Speech Rate and Salient Syllables Position in Spontaneous Speech of Children with Autism Spectrum Disorder

Valentina Saccone
University of Florence
valentina.saccone@unifi.it

Abstract

The study employs a semi-automatic approach to analyze speech rate in spoken Italian, aiming to identify acoustic parameters associated with perceptual atypicality in the speech of children diagnosed with Autism Spectrum Disorder (ASD). The research focuses on a dataset comprising recordings of semi-spontaneous interactions, in comparison with interviews of Typically Developing (TD) children. A detailed examination of speech rate variability is conducted, progressing from assessing overall speech rate in conversation to the analysis of individual utterances. Furthermore, salient syllables within utterances are identified using an automatic procedure through the Salient Detector Praat script and analyzed for stress position. The study highlights specific speech style, including rapid-telegraphic and reading-performed speech. Additionally, it reveals a higher speech rate with the increasing length of utterance when <10 syllables; conversely, a speech rate diminishing in 20-25 syllables utterances, suggesting potential difficulty in producing longer utterances associated with increased cognitive load.

Keywords: Autism Spectrum Disorder, speech rate, salience

1. Introduction

Autism Spectrum Disorders (ASD) encompass a diverse range of neurodevelopmental conditions characterized by impairments in social interaction, language, and communication, as well as restricted interests and repetitive behaviors, as outlined in the DSM-5 (American Psychiatric Association, 2013). While the clinical presentation of ASD is highly varied, certain atypical communicative-linguistic features are frequently observed, though they lack definitive diagnostic significance when considered independently. Among these patterns, prosody emerges as a linguistic domain of particular interest. Altered prosodic features include aspects such as rhythm, affective expression, pragmatic functions, and syntactic structures (McCann et al., 2007; Eigsti et al., 2011; McAlpine et al., 2014; Fusaroli et al., 2017).

Notably, durational variability and speech rate contribute to the perception of atypical speech patterns in autism. Previous literature suggests that individuals with ASD exhibit a slower speech rate (Patel et al., 2020), and variations in durational cues associated with prosodic phrasing and stress (Fosnot and Jun, 1999; Paul et al., 2008). Studies have highlighted phenomena such as the elongation of stressed syllables (Paul et al., 2008; Byrd and Salzman, 2003; Patel et al, 2020), there is a scarcity of research within the context of naturalistic speech.

This analysis aims to advance our understanding of speech rate and rhythm in spontaneous speech, laying the groundwork for more comprehensive investigations. It is part of a broader inquiry into various aspects of prosody, including intonation modulation, intensity variation, speech rhythm, information structure (Chiti et al., forthcoming) and the interplay between prosody and co-speech gestures (Saccone et al., 2023).

Previous research on typically developing populations has demonstrated a correlation between increased cognitive load and a reduced speech rate (Griffin and Williams, 1987; Huttunen et al., 2011), as well as an association between longer utterances and heightened speech rate, often quantified in terms of phonological word count (Darling-White and Banks, 2021). These findings serve to frame the present study.

Our hypothesis is that the variability of the speech rate inside the utterance and the position of salient syllables, more than durations, allow recognizing different atypical trend.

2. Dataset

The dataset under analysis comprises audio recordings of 9 children[1] diagnosed with ASD (7 males and 2 females, aged between 8 and 12 years) who exhibit intelligible language skills and Italian as their native tongue[2]. All participants are from the Pistoia area in Tuscany. The recordings were captured during therapy sessions conducted in a semi-spontaneous set, encompassing small talks in the form of interviews about school, hobbies, and holidays,

[1] Children are referred to here with an ID composed of a numeral and a letter indicating gender (e.g., 1M, 3F).
[2] For privacy reasons, it was not possible to collect the test results that determined the diagnosis of autism in the children of the sample. To ensure consistency in language levels, the selection was carried out by therapists, psychologists, speech therapists, and educators who are responsible for the children, all of whom belong to the same non-profit organization for therapeutic treatments.

