

Prosody, gesture, and self-adaptors.

A case study of Autism Spectrum Disorder for large corpora collection

Valentina Saccone[°], Giorgina Cantalini² Massimo Moneglia^{°1}

[°]Università di Firenze, ^{*}Scuola di Teatro Paolo Grassi, Milano

Individuals with Autism Spectrum Disorder (ASD) display distinctive speech patterns and bodily movements. This pilot study examines spontaneous interactions between an individual with ASD and a typically developing peer (age 19), incorporating monological and dialogical contexts. The analysis, grounded in the Language into Act Theory framework, explores the information structure of the speech and linguistic parameters influenced by prosody, such as utterance boundaries, information structure, speech disfluency, mean length of prosodic units, and speech rate. The study also employs Kita's model to analyze bodily movements, including gestures and self-adaptors, and their temporal relation with speech. Notable findings reveal that ASD speech is characterized by a monotonous information structure and prosodic contour, featuring slower and longer units with a limited rate variation and information type. On the gestural side, the ASD subject exhibits fewer gestures and more self-adaptors, with some instances of asynchrony between gestures and speech. This pilot study serves as a foundational step for a broader corpus-based project dedicated to exploring the development of pragmatic skills in individuals with ASD.

Keywords: multimodality; Autism Spectrum Disorder; gestures; prosody; self-adaptors

1. Introduction

Individuals with Autism Spectrum Disorder (ASD) who have rich linguistic communication abilities show little problems regarding syntax and lexicon (Janke

¹ Valentina Saccone provided the prosodic analysis, directed the research, and wrote this paper. Giorgina Cantalini provided the data source and the fine-grained annotation of gestures and adaptors. Massimo Moneglia provided the prosodic and informational annotations and achieved gesture analysis.

& Perovic, 2017; Rescorla & Safyer, 2013) but tend to have difficulties in how utterances are performed (Lord & Paul, 1997). Since early studies (Kanner, 1943 and followings; Wing, 1981), prosodic monotony and a significant reduction in gestural expression have been noted in ASD, and presently, prosodic differences are described as robotic and lacking intonation modulation (McCann et al., 2007; Eigsti et al., 2011; Fusaroli et al., 2017). A significant difference with neurotypical subjects was found in various aspects of stress perception and production (Augustyn & Volkmar, 2005). For example, individuals with ASD could produce categorically accurate patterns of statements and questions. However, listeners perceived their prosodic contours as odd, and acoustic measurements showed alterations in duration and pitch and greater variability in fundamental frequency contours compared to typically developing peers (Filipe et al., 2014). To create a standard set of typical acoustic features for Italian children with ASD, relevant studies have been carried out on parameters calculated on the speech flow, such as pitch slope, loudness, and voice quality features such as jitter, shimmer, and HNR (Beccaria et al., 2022; Biancalani et al., 2023).

Research and clinical description have reported differences in gesture quality and synchronization since Hans Asperger's original study of the disorder, which noted the "large," "clumsy," and "inappropriate" gestures of the patients (Asperger, 1944). A deficit in the development of non-verbal behaviors is a characteristic of ASD young children (Mundy et al., 1986). Once language capacity is reached, a general decrease in gesture quantity and a reduced variety are often observed (Colgan et al., 2006; Wetherby & Prutting, 1984), along with issues concerning synchrony with speech (So et al., 2015). In parallel, the high frequency of self-adaptors, such as fidgeting during the communication exchange, has also been noted (Froiland & Davison, 2016). Consequently, reduced communicative effectiveness and a diminished contribution of multimodality to social interaction emerge (Duffy & Hally, 2011).

From a clinical point of view, absence or infrequency of gestures are considered symptomatic in ASD diagnostic measures such as the ADOS (Lord et al., 2002), the Autism Diagnostic Interview (ADI; Lord et al., 1994), and the M-CHAT (Robins et al., 2001).

However, de Marchena and Eigsti (2010), in their study on narratives in adolescents with ASD and neurotypical peers, found large individual differences within each group, with some participants producing as few as two gestures during their story and some producing as many as 23. Crucially, the gestures of ASD subjects were poorly synchronized with the semantically related speech

(preceding or following the relevant speech of around 333 ms average), causing a reduction of effectiveness in communication.

Gesture and speech prosody arise dynamically from an underlying common thought process (McNeill, 2005) or, alternatively, from separate but parallel processes coordinated during production (Chu & Kita, 2008; Kita, 2000; Kita & Özyürek, 2003). Gestures are timed closely with speech and rhythmically coordinated (Loehr, 2007). According to our corpus of adult neurotypical Italian subjects (Cantalini et al., 2020), gestures accompany 90% of the speech flow and are highly synchronous.

Beyond the synchronization with lexical information, gestures also synchronize with prosody at various levels; in particular, prosodic prominence turns out to synchronize with the expressive phase of gestures (Swerts & Kraemer, 2010; Loehr, 2012; 2014; Esteve-Gibert & Prieto, 2013; Shattuck-Hufnagel & Ren, 2018). Also, prosodic boundaries have been recently identified for their role in synchronization (Shattuck-Hufnagel et al., 2010; Esteve-Gibert & Prieto, 2013; Rohrer et al., 2019; Cantalini & Moneglia, 2020; Rohrer, 2022).

Prosody and gestures convey pragmatic information as a function of embodied cognition (Sparaci, 2008). Prosody expresses the interactive value of the speech activity (Cresti 2000, 2020), while gesture expresses the richness of the mental states from which the linguistic performance originates.

Making clear the differentials between ASD and neurotypical in spontaneous, interactive contexts when idea processing is synchronous to the speech activity may shed light on our understanding of the basic process of embodiment in language production.

The pilot study presented in this paper focuses on observing prosody and gestures in spontaneous interactions between one individual affected by Autism Spectrum Disorder (ASD) and one typically developing (TD) peer. The study's objective is to set up an adequate annotation procedure to capture fine-grained differences in prosodic performance, gesture frequency and quality, and gesture prosody synchronization. The pilot aims to provide the foundation for a large-scale project to develop pragmatic skills in ASD.

To this end, the Language into Act Theory perspective (Cresti, 2000; Cresti & Moneglia, 2018), which links prosodic cues to pragmatic function, will be adopted. In parallel, for the analysis of gesture flow, following our previous studies (Cantalini et al., 2020; Cantalini & Moneglia, 2020), we will employ standard models (Kita et al., 1998; Ladewig & Bressemer, 2013; Bressemer et al., 2013).

In 2, we will present the dataset and the prosody and gesture annotation methodology. Sections 3 and 4 will analyze data retrieved from the annotation,

assessing whether the procedure enables the measurement of prosodic and gestural performance in ASD and neurotypical subjects, capturing the atypicality of ASD at both levels. Finally, in the discussion section, we will evaluate the results of the analysis, emphasizing potential distinctive features in prosodic and gestural behavior and their alignment.

2. Methodology

2.1 Setting and Dataset

The dataset consists of an interview with two students at the School of Translation Studies in Milan by their teacher. It is designed to capture monologic and dialogic data in a friendly, interactive environment. The first subject (R) is a 19-year-old ‘Highly activation’ ASD, who has undergone behavioral therapy since the age of 6 and is proficient in language use. The second same-age student (A) is a neurotypical female². Both students, attending the same classes taught by the teacher, provided their consent to be interviewed and audio-video recorded for research purposes. The interview took place within the school theater lab, where the video recording set up is always present. Subjects are seated in front of the camera, oriented 45° toward each other, while the teacher remains out of the frame (Figure 1).



Figure 1. Recording setting

The recording was taken in 2021 during the COVID-19 pandemic, lasting 8 minutes and 18 seconds. Both students answered two questions on topics they were concerned about. Subsequently, the teacher prompted them to question each

² No metadata about test scores are available for the two participants.

other, resulting in a monologic and a dialogic part where the two subjects directly interact with each other and the teacher.

