Facial affect production has been demonstrated to be equally difficult for individuals with ASD in both imitation and expression tasks [11]. Also, facial expressions of emotion are rated as more intense and less natural in individuals with ASD [12] by both neurotypical raters as well as raters with ASD [13].

Prior work has demonstrated the objective significance of automated quantitative assessment of atypical speech production and facial expression in ASD [14, 15, 16]. Such automated assessments can help clinicians objectively quantify skills in this domain of functioning for which no standardised, age-normed, validated measures are available. [17, 18, 19].

Moreover, the coordination of facial and vocal expressions during emotional speech production is less coordinated in ASD [20]. Due to such cross-domain atypicalities in ASD, a multimodal framework approach has been suggested to provide useful quantitative insights in early diagnosis and categorisation of ASD [21, 22, 23]. In this paper, we explore the interplay between multiple neuropsychological subsystems through examination of automatically-extracted voice and facial metrics during a novel conversational task and a comprehensive set of clinically-validated objective measures. We probe the feasibility of a scalable, low-cost and remotely-administrable multimodal conversational platform for this purpose. Such technology could potentially assist clinicians and researchers in gathering relevant diagnostic information and in longitudinally monitoring children with developmental disorders.

2. Multimodal Conversational Platform

NEMSI or NEurological and Mental health Screening Instrument is a cloud-based multimodal dialogue technology in which participants engage in a conversation with a virtual agent and speech and facial behaviours are elicited through a variety of exercises. Data analytics modules automatically extract relevant speech and facial metrics from the captured audiovisual data in real time and store them in a database. These metrics, with a detailed participant-wise, session-wise and task-wise breakdown, can be accessed by researchers via a user-friendly dashboard. A more detailed description of NEMSI, conversational protocols and a schematic of the software and hardware architecture of the system can be found in our previous publications [24, 25, 26].

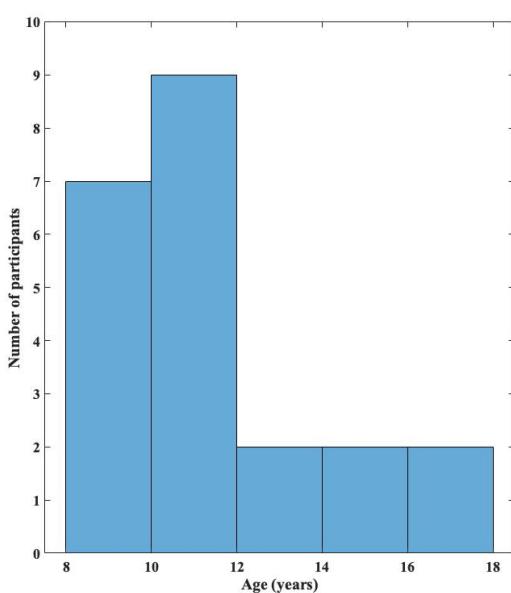


Figure 1: *Distribution of age in the cohort*

3. Methods

The current study was conducted on site at the University of California, San Francisco and was approved by the university's Institutional Review Board. Informed consent was obtained from the participants' guardians along with a written assent from the participants. Data from 22 participants (10 females, mean age: 11.37 ± 2.47 years, see Figure 1) diagnosed with ASD by a licensed clinical psychologist was included in the analysis.

3.1. Standardised Clinical Instruments

All participants / guardians underwent a thorough neuropsychological evaluation battery including the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) [27], the Autism Diagnostic Interview - Revised [28], the Behavior Assessment System for Children, Third Edition (BASC-3) [29], the Diagnostic Analysis of Nonverbal Accuracy (DANVA) [30, 31] and NEPSY-II [32]. The ADOS-2 scores for 11 participants were collected during the COVID-19 pandemic and were deemed invalid because the participants wore a mask during the assessment. These scores were excluded from the analysis. Additionally, participants also took part in a motor skills and fine motor dexterity assessment, including bilateral manual grip strength assessed with a Lafayette Hand Dynamometer Model J00105 [33], finger tapping speed assessed with a standard board-mounted finger tapper with counter [34], and the Grooved Pegboard Test [35]. All scores were age scaled based on the referenced norms.