Proceedings of the RaPID-5 @LREC-COLING 2024, pages 9–15

the narration of the story "Frog, Where are you?" (Mayer, 1969), and board game sessions. The analysis focuses on a subset of 20[3] utterances per speaker, totaling 180 utterances.
For comparison, data from a sample of 180 utterances by typically developing children (TD) were processed. The audio recordings were collected during interviews conducted in elementary and middle schools located in Florence, Prato, and Pistoia (all within Tuscany), ensuring demographic alignment with the ASD group in terms of regional language and age (8-12). The interview topics included class trips, sports, music, lunch breaks, diversity, and inclusion. The TD group consisted of 50 children (29 males and 21 females), each contributing a maximum of four utterances. This variation was deliberately chosen to maximize diversity within the group, reflecting preference for representing speech variability[4].

3. Methods

All speech samples were analyzed using Praat software (Boersma and Weenink, 2024).The audio was transcribed and segmented into prosodic units and utterances, based on the Language into Act Theory (Cresti, 2000; Moneglia, 2005). Additionally, automatic procedure in the Praat's Vocal Toolkit script was utilized to mark syllabic regions, which were then manually revised.

To examine durational variation of the speech, the analysis proceeded through three steps:

i. Calculation of mean speech rate (syll/sec) for each speaker in relation to the Medium Length of Utterance (MLU);

ii. Measurement of speech rate variation;

iii. Application of the Salience Detector Praat script to identify salient syllables within utterances, subsequently annotated by stress position (pre-stressed, stressed, or post-stressed syllables).

To ensure comparability, exclusion criteria were applied to maintain consistency in utterance length and illocutionary type (only assertion, only one for utterance[5]), while avoiding interruptions or overlaps with the interviewer. The selection ensures that the difference in discourse type (e.g., interview vs. narration) has the least possible impact on the sampling regarding informational and prosodic phenomena. Utterances were divided into four length categories based on syllable count (not words) to reduce variability: A) 3-4 syllables; B) 8-10 syllables; C) 14-16 syllables; D) 20-25 syllables[6].

The script used in step iii. was developed by Plinio Barbosa (Barbosa et al., 2019; Barbosa, 2022) employing a Table of Real with mean duration values for each sound of the Italian language, specifically focusing on Tuscan regional Italian[7]. This script facilitates language-independent automatic detection of syllable-sized normalized duration peaks to study prominence and boundary, identifying prosody-related acoustic salience in each utterance[8].

Finally, to complement the analysis with perceptual evidence, a perception test was conducted with a panel of 20 native Italian speakers using a sample of 40 utterances (25 from the ASD group and 15 from the TD group). Utterances were randomly selected from the samples and presented randomly within the test. The panel was then asked to identify whether each sounded like spontaneous and whether any anomalies were perceived in the prosody of the recordings. Thus, the utterances were judged individually without knowledge of the speakers' identities.

4. Analysis

4.1 Speech rate and MLU

Initially, in the ASD group, we quantified the speech rate as the number of syllables per second (syll/sec) and observed its correlation with the Medium Length of Utterance (MLU) measured in words[9]. The findings are depicted in Figure 1.

[3] The minimum of 20 utterances is consistent with previous research, such as that conducted by Patel et al. (2020). The chosen utterances are mainly from the interview section of the recordings.

[4] Preliminary tests on the sample did not reveal significant variations based on gender in the parameters we are extracting.

[5] Following the Language into Act Theory, it is possible to identify prosodically terminated sequences which host more than one illocution (Moneglia and Raso, 2014; Panunzi and Saccone, 2018; Saccone, 2022).

[6] Group D encompasses a broader range of length of utterance than the other categories. This was a necessary choice due to the scarcity of long utterances for some of the children in the ASD group.

[7] The Table of Real for the Italian language is the result of collaborative work by the author in conjunction with Marcelo Vieira (McGill University) and Plinio Barbosa (University of Campinas).

[8] The local maxima of smoothed z-scores are selected as salient segments, delineating the end of corresponding stress groups.

[9] To measure the MLU, the dataset is expanded to include the total recordings (nearly 15 minutes per child).