Structure of the interview

- Questions by the teacher to both subjects:
 - ‘What did you think when I asked you to help me in my research?’
 - ‘When did you realize it was harmless and did not concern your evaluation?’
 - ‘Are you worried about the pandemic situation?’
 - ‘Ask your partner any question.’
- A asked R whether he had already seen the new building where the school will be moving shortly.
- R asked A which examination she was more concerned about.

2.2 Prosodic annotation

The Language into Act Theory (L-AcT) (Cresti, 2000; Cresti & Moneglia, 2018) focuses on the pragmatic role played by prosody in speech organization, and it is specifically designed for spontaneous speech corpora analysis). According to L-AcT, the Utterance is the primary referring unit for the analysis of speech and results from pragmatic activities by the speaker; it is autonomous and conveys an illocutionary act (Austin, 1962).

The framework provides explicit methods for speech segmentation into utterances (Moneglia, 2005) and for the annotation of information structure that are based on the hypothesis of a systematic correspondence between prosodic units and information functions (Cresti, 2000; Moneglia & Raso, 2014).

Experts achieve segmentation through perceptual judgments. Prosodically terminated sequences (TS) are assumed to correspond to the utterance boundaries (Izre’el et al., 2020). Subsequently, TSs are segmented into prosodic units. The methodology relies on the following principles³:

- Utterances performing a pragmatic activity correspond to sequences of prosodic units ending with a perceptively relevant terminal prosodic boundary.
- Non-terminal prosodic boundaries within a TS identify Prosodic units.

³ The annotation of the terminal and non-terminal prosodic boundaries follows from perceptual evidence. Various corpus-building initiatives have validated the annotation schema (Danieli et al., 2004; Moneglia et al. 2010; Amir et al. 2004; Izre’el & Mettouchi 2015; Panunzi et al. 2020).

- Prosodic units are characterized by perceptively relevant prosodic movements ('t Hart et al. 1990).
- Prosodic units correspond to units of information (IUs) (Chafe, 1994) and bear functional values defining the Information structure of the utterance (Cresti, 2000).

Beyond speech segmentation criteria based on prosodic cues, the main contribution provided by this approach is the idea that within the utterance, one specific IU type, the Comment (COM), bears the illocution and, for this reason, is its core structuring element. The COM carries the pragmatic and prosodic autonomy and is necessary and sufficient to form an utterance. The prosodic contour of the COM widely varies as a function of the illocutionary act performed. It can be described as a *root* unit ('t Hart et al., 1990).

The *utterance* can be *simple*, consisting of only one COM IU, possibly scanned in prosodic units with no information function. An utterance is *compound* when composed of several IUs, giving rise to a prosodic pattern where each IU conveys a pragmatic value. The following is the minimal tagset of IU types adopted for this research.

Textual Units

- *Comment* (COM): IU expressing the illocutionary information (necessary and sufficient to perform an utterance).
- *Multiple Comment* (CMM): IU expressing the illocutionary information within a prosodic model, giving rise to an illocutionary pattern.
- *Bound Comment* (COB): IU weakly expressing illocutionary force within an adjunctive process.
- *Topic* (TOP): IU specifying what domain the illocutionary act is about.
- *Parenthesis* (PAR): IU that adds information to the utterance on a secondary plane.
- *Appendix* (APC/APT): IU that integrates the information of the Comment or Topic units with unnecessary information.
- *Scanning Unit* (SCA): prosodic unit marked by a non-terminal boundary that only scans the lexical content of one IU, not playing any Information function.

Brief Units

- *Locutive introducer* (INT): signals the onset of reported speech or thought.
- *Discourse Connector* (DCT): signals the link of the utterance with previous information in the discourse or the context or within Comments.

- *Dialogical*: IUs of different types with no semantic content devoted to the management of the interaction (Discourse Markers)

A TS performs one utterance if it contains one illocutionary unit COM. However, TSs can contain more than one illocutionary unit. In this case, they can be patterned in a prosodic model to give rise to rhetorical effects (Saccone et al., 2018; Saccone & Panunzi, 2020) or can be performed in an un-patterned sequence (Saccone, 2022). Accordingly, the theory foresees two types of TSs beyond the utterance with different pragmatic values:⁴

- The *utterance*, which corresponds to one *linguistic act* and can be *simple* (consisting only of one IU) or *compound* when structured by more than one IU;
- The *illocutionary pattern*, a TS which structures several linguistic acts (usually two), performed within a prosodic model, to express rhetorical relations (*reinforcement, list, comparison, tag questions, double questions, etc.*);
- The *stanza*, a TS made up of a sequence of weak illocutionary acts that follow the flow of thought. It develops through an additional process not corresponding to the execution of a prosodic model.

Information units are characterized by their prosodic form and distribution with respect to the COM. Considering textual units, TOP units always precede the COM and have a *prefix* prosodic contour ('t Hart et al., 1990; Cavalcante, 2016). The APC integrates the COM and necessarily follows it. In 't Hart's terms, APC is performed with a suffix prosodic contour and does not bear functional prosodic prominence (Cresti, 2021). Parenthesis is characterized by a lowering of the f0 mean with respect to the preceding and following IUs (Saccone & Trombetta, 2021; Saccone & Panunzi, 2023). Finally, Scanning units do not bear any prosodic prominence.

Prosodic units are characterized by *perceptively relevant prosodic movements* whose relevance is connected to the functional value conveyed. In particular, *prefix* and *root* prosodic contours can comprise a *preparation*, a *nucleus*, and a *tail*. The nucleus corresponds to the minimal prosodic contour sufficient to achieve the information function of the unit.

TSs have been transcribed and aligned to the speech signal in WINPITCH (Martin, 2004). Each IU has been annotated with its Informational value and

⁴ See Moneglia & Raso (2014) for details on the reference unit's types in L-AcT.

aligned to the signal in a dependent tier. Annotated files have been converted into Praat files (Boersma & Weenink, 2021), imported into ELAN (2023), and reconciled with gesture annotation. The acoustic signal is also processed through semi-automatic tools to create annotation tiers for the classification of sounding/silence segments, syllable division, and extracting acoustic features. This analysis is conducted using Praat software for prosodic analysis.

2.3 Gesture annotation

Gestures and prosody have been annotated independently one from the other by expert annotators according to the standard models proposed by Kita et al. (1998), along with additions from Ladewig & Bressemer (2013), Bressemer et al. (2013), Lausberg (2013).

The model foresees a hierarchy of gestural elements that develop around a mandatory Expressive nuclear phase (*stroke*) in which a peak of energy constitutes the semantic part of the gesture. The Expressive phase may be either simple or compound, dynamic or stationary. Accordingly, Co-speech gestures have been segmented into the following hierarchic levels.

- *Gesture Unit*: a sequence of one or more gestures between two rest positions.
- *Gesture Phrase*: the minimal gestural linear pattern constituted by *Phases* of a gesture around a prominence (*Stroke*).
- *Phases*: the *Stroke* compulsory root, optionally preceded by *Preparation* and followed by *post-stroke Hold* or *Retraction*.

Figure 2. represents these hierarchical relations.

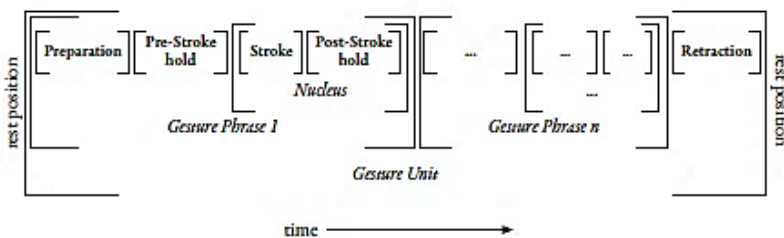


Figure 2. Gesture unit, phrases, and phases (from Andrén, 2010)

Gestures are distinguished from other hand movements since only gestures contribute to communication and have a structure comprising a peak of energy in their expressive part.