3.2. Affect Production Task

The Affect Production Task (APT) was administered via Modality.AI's NEMSI platform. Participants were prompted to produce four emotions (happy, sad, angry and afraid) through five tasks:

1. Monosyllabic Emotion Production (8 turns, 2 turns per

emotion): a video stimulus prompted the users to say "oh" in a way that conveyed the specified emotion during every turn;

2. Sentence-length Emotion Production (8 turns, 2 turns per emotion): a video stimulus prompted the users to say "I'll be right back" in a way that conveyed the specified emotion during every turn;
3. Emotion-eliciting Situations (16 turns, 4 turns per emotion): the virtual agent narrated a situation and a corresponding emotional response at the end of which a picture depicting the situation was shown and the user was prompted to say "oh" in a way that conveyed the specified emotion that was situationally appropriate;
4. Monosyllabic Repetition (16 turns, 4 turn per emotion): an audio recording of "oh" in one of the four emotions was played during every turn and the user was asked to repeat the monosyllabic production in the same emotional inflection;
5. Facial Repetition (8 turns, 2 turns per emotion): a video prompted the user to make a face representing one of the four emotions like the person in the video during every turn.

3.3. Analytics

Speech metrics (Fundamental Frequency in Hz, Articulation Loudness in dB) and facial metrics (Eyebrow Height in pixels, Eye Openness in pixels, Lip Aperture in pixels, Mouth Surface Area in pixels², Jaw Velocity in pixels/frames and Jaw Acceleration pixels/frames²) were calculated in real time [25] and displayed on a dashboard accessible to the clinicians and the researchers involved in the study. Facial metrics in pixels were normalised within every subject by dividing the values by the inter-lachrymal distance in pixels for each subject.

3.4. Statistical Methods

Linear correlations were run between measures from standardised neuropsychological assessments and the speech and facial metrics derived from our platform. To control for false positives due to multiple comparison, the Benjamini-Hochberg procedure [36] was used to adjust the p-values using a false discovery rate of $\alpha = 0.05$. Findings that survived this adjusted significance threshold were included in this paper.

4. Observations

4.1. Language Level and IQ

The Clinical Evaluation of Language Fundamentals- Fifth Edition (CELF-5) [37] and the Weschler Intelligence Scale for Children - Full Scale IQ (WISC - FSIQ) [38] scores for the cohort are reported in Table 1.

4.2. Speech Metrics

We found significant positive correlations between the DANVA Adult Receptive Paralanguage Subscore (APSS) and the fundamental frequency of the participants' speech during the Emotion-Eliciting Situation subtask in the APT (Happy: $r^2 = 0.41$ and Sad: $r^2 = 0.54$) and also in the Monosyllabic Emotion Production subtask (Happy: $r^2 = 0.56$ and Angry: $r^2 = 0.44$). We also found positive correlations between the DANVA Child Receptive Paralanguage Subscore (CPSS) and the fundamental frequency of the participants' speech during the Emotion-

Table 1: *Language level and IQ measures in the cohort*

Measure	Range	Mean \pm SD
CELF-5 Core Language Score	59 - 143	100.77 \pm 22.06
CELF-5 Expressive Language Index	59 - 131	100.09 \pm 20.48
CELF-5 Receptive Language Index	58 - 141	100.73 \pm 23.08
WISC - FSIQ	50 - 136	102.95 \pm 19.79

Table 2: *Speech acoustic measures of fundamental frequency (F0) and articulation loudness extracted automatically correlate with the participants' accuracy at recognition of facial and vocal affect in audio recordings of adults and children*

Objective Clinical Measure	Extracted Analytics	Prompt	Sample Size	r^2	p-value
DANVA APSS	Fundamental Frequency	Emotion-Eliciting Situation HAPPY	15	0.41	0.0099
DANVA APSS	Fundamental Frequency	Monosyllabic Emotion Production HAPPY	19	0.56	0.0002
DANVA APSS	Fundamental Frequency	Emotion-Eliciting Situation SAD	17	0.54	0.0007
DANVA APSS	Fundamental Frequency	Monosyllabic Emotion Production ANGRY	18	0.44	0.0028
DANVA CPSS	Fundamental Frequency	Emotion-Eliciting Situation HAPPY	16	0.55	0.0016
DANVA CPSS	Fundamental Frequency	Monosyllabic Emotion Production HAPPY	19	0.48	0.0010
DANVA CPSS	Fundamental Frequency	Emotion-Eliciting Situation SAD	17	0.45	0.0034
DANVA CPSS	Fundamental Frequency	Monosyllabic Repetition AFRAID	16	0.36	0.0108
DANVA CPSS	Articulation Loudness	Sentence-length Emotion Production HAPPY	16	0.34	0.0180

Table 3: *Jaw kinematic measures (velocity and acceleration) extracted automatically correlate with motor speed of the dominant hand*