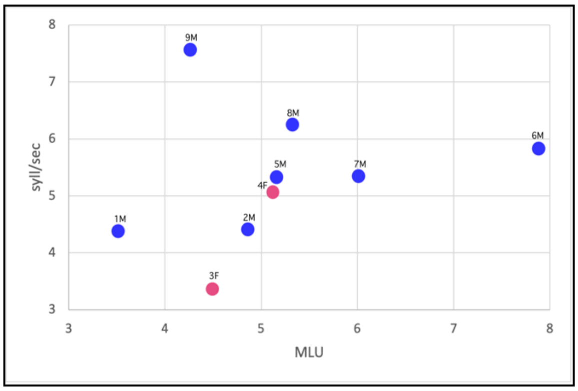

Figure 1: Mean speech rate - ASD group (Chiti et al., forthcoming).

The trend observed in the ASD group aligns with expectation from TD children's literature (see, among others, Darling-White and Banks, 2021): an increase in mean speech rate correlates with a higher MLU. Notably, participant 9M deviates from this general trend, exhibiting short utterances and fast speech, reflecting a rapid and telegraphic speech style.

Random sample testing reveals variability in speech rate within utterances, indicating an atypical speed perception not fully captured by average calculations per speaker. Steps ii. and iii. delve deeper into this aspect.

4.2 Speech rate variation

Furthermore, we examined speech rate for each utterance in the chosen categories (A, B, C, and D) correlated with utterance length, uncovering deviations from the TD pattern. Figure 2 illustrates the mean speech rate (syll/sec) across different utterance lengths for each participant of the ASD group and aggregates it for the TD group. Data is divided into four columns corresponding to utterance length, with darker colors indicating longer utterances.

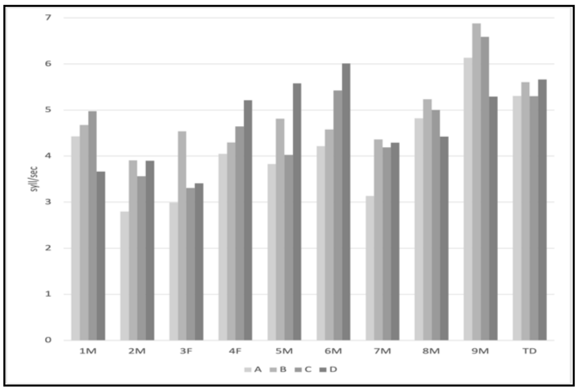

Figure 2: Mean speech rate categorized by length of utterance.

The dataset as a whole shows an increase in speech rate from category A to B. Only participants 4F and 6M exhibit a consistently increasing trend across all the categories, while in 2M, 3F, 5M, and 7M, the increase is interrupted by a peak in category B, as well as for the TD group. Interestingly, some participants of the ASD group (1M, 8M, and 9M) display the lowest values in column D (the darker), suggesting potential difficulty in producing longer utterances associated with increased cognitive load, as proposed by Griffin and Williams (1987) and Huttunen et al. (2011).

Examining the standard deviation of the mean speech rates reveals variability among participants. Categorizing results into A, B, C, and D sheds light on the relationship between speech rate variability and utterance length (Figure 3).

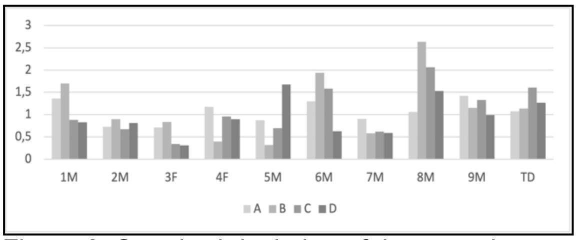

Figure 3: Standard deviation of the speech rate categorized by length of utterance.

Even for this parameter, the behavior of the ASD group does not exhibit a common trend. In 2M, 3F, and 7M, values never exceed 1, indicating consistency in the samples. Conversely, peaks in variability indicate dispersed data within the utterance, particularly evident for participants 1M, 6M, 8M, and 9M, where two or more columns show a deviation >1. Notably, participant 8M exhibits high variance across all columns, indicating pronounced variability in speech rate irrespective of utterance length[10].