The annotation procedure is supported by acknowledging the functions of different gestural components, whose definitions (listed in Figure 3 taken from Ladewig & Brassem, 2013) work as operative instructions.

gesture phase	function within sequential embedding
preparation	<ul style="list-style-type: none"> – prepare the hands for the execution of the stroke – move the hands to a particular position in gesture space – assume a particular configuration with which the stroke begins
retraction	<ul style="list-style-type: none"> – transition from stroke to possible rest position
stroke	<ul style="list-style-type: none"> – forms the center of a gestural unit (“nucleus”, Kendon 2004)
hold	<ul style="list-style-type: none"> – neighbors stroke (either precedes or follows it) – may selectively stand alone – belongs to the center of a gestural unit (“nucleus”, Kendon 2004)
rest position	<ul style="list-style-type: none"> – default condition

Figure 3. (Ladewig & Brassem, 2013:1075)

However, in dealing with ASD subjects where gesticulation is known as reduced, hand movements have scarce perceptive relevance, and their contribution to communication is not evident. Therefore, the annotation strategy requires the maximum granularity of observation to be effective, as the examples in the following section will make clear.

Gesture Units, Gesture Phrases, Strokes, Preparation, Hold, and Retraction have been aligned to the acoustic signal in ELAN in dependent tiers for each participant. Transcription and prosodic annotation of the speech flow were unavailable to annotators during gesture annotation.

Gesture types have been classified according to traditional typologies which integrates the categories proposed by McNeil (1992) with categories derived from the functional approach by Kendon,(2004). The following are the semiotic labels used to this end:

- *Emblems*: highly conventionalized gestures.
- *Iconic* (representational): gestures depicting objects or actions.
- *Deictic*: pointing to person or space/time relations.
- *Batonic*: gestures that accompany the rhythm of speech units.
- *Pragmatic*: recurrent gestures depicting an underlying metaphor supplementary to speech units.
- *Interactive*: gestures stressing the role of the interlocutor with respect to the utterance.

Crucially, the annotation also encompasses involuntary self-touching movements, referred to as *self-adaptors* (Kendon, 2004). Self-adaptors are “manipulations of one’s body parts or other objects, being peripheral or nonessential to central ongoing events or tasks (Mehrabian & Friedman, 1986) and do not convey meaning contrary to gestures co-occurring with speech.

Self-adaptors have been frequently noted in ASD speech may indicate discomfort. The literature reports correlations between self-adaptors and the perceived emotional state of the speaker (Lin et al., 2021; Mahmoud, 2013; Mehrabian & Friedman, 1986; Neff et al., 2011). Their processing is harder than that of iconic gestures related to the semantic content of the speech (Chui, 2018). This behavior is continuous and cannot be segmented into distinct units. Self-adaptors are distinguished in this annotation from generic body movements since they specifically regard hand/finger movements and touching where the movement made by hands is defined (they move, for example, to the head, scratch, rub, and adjust their sleeve). Generic hand movements made during the recording have been annotated, although distinguished from self-adaptors. Self-adaptors have been annotated throughout the recording when subjects were speaking or silent.

The following are typical self-adaptors:

- Self-manipulators like scratching one’s leg (Chan et al., 2016).
- Bringing a hand to the mouth/head when looking for a word or “thinking gesture” (Breckinridge et al., 2017), also known as “Butterworths”, McNeill, 2005).
- Fidgeting could be caused by physical discomfort (Mahmoud et al., 2013).
- Nervous ticks.
- Small hand or finger movements.

General observations regarding the synchronization of co-speech gestures and self-adaptors with prosodic units and utterances have been derived once both speech segmentation in TSs and IUs and gestures are annotated and reconciled in ELAN. The following figure reports the annotation grid in ELAN used for this pilot, respectively, for R and A. The interviewer, who is out of the scope of the camera, has received only the transcript divided into TS.

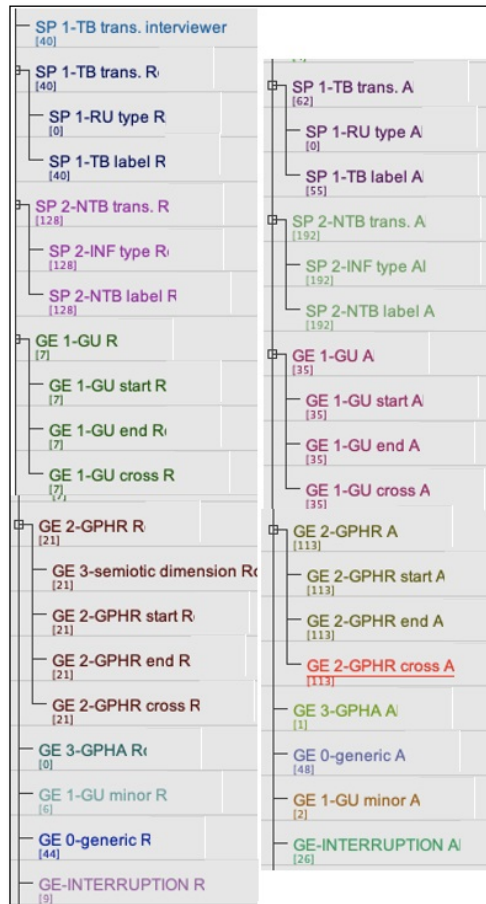


Figure 4. Gesture / Prosody annotation grid in ELAN

For the Speech signal of each participant (SP), the annotation identifies the transcript of each TS and its Type according to L-AcT. The Information function is annotated in a dependent set of tiers for each sequence ending with an NTB (non-terminal boundary) within each TB (terminal boundary). Gestures (GE) are annotated considering Gesture Units, Gesture Phrases, their semiotic dimension, and Self-adaptors. Each gesture unit (GU 1) is annotated, marking its starting and ending points. Each GU and each GPHRS (GE 2) is annotated, marking their starting and ending points. GPHRs are annotated according to their semiotic dimension (GE 3). Self-adaptors are identified in GE 0. To help the retrieval of the relation between prosodic boundaries of IUs (TS and NTB) with respect to

GUs and GPHRs, specific tiers (GU cross and GPHR cross) report whether the gesture crosses TB or NTB.

4. Results of speech annotation

4.1 Information structure

The speech of A and R has been segmented based on L-AcT, dividing each turn into TS and subsequently into IUs. Table 1 illustrates the quantitative data for the two speakers.

Table 1. The speech production of A and R at a glance

	A	R
<i>Turns</i>	28	23
<i>Terminated Sequences</i>	55	40
<i>Verbal TS</i>	40 (72%)	28 (70%)
<i>Information Units</i>	192	128
<i>Textual IU</i>	161 (85%)	109 (84%)
<i>Non-textual IU</i>	31 (15%)	19 (16%)

No relevant quantitative differences emerge, ensuring the comparability of the two sets of TSs.

Focusing on Terminated Sequences, we classified them based on their type (utterance, illocutionary pattern, and stanza) and structure (simple or compound utterances) (Table 2):

Table 2. Overview of Terminated Sequences of A and R

Terminated Sequences		A	R
<i>Utterance</i>	tot.	34	33
	Simple	18	18
	Comp_text	12	11
	Comp_dial	4	4
<i>Illocutionary Pattern</i>		11	4
<i>Stanza</i>		10	3
TOT		55	40

Both participants produce nearly the same amount of Simple (A, R: 11) and Compound Utterances (A: 12+4; R: 11+4), which thus form a comparable sample

for the two speakers. Compound utterances have been divided into two groups: the textual one (Comp_text) collects utterances composed by the Comment, which bears the illocution, one or more textual information units, and possible dialogic information units; the dialogic one (Comp_dial) collects utterances composed by the Comment and dialogic information units only, which are not a signal of complexity. In Comp dial one, IU refers to the interlocutor and does not give rise to informational relations between IUs.