Objective Clinical Measure	Extracted Analytics	Prompt	Sample Size	r^2	p-value
Finger Tapping Dominant Hand	Maximum Jaw Velocity	Monosyllabic Repetition SAD	13	0.44	0.0139
Finger Tapping Dominant Hand	Maximum Jaw Velocity	Monosyllabic Repetition AFRAID	13	0.35	0.0317
Finger Tapping Dominant Hand	Maximum Jaw Velocity	Emotion-Eliciting Situation AFRAID	13	0.35	0.0345
Finger Tapping Dominant Hand	Maximum Jaw Acceleration	Emotion-Eliciting Situation AFRAID	13	0.42	0.0164

Eliciting Situation subtask (Happy: $r^2 = 0.55$ and Sad: $r^2 = 0.45$), the Monosyllabic Emotion Production subtask (Happy: $r^2 = 0.48$) and the Monosyllabic Repetition subtask (Angry: $r^2 = 0.36$). Fundamental frequency of speech during all these tasks did not correlate with participants' age so the effect of age on these correlations can be ruled out. Additionally, the DANVA CPSS score was also positively correlated with Articulation Loudness during the Sentence-length Emotion Production subtask (Happy: $r^2 = 0.34$). Please refer to Table 2 for more details.

4.3. Facial Metrics

Motor speed of the dominant hand assessed using a finger tapping task was found to be positively correlated with participants' maximum jaw velocity during the Monosyllabic Repetition subtask in the APT (Sad: $r^2 = 0.43$ and Afraid: $r^2 = 0.37$) and during the Emotion-Eliciting Situation subtask (Afraid: $r^2 = 0.36$). Motor speed of the dominant hand was also positively correlated with maximum jaw acceleration during the Emotion-Eliciting Situation subtask (Afraid: $r^2 = 0.42$). Please refer to Table 3 for more details.

5. Discussion

In this study, we found that speech metrics derived by NEMSI, a cloud-based multimodal dialogue technology, are significantly correlated with recognition of vocal affect assessed via the Diagnostic Analysis of Nonverbal Accuracy (DANVA) task. We also found correlations between jaw kinematics and motor speed of the dominant hand as assessed by a standardised finger

tapping task.

The DANVA receptive paralanguage subtasks assess the participants' ability to identify vocal emotional communication, with higher scores indicating competence in recognition of vocal affect. Participants who were more competent in identifying the emotion from adult and child voice recordings produced emotional speech (monosyllabic "oh") with a higher fundamental frequency. A previous study has found an increased pitch range in high-functioning autism [9]. Participants with better DANVA CPSS scores also produced louder sentence-length happy emotional speech. Increased articulation loudness is indeed associated with happy emotional speech [39]. Our findings may suggest that high-performing participants on the vocal affect recognition task may possess greater ability to modulate the acoustic properties of emotional speech whereas more impaired participants are unable to do so. This may suggest that there is a relationship between the recognition and production of the acoustic properties that are relevant for signifying emotion. Indeed, previous research has identified a relationship between basic auditory processing and vocal affect recognition [40, 41, 42].

The correlations between jaw movement and motor dexterity of the dominant hand may be a manifestation of global motor impairments in individuals with ASD [43, 44]. Motor deficits in ASD have been attributed to abnormal sensorimotor integration [45, 46] due to increased sensory noise [47] or dysfunctional representation of internal models of action [48]. A previous study has found no differences in finger-tapping accuracy between children with ASD and typically-developing children but there was increased variability in temporal processing

parameters in the ASD group [49]. Correlations between jaw movement and finger tapping suggest global altered temporal processing in ASD which may also explain poor cross-modal coordination during emotional speech production [20]. Indeed, prior work has suggested a coupling between speech motor coordination and fine motor skill systems in ASD [50].

Our preliminary results indicate that speech and facial metrics extracted from our APT task administered using Modality.AI's NEMSI platform can tap into deficits of various physiological subsystems in children with ASD. The current study was conducted in a laboratory setting but the conversational platform can also be accessed remotely via a weblink [26]. Thus, our technology provides the opportunity for clinicians and researchers to monitor children with developmental disorders in a non-clinical, non-laboratory setting. Since deficits in emotion recognition have been observed cross-culturally in individuals with ASD [51], our technology can also be leveraged to compile a global corpus of speech and facial metrics during affect recognition and production in ASD. A conversational task where participants are prompted to produce a monosyllabic “oh” may also prove useful in a low-functioning population. Furthermore, it has been shown that children with ASD can be trained to produce qualitatively-improved emotional facial expressions [52]. Remote monitoring may also facilitate a longitudinal, quantitative analysis of such training paradigms.