Data for the TD group aggregates measurements for 50 different speakers; thus, high variability is expected. Consequently, we cannot directly compare TD values with the participants of ASD group values. Nevertheless, they fall below the ASD average for columns A and B. Even though further investigation is necessary, this suggests that the ASD values are high. Additionally, in TD, the ratio between the columns can still be interpreted, revealing a trend of broader variation in speech rate in columns C and D (14-25 syllables utterances) compared to A and B (3-10 syllables utterances), a trend not repeated in the ASD group.

4.3 Salience position inside the utterance

Speech rhythm variability is associated with vowel lengthening and prosodic prominences, prompting us to use a script to identify salient

[10] This result is to be connected to a peculiar characteristic of 8M not covered here, that is the high presence of pauses in his speech.

syllables within utterances. Salience position typically corresponds to stress presence and to pragmatic prominences, with unexpected saliences possibly indicating non-pragmatic effects. Each utterance's script output lists syllables with duration measurements and z-score deviations, highlighting salient syllables (often one for A, two for B, and varying numbers for C and D). Elongations and pre-pausal vowel lengthening, associated with specific pragmatic cues, are separately tagged and not discussed here[11].

Annotating salient syllables by stress position reveals diverse speech trends. In TD group expectations, stress-saliences predominate, alongside post-stress-saliences, with fewer pre-stress saliences due to the spontaneous nature of speech. Results are shown in Figure 4.

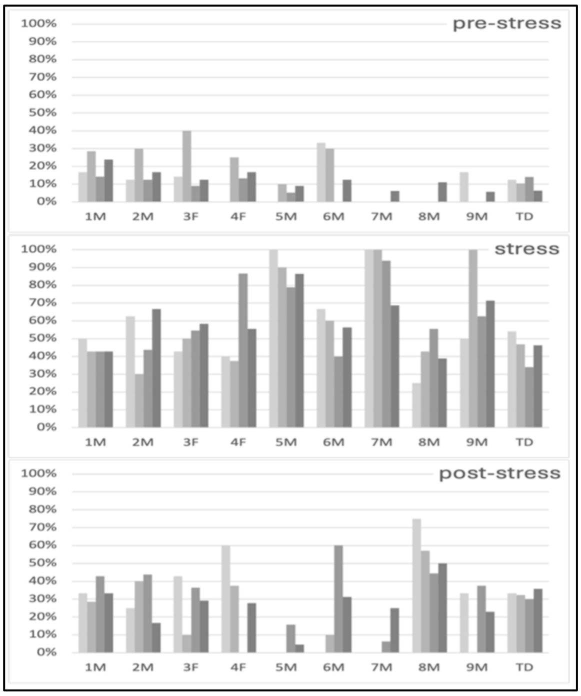

Figure 4: Salience position.

In the TD group, the length of utterance does not affect speech behavior; as expected, half of the saliences occur in stress-position and nearly 35% in post-stress position. Although pre-stress saliences are present in a lower percentage, they still contribute to the overall pattern. As already noted, the TD group comprises various speakers, making the homogeneous pattern more relevant.

In contrast, the ASD group exhibits diverse deviations from this trend. Analysis segmented by utterance length reveals varied behaviors, indicating a lack of uniformity in rhythmic patterns across the speech flow. Participant 8M displays a unique behavior, characterized by a prominent presence of post-stress saliences. Moreover, participants 5M and 7M demonstrate a distinctive trend, with stress-position saliences reaching peaks of 100%, while pre- and post-stress saliences are rarely observed. This phenomenon diminishes the natural spontaneity of speech, resembling a style akin to reading aloud. Comparative studies between spontaneous and read speech corroborate these findings, suggesting increased variation in the position of salient syllables during spontaneous speech (Nakamura et al., 2008; Furui, 2003).

To account for these variations, a K-means cluster analysis was conducted with k=4 (chosen via the elbow curve). Figure 5 illustrates the resulting clusters.

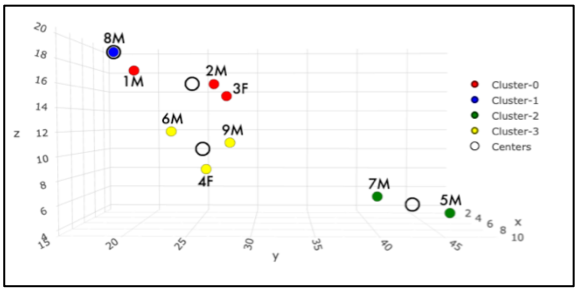

Figure 5: K-means cluster analysis.