On the contrary, Illocutionary Patterns and Stanzas are quantitatively different in the two speakers. R produces fewer of both types, mainly structuring his turns with only one illocution per each TS (A: 11, 10; R: 4, 3).

Looking at the IUs, Figure 5 shows the units that compose the TSs grouped based on the type of TS. The plots report a row number of recorded items, which are directly comparable concerning simple and compound Utterances due to the parity of COM units.

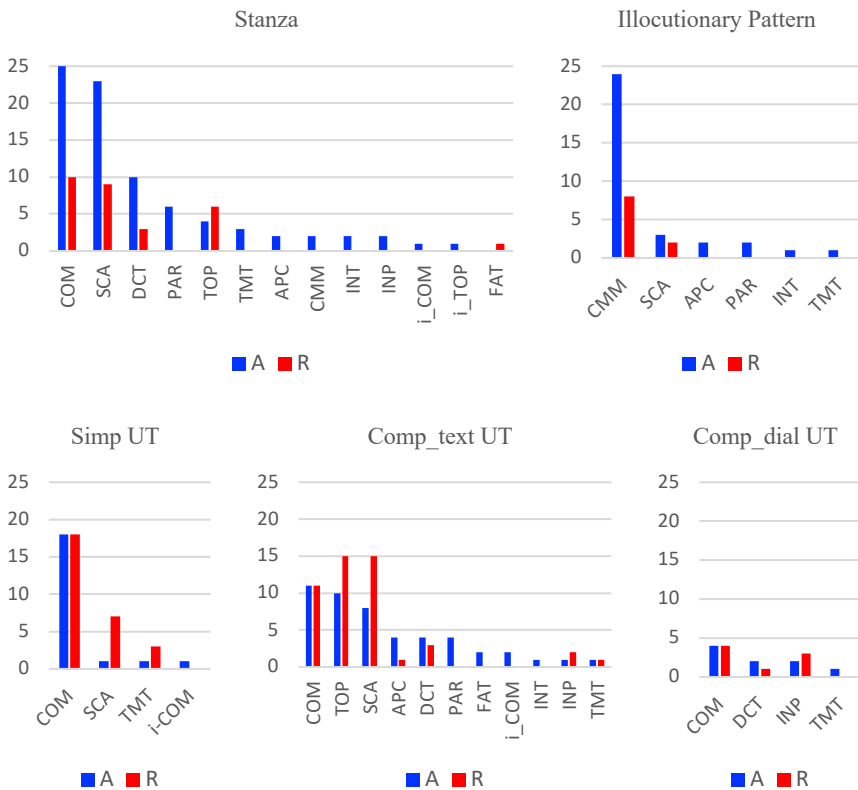


Figure 5. Information structure of TSs per each type.

R's speech shows less variety in IUs, structuring the TSs without the recurrence of Appendixes and Parentheses (only 1 APC is found in R's Compound UT), which, on the contrary, are present in A's Stanzas, Illocutionary Patterns, and Compound-textual Utterances. Specifically, the absence of PAR in an interview, where the speaker can take time to answer without the hit and response of a quick dialogue, deviates from spontaneous speech data (Saccone, 2022; Saccone & Panunzi, 2023).

This is not to say that R does not articulate his speech. The presence of TOP units manifests the capability of structuring the information patterns, and it is particularly noteworthy as the Topic serves as the primary means of organizing information in relation to the Comment (Cresti & Moneglia, 2018)⁵.

However, the primary structuring strategy of R is the Scanning of the locative content into units (SCA) in connection to the limits of duration and syllabic length, such as not surpassing a canonical size of seven syllables (Martin, 2009; Miller & Weinert, 1998) rather than to the pragmatic architecture and textual hierarchy of the speech. In fact, with SCA, the relation between prosodic units does not give rise to information patterns, and the locutive content is fully compositional.

In particular, the internal structure of the 3 R's Stanzas is minimally articulated. His 3 Stanzas are chains of 3-4 Bound Comments, with Topic as the only other textual unit. Conversely, for what concerns A, her Stanzas are composed of 2 to 4 Bound Comments, together with Topic, Appendix, and Parenthesis as other textual units.

Comment units are often prosodically scanned for both speakers in more than one Tone Unit. This usually happens in Stanzas but occurs in the other types of TS, too. Table 3 focuses on the distribution of Scanning units (SCA).

Table 3. SCA units in different types of TSs

SCA	A	R	SCA of COM	A	R
<i>Simple UT</i>	0	7 (25%)	<i>Simple UT</i>	0	7 (38.9%)
<i>Comp_dial UT</i>	0	0	<i>Comp_dial UT</i>	0	0
<i>Comp_text UT</i>	9 (18.8%)	15 (31.3%)	<i>Comp_text UT</i>	0	9 (81.8%)

⁵ With the necessary approximation due to the limited sample size, especially concerning stanzas, the presence of Topic in A (26% of utterances; 30% of stanzas) and R (30% of utterances; 2 on 3 stanzas) follow the general trend observed for Italian monologues in DB-IPIC (Panunzi & Gregori, 2012) (21.5% of utterances; 34.5% of stanzas).

<i>Illocutionary Pattern</i>	3 (9.1%)	2 (20%)	<i>Illocutionary Pattern</i>	1 (4.2%)	2 (25%)
<i>Stanza</i>	23 (28%)	9 (31%)	<i>Stanza</i>	20 (80%)	4 (40%)
TOT.	35 (18.2%)	33 (25.8%)	TOT.	21 (25.6%)	22 (43.1%)

As shown in Table 3, the total number of SCA units is similar for the two participants (A: 35; R: 33), but the distribution differs⁶. A has no SCA units in Simple UT, and 23 SCAs are exclusively found in Stanzas, following the expected trend (Saccone, 2022). In contrast, R's SCAs are primarily in Compound UTs (15) and Simple UTs (7). SCAs in Simple UTs, necessarily scanning COM units, together with the recurrence of time-taking units such as vocalizations, reflect flattening in the utterance and a sense of monotony for the auditory.

4.2 Prosodic features

The segmentation in TSs and IUs served as a base for the acoustic and prosodic analysis. To this end, the audio was processed through Praat software.

The acoustic measures pertain:

- duration of the stretch of speech and pauses;
- mean length of TSs and IUs in syllables;
- speech rate of TSs and IUs in syllables/second;
- f0 range of IUs.

4.2.1. *Duration of the stretch of speech and pauses*

Firstly, we divided the speech flow into silent and sounding segments using an automatic procedure. We then manually verified and measured the duration of the segments for the two speakers, with a cutoff point of 150 ms for the minimum pause duration.

The segmented pauses were classified based on their position relative to the information structure of the turn⁷: between IUs of the same TS (IU-pauses), between TS of the same turn (TS-pauses), at the beginning of a new turn (T-pauses). Table 5 displays the measurements.

⁶ The calculation was performed in the first case (SCA) on the total units and, in the second case (SCA of COM), on the total Comments.

⁷ The analysis and classification of pauses is based on Saccone & Trillocco (2022) and Saccone et al. (2023) and their empirical works conducted on schizophrenic speech, which, contrary to the data of both A and R, show a massive presence of very long pauses in TS- and T-position, the second reaching a length of more than 20 seconds.