Future studies would involve collecting data from larger cohorts of children with ASD along with data from typically-developing children to examine whether differences in vocal and facial affect production between the two groups can be captured through automatically-extracted metrics.

6. Acknowledgements

This study was supported by the National Institutes of Health grant K23 DC016637 and Autism Speaks grant 11637 awarded to Carly Demopoulos.

7. References

- [1] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.
- [2] M. J. Maenner, K. A. Shaw, J. Baio *et al.*, “Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2016,” *MMWR Surveillance Summaries*, vol. 69, no. 4, p. 1, 2020.
- [3] A. S. Winter, C. Fountain, K. Cheslack-Postava, and P. S. Bearman, “The social patterning of autism diagnoses reversed in california between 1992 and 2018,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 295–30 302, 2020.
- [4] M. K. Khowaja, A. P. Hazzard, and D. L. Robins, “Sociodemographic barriers to early detection of autism: screening and evaluation using the m-chat, m-chat-r, and follow-up,” *Journal of autism and developmental disorders*, vol. 45, no. 6, pp. 1797–1808, 2015.
- [5] R. B. Grossman, L. R. Edelson, and H. Tager-Flusberg, “Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism,” *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 3, pp. 1035–1044, 2013.
- [6] L. M. Lozier, J. W. Vanmeter, and A. A. Marsh, “Impairments in facial affect recognition associated with autism spectrum disorders: a meta-analysis,” *Development and psychopathology*, vol. 26, no. 4, pp. 933–945, 2014.
- [7] S. Griffiths, C. Jarrold, I. S. Penton-Voak, A. T. Woods, A. L. Skinner, and M. R. Munafò, “Impaired recognition of basic emotions from facial expressions in young people with autism spectrum disorder: Assessing the importance of expression intensity,” *Journal of autism and developmental disorders*, vol. 49, no. 7, pp. 2768–2778, 2019.
- [8] S. M. Fosnot and S. Jun, “Prosodic characteristics in children with stuttering or autism during reading and imitation,” in *Proceedings of the 14th international congress of phonetic sciences*, 1999, pp. 1925–1928.
- [9] A. Nadig and H. Shaw, “Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners,” *Journal of Autism and Developmental Disorders*, vol. 42, no. 4, pp. 499–511, 2012.
- [10] D. J. Hubbard, D. J. Faso, P. F. Assmann, and N. J. Sasson, “Production and perception of emotional prosody by adults with autism spectrum disorder,” *Autism Research*, vol. 10, no. 12, pp. 1991–2001, 2017.
- [11] K. A. Loveland, B. Tunali-Kotoski, D. A. Pearson, K. A. Brelsford, J. Ortegon, and R. Chen, “Imitation and expression of facial affect in autism,” *Development and Psychopathology*, vol. 6, no. 3, pp. 433–444, 1994.
- [12] D. J. Faso, N. J. Sasson, and A. E. Pinkham, “Evaluating posed and evoked facial expressions of emotion from adults with autism spectrum disorder,” *Journal of autism and developmental disorders*, vol. 45, no. 1, pp. 75–89, 2015.
- [13] R. Brewer, F. Biotti, C. Catmur, C. Press, F. Happé, R. Cook, and G. Bird, “Can neurotypical individuals read autistic facial expressions? atypical production of emotional facial expressions in autism spectrum disorders,” *Autism Research*, vol. 9, no. 2, pp. 262–271, 2016.
- [14] F. B. Pokorny, B. Schuller, P. B. Marschik, R. Brueckner, P. Nyström, N. Cummins, S. Bölte, C. Einspieler, and T. Falck-Ytter, “Earlier identification of children with autism spectrum disorder: An automatic vocalisation-based approach,” in *INTERSPEECH*, 2017, pp. 309–313.
- [15] E. Zane, Z. Yang, L. Pozzan, T. Guha, S. Narayanan, and R. B. Grossman, “Motion-capture patterns of voluntarily mimicked dynamic facial expressions in children and adolescents with and without asd,” *Journal of autism and developmental disorders*, vol. 49, no. 3, pp. 1062–1079, 2019.
- [16] S. Cho, M. Liberman, N. Ryant, M. Cola, R. T. Schultz, and J. Parish-Morris, “Automatic detection of autism spectrum disorder in children using acoustic and text features from brief natural conversations,” in *INTERSPEECH*, 2019, pp. 2513–2517.
- [17] M. Li, D. Tang, J. Zeng, T. Zhou, H. Zhu, B. Chen, and X. Zou, “An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder,” *Computer Speech & Language*, vol. 56, pp. 80–94, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0885230817303601>
- [18] M. Leo, P. Carcagnì, C. Distante, P. Spagnolo, P. L. Mazzeo, A. C. Rosato, S. Petrocchi, C. Pellegrino, A. Levante, F. De Lumè *et al.*, “Computational assessment of facial expression production in asd children,” *Sensors*, vol. 18, no. 11, p. 3993, 2018.
- [19] A. Bangerter, M. Chatterjee, J. Manfredonia, N. V. Manyakov, S. Ness, M. A. Boice, A. Skalkin, M. S. Goodwin, G. Dawson, R. Hendren *et al.*, “Automated recognition of spontaneous facial expression in individuals with autism spectrum disorder: parsing response variability,” *Molecular autism*, vol. 11, pp. 1–15, 2020.
- [20] T. Sorensen, E. Zane, T. Feng, S. Narayanan, and R. Grossman, “Cross-modal coordination of face-directed gaze and emotional speech production in school-aged children and adolescents with asd,” *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.
- [21] M. D. Samad, N. Diawara, J. L. Bobzien, J. W. Harrington, M. A. Witherow, and K. M. Iftekharuddin, “A feasibility study of autism behavioral markers in spontaneous facial, visual, and hand movement response data,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 353–361, 2017.
- [22] C.-P. Chen, X.-H. Tseng, S. S.-F. Gau, and C.-C. Lee, “Computing multimodal dyadic behaviors during spontaneous diagnosis interviews toward automatic categorization of autism spectrum disorder,” in *INTERSPEECH*, 2017, pp. 2361–2365.
- [23] J. Chen, M. Liao, G. Wang, and C. Chen, “An intelligent multimodal framework for identifying children with autism spectrum disorder,” *International Journal of Applied Mathematics and Computer Science*, vol. 30, no. 3, pp. 435–448, 2020.