The clusters are divided as follow: Cluster-0 (in red) encompasses participants 1M, 2M, and 3F; Cluster-1 (in blue) comprises solely participant 8M; Cluster-2 (in green) connects participants 5M and 7M; Lastly, Cluster-3 (in yellow) consists of participants 4F, 6M, and 9M.

What distinguishes 8M (Cluster-1) is a diminished speech rate in longer utterances and above all, high variance in speech regardless of utterance length. Additionally, the cluster analysis highlights systematic differences in the position of salient syllables compared to the other participants.

Cluster-2 is positioned furthest from the other clusters, which is consistent with the evaluations proposed in step ii.

4.4 Perception test

The perception test aimed to validate the atypia identified in the acoustic analysis of speech rate and salience position. Respondents were guided by various prosodic factors besides speech rate, as intonation, and intensity, providing insights into communicative and linguistic significance. Results are depicted in Figure 6 ordered by decreasing value of spontaneity (with TD highlighted in yellow).

[11] At the pre-pausal boundaries, prosodic events, such as pre-pausal lengthening, have been observed in studies by Sorianello (1994); Rao (2010); Kentner et al. (2023), among others.

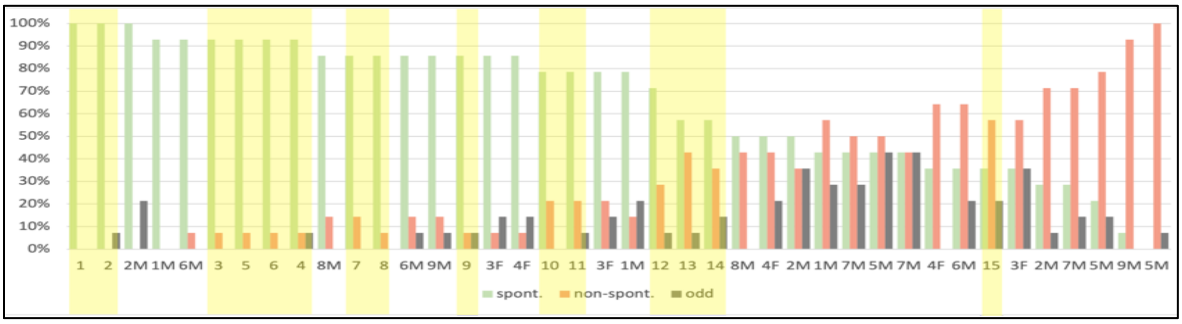

Figure 6: Perception test results.

On x axis, the ID of the speaker for the ASD group and consecutive numbers for TD utterances[12].

The identification of the TD utterances as spontaneous speech clearly emerges (green columns often >80%), with a few instances of sense of "oddity" perceive (signaled by ≤21% of the panel). Conversely, responses for the ASD group varied widely, with a relevant portion (≥ 40% of the panel for half of the sample) being classified as non-spontaneous, and reaching levels of 70-100% for 5M, 7M , and 9M. Instances of indicating "odd" (27 cases) were more likely to occur for the ASD group (19 cases) and in conjunction with non-spontaneous evaluation (25 cases).

Participants 2M, 5M, 7M, and 9M displayed the highest rates of non-spontaneous categorization (>70%). 5M and 7M also recording the highest rates of "odd" (>40%).

Non-spontaneous perceptions were primary associated with a sense of read speech or preparedness akin to performed speech.

5. Discussion

This study delves into the speech rate, examining its variability from overall speech rate in conversation to the analysis of individual utterances. While previous studies on Italian-speaking individuals with ASD have focused on quantifying vocalic quantity (Fantini et al., 2023), this research introduces novel aspects by examining the variation in internal utterance lengths and the placement of salient syllables and using ecological settings for recording.

As expected from the spectrum heterogeneity, the autistic children of our sample do not exhibit a common trend. Nonetheless, our studies highlight variation in specific parameters compared to typical developing individuals.