We consider the stretch of speech as the duration of TSs (i.e., the sum of sounding segments and IU pauses). For each type of pause, the number in parentheses represents the percentage of units involved in pausation (IU, TS, or T), while squared brackets contain the range of pause duration (minimum-maximum). The table also reports the pause-to-speech ratio, calculated as the sum of TS- and IU-pauses duration divided by the stretch of speech duration.

Table 5. Duration of speech and pauses

duration	A	R
<i>stretch of speech</i>	3:35	3:00
<i>IU-pauses</i>	12.66 s (23.4%) [152-1357 ms]	8.19 s (26.1%) [159-1336 ms]
<i>TS-pauses</i>	5.99 s (43.8%) [186-872 ms]	0.84 s (16.7%) [299-544 ms]
<i>T-pauses</i>	4.59 s (43.5%) [224-791 ms]	10.86 s (75.0%) [209-1335 ms]
<i>pause/speech</i>	0.09	0.05

No differences were found for IU pauses, while the duration and frequency of pauses in the other two groups differed. A has more and longer pauses between TS than R, while the opposite happens for T-pauses. In line with the literature on comparative studies between TD and ASD children and adolescents (Gorman et al., 2016; Irvine et al., 2016; McGregor & Hadden, 2018), the low number of A's T-pauses can be read in correlation with a higher presence of filled pauses, time-taking vocalization, and discourse markers isolated in specific IUs (*Dialogical Units*) at the beginning of the turn (A: 40%; R: 30% of turns). Note that filled pauses in dialogue fulfill pragmatic functions, facilitating understanding and the flow of conversation, keeping the turn, signaling politeness and attention, or foreshadowing the duration and informativeness of upcoming complex utterances (Fox Tree, 2001; Corley & Hartsuiker, 2003).

4.2.2. Mean length and speech rate of TSs and IUs

We chose to measure the length of the different units in terms of phonological syllables with a semi-automatic procedure based on Praat tools and their speech rate in terms of syllables/second. Table 6 summarizes the resulting measures for each type of TS, with values in brackets indicating standard deviation.

Table 6. Mean length and Speech rate of TSs

	A	R
<i>mean length</i>		

<i>Simple UT</i>	5.75 syll (5.03)	9.56 syll (9.98)
<i>Compound UT</i>	24.56 syll (20.86)	26.07 syll (16.91)
<i>Illocutionary Pattern</i>	15.09 syll (18.50)	11.50 syll (12.07)
<i>Stanza</i>	50.30 syll (18.82)	68.67 syll (12.66)
		speech rate
<i>Simple UT</i>	5.05 syll/s (2.29)	3.61 syll/s (1.69)
<i>Compound UT</i>	5.47 syll/s (1.41)	4.49 syll/s (1.03)
<i>Illocutionary Pattern</i>	4.71 syll/s (1.82)	4.05 syll/s (1.54)
<i>Stanza</i>	5.95 syll/s (0.80)	4.34 syll/s (0.53)

Data are specifically relevant for Simple and Compound UT since A's and R's speech includes the same number of items for these types of TS⁸. The general trend is that R's TSs are longer and slower than A's⁹ (Table 6). This is particularly evident in Simple UT (A: 5.75 syll; R: 9.56 syll).

Looking at the speech rate, a lack of variation emerges comparing R to A in each type of TS, as shown by the lower rate of standard deviation in R for each mean value reported in the table.

For a more fine-grained analysis, the same measures were taken for IUs (Table 7).

Table 7. Mean length and Speech rate of IUs

	A	R
		mean length
<i>IU</i>	6.22 syll (4.52)	6.56 syll (5.21)
<i>Textual IU</i>	6.95 syll (4.52)	7.40 syll (5.19)
<i>Textual IU in SimpleUT</i>	6.17 syll (4.94)	6.78 syll (5.89)
		speech rate
<i>IU</i>	4.81 syll/s (2.00)	4.04 syll/s (1.42)
<i>Textual IU</i>	6.24 syll/s (2.22)	4.63 syll/s (1.66)
<i>Textual IU in SimpleUT</i>	5.34 syll/s (2.20)	3.65 syll/s (1.65)

As expected, the trend is the same as the previous findings, and again, R's IUs are longer and slower than As.

The gap between speakers increases by narrowing down the sample of IUs under consideration from total to textual and textual in Simple UTs. The standard deviation of the speech rate is again lower for R's measures. Furthermore, the speech rate is at its maximum gap in textual IUs.

⁸ The mean length of Illocutionary Patterns is the only parameter that does not follow the observed trend. This does not change the picture since the number of Illocutionary Patterns is not comparable between the two speakers and is not considered.

⁹ However, it is worth noting that A's speech is perceived as fast by Italian native speakers.

4.2.3. *F0 range*

Based on the acoustic quality and consistency in illocution¹⁰, we selected COM units from the speech flow of the two participants, resulting in a comparable set for simple utterances (A: 10; R: 7) and compound utterances (A, R: 11). Data are collected in Table 8.

Table 8. *f0* range in COM units

f0 range (st)	A			R		
	mean	sd	items	mean	sd	items
<i>Simpl UT</i>	15.30	8.16	10	11.33	3.91	7
<i>Simpl UT_holo</i>	2.45	-	1	7.29	4.21	5
<i>Comp UT</i>	11.50	7.19	11	8.48	2.28	11
<i>Stanza</i>	12.67	7.93	8	16.45	-	1
TOT.			30			24

The *f0* range was measured in Hertz and semitones (st) for each illocutionary unit. Table 8 presents mean values in st to allow a direct comparison between the two speakers.

Comparing the mean of the *f0* range for A and R, it is possible to see a reduction both in Simple (A: 15.30 st; R: 11.33 st) and Compound UTs (A: 11.50 st; R: 8.48 st) in R, which reports a flatter prosodic contour and suggests a fewer illocutionary variation.

Standard deviation values highlight a notable lack of variation when comparing R to A in Simple (A: 8.16; R: 3.91) and Compound UTs (A: 7.19; R: 2.28). It is noteworthy that 't Hart (1981) found the Just Noticeable Difference (JND) for speech to average between 1 and 2 st, with differences of more than 3 st being relevant in communicative situations.

The lower dispersion of R's measures can be noticed in Figure 6, which displays the *f0* range (st) of each selected COM unit (A: 30; R: 24). The blue dots (A) cover a broader area on the plot compared to the red dots (R), depicting that the values are spread out over a broader range for A.

¹⁰ Interrogative Comments and holophrases are treated separately due to their prosodic peculiarities.

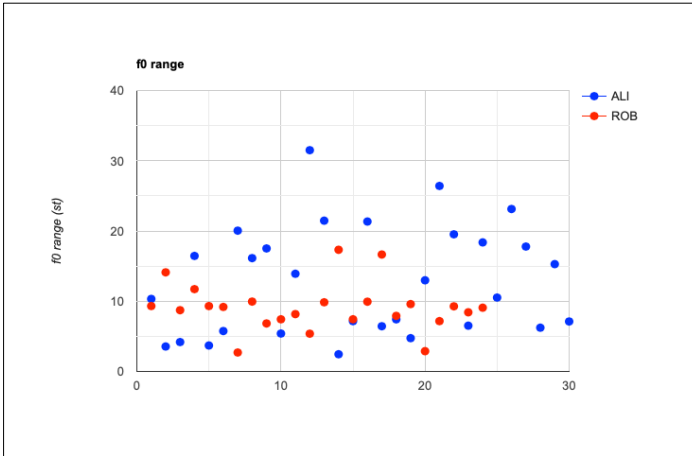


Figure 6. f0 range in COM units

3. Results of gesture annotation

3.1 Quantitative and qualitative reduction of gesticulation in ASD

Table 10 sketches the consistency of the annotated dataset.