[24] D. Suendermann-Oeft, A. Robinson, A. Cornish, D. Habberstad, D. Pautler, D. Schnelle-Walka, F. Haller, J. Liscombe, M. Neumann, M. Merrill *et al.*, “Nemsi: A multimodal dialog system for screening of neurological or mental conditions,” in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 2019, pp. 245–247.

[25] M. Neumann, O. Roesler, D. Suendermann-Oeft, and V. Ramanarayanan, “On the utility of audiovisual dialog technologies and signal analytics for real-time remote monitoring of depression biomarkers,” in *Proceedings of the First Workshop on Natural Language Processing for Medical Conversations*, 2020, pp. 47–52.

[26] V. Ramanarayanan, O. Roesler, M. Neumann, D. Pautler, D. Habberstad, A. Cornish, H. Kothare, V. Murali, J. Liscombe, D. Schnelle-Walka *et al.*, “Toward remote patient monitoring of speech, video, cognitive and respiratory biomarkers using multimodal dialog technology,” *Proc. Interspeech 2020*, pp. 492–493, 2020.

[27] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. Dilavore, A. Pickles, and M. Rutter, “The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism,” *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[28] C. Lord, M. Rutter, and A. Le Couteur, “Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders,” *Journal of autism and developmental disorders*, vol. 24, no. 5, pp. 659–685, 1994.

[29] C. Reynolds and R. Kamphaus, “Behavior assessment system for children—third edition (basc-3),” *Bloomington, MN: Pearson*, 2015.

[30] S. Nowicki and M. P. Duke, “Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale,” *Journal of Nonverbal behavior*, vol. 18, no. 1, pp. 9–35, 1994.

[31] K. M. Baum and S. Nowicki, “Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity,” *Journal of Nonverbal Behavior*, vol. 22, no. 2, pp. 89–107, 1998.

[32] B. L. Brooks, E. M. Sherman, and E. Strauss, “Nepsy-ii: a developmental neuropsychological assessment,” *Child Neuropsychology*, vol. 16, no. 1, pp. 80–101, 2009.

[33] R. M. Reitan and D. Wolfson, *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Reitan Neuropsychology, 1985, vol. 4.

[34] R. Reitan and D. Wolfson, “The halstead-reitan cognitive test battery: Theory and clinical interpretation,” 1993.