Children in ASD dataset show a lack of uniformity in rhythmic patterns across the speech flow (particularly for participant 8M), and an increase in speech rate in 2 to 10 syllables utterances (A and B categories). On the other hand, low values recorded for the speech rate in 20-25 syllables utterances (category D) suggest potential difficulty in producing longer utterances associated with increased cognitive load. Being equal the number of syllables, future research should investigate the interplay of speech rate with the number of prosodic-pragmatic unit inside the utterance.

Additionally, according to the perception test, the speech of our ASD sample results mostly as non-spontaneous to the listener.

The study of speech rate and saliences also highlighted specific speech style, as for the rapid and telegraphic tone of participant 9M (with short and fast utterances), and the reading-performed one of participants 5M and 7M (with stress-position saliences reaching peaks of 100%).

To enhance the robustness of our findings, future research should expand the study to a larger population, while also collecting more specific metadata for each participant (such as scores from autism assessment tests, verbal and non-verbal IQ). It is important to note the time-consuming nature of syllable-by-syllable transcription, which has limitations and could be optimized using automated procedures.

Given the differences between the two groups, the TD group should not be understood as a control group for the ASD group presented here, but rather as a comparison with a neurotypical trend as varied as possible, in order to emphasize the points of variation within the ASD sample. Future research may focus on balanced samples, which were not available to us at the time of this study.

Moreover, integrating the analysis of pausing and enhancing the role of elongations and pre-pausal vowel lengthening with specific pragmatic

[12] For the sake of simplicity in representation, consecutive numbers have been assigned to the TD group, but this does not imply that the utterances were selected in an ordered manner. As outlined in the methods, the selection procedure was random, with each label corresponding to one random utterance. Moreover, the difference in the number of utterances between the two groups does not affect the results, which are not considered in absolute numbers.

effects could provide valuable insights into speech rhythm and its variations.

As pointed out in Patel et al. (2020), while listeners may discern clear distinctions in prosody, these may not always be reflected in basic or easily measurable acoustic properties. For instance, the simple measurement of average speech rate may not fully capture the atypical nature of speech in individuals with ASD. Nonetheless, the durational characteristics of speech remain relevant for clinicians, as therapeutic interventions targeting rhythm and timing stability have shown promising improvements across both speech and motor domains in ASD (Franich et al., 2020).

6. Ethical statement

The audio recordings of the ASD group were collected during a Speech Therapy BA thesis, and informed consent, signed by parents or legal guardians, was obtained from each participant. The audio recordings of the TD group were selected with the assistance of schoolteachers from publicly accessible materials on school platforms and public channels.

7. Bibliographical References

APA-American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, DSM-5*. Arlington.

Barbosa, P. A., Camargo, Z. A. and Madureira, S. (2019). Acoustic-based tools and scripts for the automatic analysis of speech in clinical and non-clinical settings. In H. A. Patil, M. Kulshreshtha, & A. Neustein (Eds.), *Signal and acoustic modeling for speech and communication disorders*. Berlin, Boston: De Gruyter, pp. 69–86.

Barbosa, P. (2022). *Manual de Prosódia Experimental*. Editora da Abralin. 10.25189/9788568990230.

Boersma, P. and Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.06, retrieved 25 February 2024 from http://www.praat.org/

Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2): 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2

Chiti, V., Saccone, V. and Panunzi, A. (forthcoming) L'alterazione degli aspetti pragmatici e prosodici nel Disturbo dello Spettro Autistico: analisi del parlato di bambini in età scolare. *Studi AISV 9*.

Cresti, E. (2000). *Corpus di italiano parlato*. Firenze: Accademia della Crusca.

Darling-White, M. and Banks, S. W. (2021). Speech rate varies with sentence length in typically developing children. *Journal of Speech, Language, and Hearing Research*, 64(6S): 2385–2391. https://doi.org/10.1044/2020_JSLHR-20-00276

Eigsti, I., Schuh, J., Mencl, E., Schultz, R. and Paul, R. (2011). The neural underpinnings of prosody in autism. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 18(6): 600-617.

Fantini, V., Gagliardi G., Maffia, M. and Pettorino, M. (2023) Il parlato di bambini con Disturbo dello Spettro Autistico: un'analisi ritmica. paper for the conference "La comunicazione parlata. I 20 anni del GSCP". RomE, 8-10 giugno.