Table 10. The annotated dataset

	R	A
<i>Speech time</i>	3:00	3:35
<i>Words</i>	486	632
<i>TS</i>	40	62
<i>Information Units</i>	128	192
<i>Gesture Units</i>	7	35
<i>Gesture Phrases</i>	21	114
<i>GPRS/IU</i>	0.16	0.59
<i>Gesticulation</i>	0:27	2:47
<i>% speech gesticulation</i>	15%	77.6%
<i>Self-adaptors</i>	3:37	0:40

The quantitative reduction of gesticulation in ASD face to the neurotypical participant emerges considering the number of GPRS in proportion to the information units, roughly one out of three (A: 0.59; R: 0.16). In parallel, the

percentage of speech time accompanied by gestures (A: 77.6%; R: 15%). In 3 minutes of speech, R produced 7 GUs containing 21 GPHRs. 5 GUs are filled by one GPHR. On the contrary, in 3:35 minutes of speech, A produced 35 GUs containing 113 GPHRs. 18 GUs contain a single GPHR.

The opposite trend is recorded when considering Self-adaptors, which strongly characterize the body involvement of the ASD, as expected.

The reduction of gesticulation can also be identified considering the qualities of GPHR, which, compared to those by A, turns out to be minimal movements that do not extend in time and space.

GPHRs by R are minimal hand movements performed at the lower level, centered on the belly or his lap. These are pretty peripheral positions according to McNeill’s gesture location schema (in Figure 7).

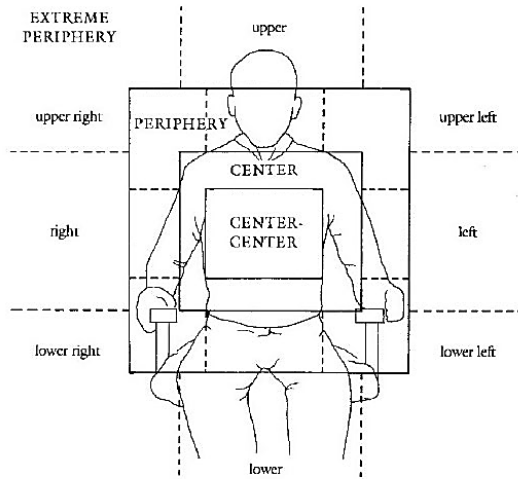


Figure 7. Gesture location (Mc Neil, 1992)

This is the case for the gestures in the first and second screenshots of Figure 8, which might be hard to perceive and, therefore, add little contribution to the communication. To study gestures with these qualities, an extremely granular annotation is needed. Nonetheless, R shows meaningful hand movements while speaking, including upper-level, less compressed gestures like the deictic gesture in the third screenshot.

Hand on the lap (low center)	Hand on the belly (under center-center)	Full gesture spam (upper left and right)
------------------------------	---	--

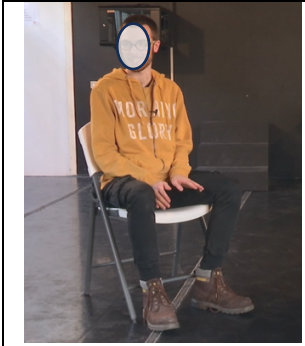
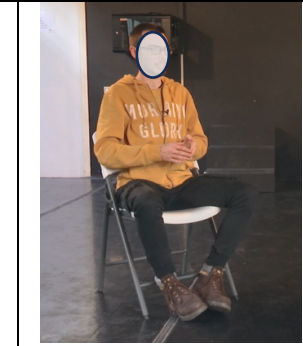

		
quindi / avevo completamente travisato / la situazione nel mio cervello	ma cosa ho sentito fino ad adesso ?	ha fatto / un salto indietro
<i>so / I had totally messed up / the situation in my head</i>	<i>what have I heard until now?</i>	<i>he took / a step back</i>

Figure 8. Reduced gesture quality in R

Compared to R, the quality of gesturing in A does not exhibit any qualitative reduction. For instance, the selection in Figure 9 illustrates large palm-up gestures in the central position moving in various directions.

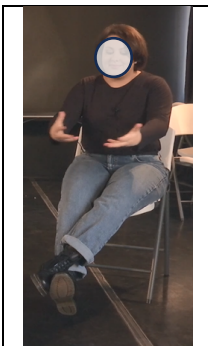

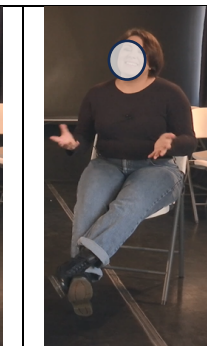
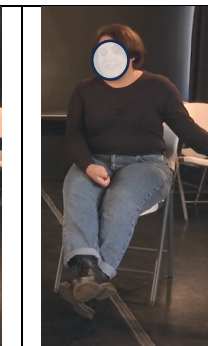





			
mi piace / presentarmi /	l'unico problema [...] è stata la logistica	ci sono arrivata	inizio a sentire / sentire tante voci
<i>I like / introducing myself</i>	<i>the only issue [...] was the logistics</i>	<i>I got there</i>	<i>I'm starting to hear / hear many voices</i>

Figure 9. Gesture quality in A

This quality of gestures could be made more objective simply considering that the average time of one GPHR is almost double in A (0.72 s in R vs. 1.21 s in A).

3.2 Gesture typology





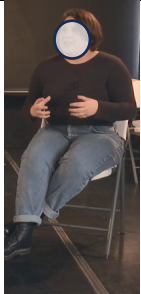


From the perspective of gesture typology, among the 21 GPHRs produced by R, we found strokes of all types except Emblems. When considering Iconic gestures, we observed that they only depict actions and never objects (Figure 10). Although data are too scarce to derive generalizations, it is notable that R, despite his quantitative reduced gesticulation, does not exhibit any impairment connected to the gesture typology frequently reported in the literature for ASD, consistent with the findings reported by Sparaci et al. (2019).

emblems	deictic	iconic	beat
not found			
	quando lei mi ha scritto in privato	mi si è illu [/] accesa la lampadina	ho pure fatto la terza dose sabato
	<i>when you wrote to me privately</i>	<i>a lightbulb turned on in my head</i>	<i>I also got the third dose on Saturday</i>
	pointing	Mimic of turning-on light	Beat on lap
pragmatic (metaphoric)		interactive	
			
quindi / avevo completamente travisato / la situazione nel mio cervello		ieri / quando [/] quando dopo essere arrivata a Brecht	

<i>so / I had totally misunderstood / the situation in my mind.</i>	<i>yesterday / when / when after arriving in Brecht</i>
	Scope on the Topic function

Figure 10. Gesture semiotic typology in R

A produces a rich series of co-speech gestures and produces instances of each type in the typology (Figure 11). Most of these gestures are pragmatic, while emblems are produced in only two cases.

Emblems	Deictic	Iconic	Iconic
			
<i>l'unica problematica / tra virgolette</i>	<i>dici che è meglio questa ?</i>	<i>riesco ad arrivare</i>	<i>un interruttore</i>
<i>the only issue / in quotes</i>	<i>do you think this one is better?</i>	<i>I can manage</i>	<i>a switch</i>
Fingers signing quotation	Positioning hands to identify a point in space	The hand moved right to the left as a mimic of a movement	Mimic of the object turning fingers
Beat	Pragmatic metaphoric	Interactive	
			
e	ci sono arrivata	no	

<i>and</i>	<i>I got there</i>	<i>no</i>
Top-down movement followed by a hold	Palm-up open hands opening movement “As you can see”. positive evaluation	Movement of the hand opened toward the interlocutor to stop his intervention

Figure 11. Gesture semiotic typology in A

3.3 Gesture Units

The qualitative attitude of gathering GPHRs within one sole arch of movements characterizes A but is also clearly testified in R.