[35] R. K. Heaton, I. Grant, and C. G. Matthews, “Differences in neuropsychological test performance associated with age, education, and sex,” *Neuropsychological assessment of neuropsychiatric disorders*, vol. 1, pp. 100–120, 1986.

[36] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[37] E. H. Wiig, W. A. Secord, and E. Semel, *Clinical evaluation of language fundamentals: CELF-5*. Pearson, 2013.

[38] D. Wechsler, *WISC-V: Technical and interpretive manual*. NCS Pearson, Incorporated, 2014.

[39] I. R. Murray and J. L. Arnott, “Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion,” *The Journal of the Acoustical Society of America*, vol. 93, no. 2, pp. 1097–1108, 1993.

[40] M. D. Lerner, J. C. McPartland, and J. P. Morris, “Multimodal emotion processing in autism spectrum disorders: an event-related potential study,” *Developmental cognitive neuroscience*, vol. 3, pp. 11–21, 2013.

[41] C. Demopoulos, J. Hopkins, B. E. Kopald, K. Paulson, L. Doyle, W. E. Andrews, and J. D. Lewine, “Deficits in auditory processing contribute to impairments in vocal affect recognition in autism spectrum disorders: A meg study,” *Neuropsychology*, vol. 29, no. 6, p. 895, 2015.

[42] C. Demopoulos and J. D. Lewine, “Audiotmetric profiles in autism spectrum disorders: Does subclinical hearing loss impact communication?” *Autism Research*, vol. 9, no. 1, pp. 107–120, 2016.

[43] E. M. Jansiewicz, M. C. Goldberg, C. J. Newschaffer, M. B. Denckla, R. Landa, and S. H. Mostofsky, “Motor signs distinguish children with high functioning autism and asperger’s syndrome from controls,” *Journal of autism and developmental disorders*, vol. 36, no. 5, pp. 613–621, 2006.

[44] X. Ming, M. Brimacombe, and G. C. Wagner, “Prevalence of motor impairment in autism spectrum disorders,” *Brain and Development*, vol. 29, no. 9, pp. 565–570, 2007.

[45] J. K. Kern, M. H. Trivedi, C. R. Garver, B. D. Grannemann, A. A. Andrews, J. S. Savla, D. G. Johnson, J. A. Mehta, and J. L. Schroeder, “The pattern of sensory processing abnormalities in autism,” *Autism*, vol. 10, no. 5, pp. 480–494, 2006.

[46] C. Demopoulos, H. Kothare, D. Mizuri, J. Henderson-Sabes, B. Fregeau, J. Tjernagel, J. F. Houde, E. H. Sherr, and S. S. Nagarajan, “Abnormal speech motor control in individuals with 16p11.2 deletions,” *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[47] E. Gowen and A. Hamilton, “Motor abilities in autism: a review using a computational context,” *Journal of autism and developmental disorders*, vol. 43, no. 2, pp. 323–344, 2013.

[48] C. C. Haswell, J. Izawa, L. R. Dowell, S. H. Mostofsky, and R. Shadmehr, “Representation of internal models of action in the autistic brain,” *Nature neuroscience*, vol. 12, no. 8, pp. 970–972, 2009.

[49] C. Morimoto, E. Hida, K. Shima, and H. Okamura, “Temporal processing instability with millisecond accuracy is a cardinal feature of sensorimotor impairments in autism spectrum disorder: analysis using the synchronized finger-tapping task,” *Journal of autism and developmental disorders*, vol. 48, no. 2, pp. 351–360, 2018.

[50] T. Talkar, J. R. Williamson, D. J. Hannon, H. M. Rao, S. Yuditskaya, K. T. Claypool, D. Sturim, L. Nowinski, H. Saro, C. Stamm, M. Mody, C. J. McDougle, and T. F. Quatieri, “Assessment of speech and fine motor coordination in children with autism spectrum disorder,” *IEEE Access*, vol. 8, pp. 127535–127545, 2020.

[51] S. Fridenson-Hayo, S. Berggren, A. Lassalle, S. Tal, D. Pigat, S. Bölte, S. Baron-Cohen, and O. Golan, “Basic and complex emotion recognition in children with autism: cross-cultural findings,” *Molecular autism*, vol. 7, no. 1, pp. 1–11, 2016.

[52] I. Gordon, M. D. Pierce, M. S. Bartlett, and J. W. Tanaka, “Training facial expression production in children on the autism spectrum,” *Journal of autism and developmental disorders*, vol. 44, no. 10, pp. 2486–2498, 2014.