Fosnot, S. M. and Jun, S. A. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. *14th International Congress of Phonetic Sciences*: 103–115.

Franich, K., Wong, H. Y., Alan, C. L. and To, C. K. (2020). Temporal coordination and prosodic structure in autism spectrum disorder: Timing across speech and non-speech motor domains. *Journal of Autism and Developmental Disorders*, 51(8): 2929–2949.

Furui, S. (2003). Recent advances in spontaneous speech recognition and understanding. *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR), Tokyo*: 1–6.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M. and Gaigg, SB. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, 10: 384–407. 10.1002/aur.1678

Griffin, G. R. and Williams, C. E. (1987). The effects of different levels of task complexity on three vocal measures. *Aviation, Space and Environmental Medicine*, 58(12): 1165–1170.

Huttunen, K. H., Keränen, H. I., Pääkkönen, R. J., Päivikki Eskelinen-Rönkä, R. and Leino, T. K. (2011). Effect of cognitive load on articulation rate and formant frequencies during simulator flights. *The Journal of the Acoustical Society of America*, 129(3): 1580–1593. https://doi.org/10.1121/1.3543948.

Kentner, G., Franz, I., Knoop, C. A. and Menninghaus, M. (2023). The final lengthening of pre-boundary syllables turns into final shortening as boundary strength levels increase. *Journal of Phonetics*, 97. https://doi.org/10.1016/j.wocn.2023.101225.

Mayer, M. (1969). *Frog, where are you?*. New York: Dial Press.

McAlpine, A., Plexico, LW., Plumb, AM. and Cleary, J. (2014). Prosody in Young Verbal Children With Autism Spectrum Disorder. *Contemporary Issues in Communication Science and Disorders*, 41: 120-132.

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A. and Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *Int. J.*

Lang. Commun. Disord., 42: 682–702. 10.1080/13682820601170102.
Moneglia, M. (2005). The C-ORAL-ROM resource" in *C-ORAL-ROM.* In E. Cresti, & M. Moneglia (Eds.), *Integrated reference corpora for spoken romance languages*. Amsterdam: John Benjamins, pp. 209–256.
Moneglia M. and Raso T. (2014), Notes on Language into Act Theory (L-AcT). In T. Raso, & H. Mello (Eds.), *Spoken corpora and linguistic studies*. Amsterdam, John Benjamins, pp. 468-495.
Nakamura, M., Iwano, K., Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance, *Computer Speech & Language*, 22(2): 171-184, https://doi.org/10.1016/j.csl.2007.07.003.
Panunzi, A. and Saccone, V. (2018). Complex Illocutive Units in L-AcT: an analysis of non-terminal prosodic breaks of Bound and Multiple Comments. *Revista de Estudos da Linguagem*, Belo Horizonte, 26(4): 1647-1674.
Patel, S. P., Nayar, K., Martin, G. E., Franich, K., Crawford, S., Diehl, J. J. and Losh, M. (2020). An acoustic characterization of prosodic differences in autism spectrum disorder and first-degree relatives. *Journal of Autism and Developmental Disorders*, 50(8): 3032-3045 https:// doi.org/10.1007/s10803-020-04392-9
Paul, R., Bianchi, N., Augustyn, A., Klin, A. and Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 2(1): 110–124. https://doi.org/10.1016/j.rasd.2007.04.001
Rao, R. (2010). Final lengthening and pause duration in three dialects of Spanish. *Selected proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, Cascadilla Proceedings Project, Somerville, MA, 69-82
Saccone, V. (2022). *Le unità del parlato e dello scritto mediato dal computer a confronto. La dimensione testuale della comunicazione spontanea.* Edizioni dell'Orso, Alessandria.
Saccone, V., Cantalini, G. and Moneglia, M. (2023), Prosody, gesture and self-adaptors. A case study of ASD for large corpora collection. *CHIMERA. Romance Corpora and Linguistic Studies*, 10: 211-235.
Sorianello, P. (1994). Il processo dell'allungamento prepausale: dati ed interpretazioni, *Quaderni del Dipartimento di Linguistica*, 5: 47-73.