Among 113 GPHRs performed by A, 94 are embraced within GUs, gathering more than one gesture (83% of GPHRs). Some of these GUs are very long gesture sequences, given that only 80 GPHRs are gathered in 6 GUs, while 20 of 35 Gus (57%) comprehend one gesture only.

Among the 21 GPHRs performed by R, 16 are gathered within two GUs (76%), while 5 of 7 (71%) comprehend one gesture only.

From the fine-grained observation of the language contexts produced by R in which GUs guest 6 and 10 gestures, we notice that the two-gesture series occur in the interactive part of the recording, and both end in correspondence with a TB.

3.4 Gesture prosody synchronization

The annotated dataset has derived general observations regarding synchronizing co-speech gestures with prosodic units and utterances. For the evaluation of these findings, it should be considered that, in previous studies, a robust synchronization of gesture with utterance at the higher level of the Pragmatic / Prosody relation has been observed in the multimodal communication dataset of Italian subjects (Cantalini et al., 2020; Cantalini & Moneglia, 2020). In particular, as Figure 12 shows, GPHRs rarely cross the utterance limit marked by terminal prosodic boundaries.

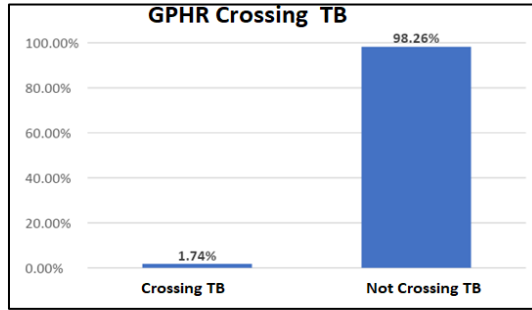


Figure 12. GPHR / TB relation in the Italian reference data set (Cantalini, 2022)

Moreover, a strong tendency of GPHRs to start and end at prosodic boundaries has also been registered. As reported in Figure 13, the onset or the end of GPHRs cooccurs with a prosodic boundary in around 75% of cases (Cantalini et al., 2020; Cantalini & Moneglia, 2020)¹¹.

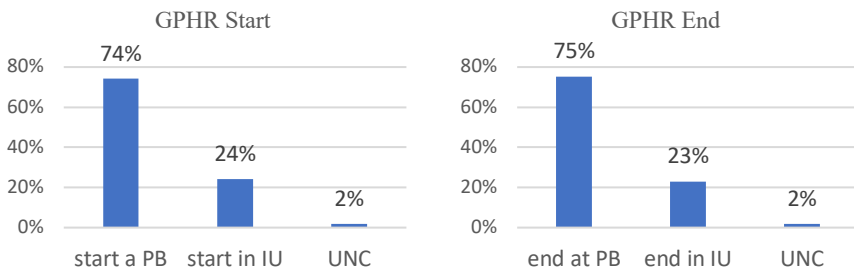


Figure 13. GPHR / Prosodic boundary relation in the Italian data set (Cantalini, 2022)

This overall correlation is consistent with the data from both R and A. Among R's 21 GPHRs, no one crosses terminal prosodic boundaries. In the case of A, only 6 GPHRs among her 114 cross the terminal boundary, but in the retraction phase in specific prosodic conditions¹².

Also, the tendency to start and end at a prosodic boundary is confirmed in both subjects. In R, among his 17 GPHRs, 13 start and 14 end at the prosodic boundary (around 80%). In A, among her 114 GPHRs, 76 start and 79 end at the




¹¹ Approximation of +/- 250 ms.

¹² The close observation of these contexts has shown, however, that in those cases, the illocutionary force conveyed by the utterance was weak, and in parallel, the annotation of the boundary quality (terminal or not terminal) was uncertain.

boundary (around 70%). Therefore, there are no signs of atypical synchronization behavior in the data from the ASD subject in this pilot when examining the relation between gestures and prosodic units.

However, in parallel to the above general data concerning gesture/prosody synchronization, the relation between gesture and linguistic content should also be considered. At the upper level of this relation, data are consistent with the synchronization principle that the linguistic anchor of gestures is always found within the TS. Beyond this general assumption, however, we may also assume that the stroke should be specifically synchronous to its content (Chui, 2005; Graziano et al., 2020).

An atypical asynchrony of gestures with the linguistic material, which should, in principle, constitute their anchor, has been noted as a peculiar feature of ASD (de Marchena & Eigsti, 2010). When the gesture alignment of R is considered in detail, this notation may be replicated, although not all through his gestural behavior. Excluding beats, which can hardly be anchored to some specific entry, 18 GPHRs by R could be, in principle, anchored to the linguistic information. Among these GPHRs, the gesture turns out delayed (starts and ends after) in two cases, and in three cases, is anticipated (starts and ends before), while in one case, it extends beyond its actual anchor (ends after). In other words, roughly one out of 3 gestures of R is asynchronous with the linguistic information they refer to. Figure 14 gives a picture of this loss of synchronicity, which scores a few milliseconds and is hardly perceived as disturbing the flow of communication. No such phenomenon emerges in the rich gestural sequences by A.

anticipated	delayed	extended
		
ha fatto / un salto indietro	<i>più o meno lo schema</i>	<i>praticamente / la tragedia greca</i>
<i>he took / a step back</i>	<i>more or less the scheme</i>	<i>basically / the Greek tragedy</i>

Pointing behind	Shaking hands up and down (approximation)	Repeated movement of the two hands from the internal to the external position (approximation)
-----------------	---	---

Figure 14. Gesture content asynchrony in R¹³

3.5 Self-adaptors

In both speakers, self-adaptors can cooccur both while speaking and during silence, never overlapping with gesticulation, and are not aligned to speech events. We did not observe qualitative differences between the two subjects except for the frequency and duration of these behaviors. For instance, see the self-adaptors behaviors reported in Figure 15.

In A, 13 instances of self-adaptors occur. 9 of them happen during speech, and two (one lasting 10 seconds) occur during silence, while two start while A is in silence and continue while A is speaking. The total duration of this behavior is around 40 seconds (17.3% of the speech time). In R, self-adaptors identified in the annotation are 17 activities, with 7 occurring while speaking, 6 covering both speaking and silence, and 4 occurring during silence. The total duration of this behavior (3:37) exceeds the speech time.

Comparing the time of gesticulation with the time of self-adaptors in the two subjects during the recording shows that self-adaptors are the body movements that most characterize ASD, while gesticulation characterizes the neurotypical subjects.

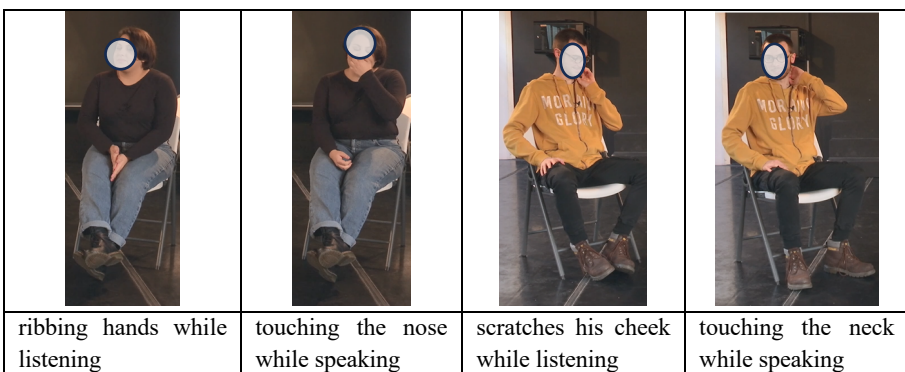


Figure 15. Self-adaptor examples in A and R

¹³ In the Italian text, the wording to which the gesture is temporally aligned is in bold, while the wording to which the gesture should be anchored is in italics.

5 Conclusions

In this paper, we presented and tested a methodological design that relies on speech and gestural frameworks grounded on pragmatic principles, allowing segmentation for gestures and speech. The pilot provides valuable insights into the spoken and gestural characteristics of ASD individuals and their comparison with the TD subjects, laying the groundwork for future research in understanding and addressing communication challenges in ASD.

In short, the analysis of the speech in terms of information structure of the subject with ASD has revealed the following atypia:

- turns are mainly structured with only one COM per TS, with a scarce presence of stanzas and illocutionary patterns, which nonetheless may occur;
- the variety of information units is limited, articulating the TSs with Topic but without Appendixes and Parentheses, which integrate pieces of information in the process of performing the utterance. In particular, the absence of PARs manifests difficulty in structuring information hierarchically across distinct textual levels, avoiding duplications of perspective within the same TS as well as modalizations of the illocution.
- the prevalence of SCAs on the other IUs means that information management is more connected to prosodic parsing of the locutive content than to the pragmatic composition of the speech.

The resulting flattening of the information structure entails a sense of monotony in the auditory, which is reflected in prosodic trends. The measures of mean length, speech rate, and f_0 range of IUs and specifically COMs show longer and slower units, a lack of variation in speech rate, and notably, in the pitch variation. This leads to prosodic contours with so slight variation that it may not show any perceptible difference between COMs (t Hart, 1981). This finding also suggests a scarce illocutionary variation in ASD. Further analysis may refine assessing f_0 measures and voice quality (Beccaria et al., 2022; Biancalani et al., 2023).

Moreover, observing pauses, we found more numerous and longer T-pauses in connection to the lower presence of filled pauses, vocalizations, and discourse markers at the beginning of a turn. It is important to note that filled pauses characterize A's speech and serve pragmatic functions in communicative events. Further research considering inter-subjective variability is needed to verify the consistency of this finding.

In addition, the presence and position of filled pauses are linked to the study of speech rhythm with potential insights into vocalic lengthening and the ratio between %V and %C (Maffia et al., 2021).

On the other hand, gestural analysis revealed a quantitative reduction in the subject with ASD, with self-adaptors being more characteristic of his behavior, with poor body involvement when transforming ideation into speech. The quantitative reduction in gesticulation is observed both in the number of GPRSs in proportion to the information units and in the percentage of speech time accompanied by gestures. The opposite quantitative trend is observed for self-adaptors, which strongly characterize the body involvement of individuals with ASD, as expected.

Upon segmenting and studying the video recording, a qualitative difference emerges between the two speakers: hand movements have limited perceptual relevance for the ASD subject, and their contribution to communication is not evident. Therefore, the annotation strategy requires the maximum observation granularity to be effective.

As expected from normative data, gestures tend to synchronize with prosodic units, specifically at prosodic boundaries. However, specific cases of asynchrony (anticipations, delays, or extensions) between semantic anchors and gestures were observed, providing a compelling starting point for future research.

References

- Amir, N., Vered, S.V., Izre'el, S. 2004. Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In: B. Bel, I. Marlien (eds.) *Proceedings of Speech Prosody 2004*, ISCA, 677–680
- Andrén Mats (2010) Children's Gestures from 18 to 30 months, PhD thesis, University of Lund.
- Asperger, H. (1944). Die " Autistischen Psychopathen" im Kindesalter. *Archiv Psychiat Nervenkrankheiten*, 117, 76-136.
- Augustyn Paul R, A, Klin A, Volkmar FR. Perception and production of prosody by speakers with autism spectrum disorders. *J Autism Dev Disord*. 2005 Apr;35(2):205-20. doi: 10.1007/s10803-004-1999-1. PMID: 15909407
- Austin JL. (1962), *How to do things with words*, Oxford, Oxford University Press.
- Beccaria, F., Gagliardi, G., Kokkinakis, D. 2022. Extraction and Classification of Acoustic Features from Italian Speaking Children with Autism Spectrum Disorders. In *Proceedings of the RAPID Workshop - 13th LREC*, pages 22–30, Marseille, France. European Language Resources Association.
- Biancalani, S., Gagliardi, G., Innocenti, M. (2023). Aspetti soprasedimentali e pragmatici dell'eloquio di bambini di età scolare con disturbo dello spettro autistico. Uno studio pilota. In M. Castagneto, M. Ravetto (a cura di) «La comunicazione parlata vercelli 2021», Roma Aracne.

- Breckinridge Church Ruth, Martha W. Alibali & Spencer D. Kelly (eds.), *Why gesture? How the hands function in speaking, thinking and communicating*, 353-377. Philadelphia/Amsterdam: John Benjamins.
- Bressem J, Ladewig SH., Müller C. (2013). Linguistic Annotation System for Gestures (LASG). In Müller C., Cienki A., Fricke E. Ladewig SH., McNeill D, Teßendorf S. (eds.), «Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38)» Vol. 1, Berlin: De Gruyter Mouton, 1098–1125.
- Boersma, P., Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.2.06, retrieved 23 January 2022 from <https://www.praat.org>.
- Cantalini, G. (2022). Corpus multimodale annotato per lo studio della gestualità co-verbale nel “parlato-parlato” e nel “parlato-recitato”. In E. Cresti, M. Moneglia (eds.) «Corpora e Studi Linguistici, Atti del LIV Congresso Internazionale di Studi della Società di Linguistica Italiana (Online, 8-10 settembre 2021)». 135-149.
- Cantalini, G., Moneglia, M. 2020. The annotation of Gesture and Gesture / Prosody synchronization in Multimodal Speech Corpora. *Journal of Speech Science*, V. 9, 1-24
- Cantalini G., Moneglia, M., Gagliardi, G., Proietti, M. 2020. La relazione gesto / prosodia e la sua variabilità. Il parlato spontaneo di contro alla performance attorale. In: A. De Meo, F. Dovetto (eds) *La Comunicazione Parlata*. Roma: Aracne, 63-89.
- Corley, M., & Hartsuiker, R. J. (2003). Hesitation in speech can... um... help a listener understand. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 25(25), 1–10.
- Cavalcante, F. A. (2016). The topic unit in spontaneous American English: A corpus- based study. Belo Horizonte: Federal University of Minas Gerais.
- Chafe, W. (1994). *Discourse, consciousness and time. The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: Chicago University Press.
- Chan, S., Khader, M., Ang, J., Chin, J., & Chai, W. (2016). To behave like a liar: Nonverbal cues to deception in an asian sample. *Journal of Police and Criminal Psychology*, 31, 165-172.
- Chu, N. & Kita, S. (2008) Spontaneous Gestures During Mental Rotation Tasks: Insights Into the Microdevelopment of the Motor Strategy, *Journal of Experimental Psychology General* 137(4):706-23, DOI: 10.1037/a0013157
- Chui K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *J. Prag* 871–887. 10.1016/j.pragma.2004.10.016
- Chui, K., Lee, C. Y., Yeh, K., & Chao, P. C. (2018). Semantic processing of self-adaptors, emblems, and iconic gestures: An ERP study. *Journal of Neurolinguistics*, 47, 105-122.
- Colgan, E. S., Lanter, E., McComish, C., Watson, L. R., Crais, E. R., Baranek, G. T. (2006). Analysis of social interaction gestures in infants with autism. *Child Neuropsychology*, 12, 307-319.
- Cresti, E. (2000). *Corpus di italiano parlato*, Volume I, in «Studi di grammatica italiana pubblicati dall'Accademia della Crusca». Firenze: Accademia della Crusca.
- Cresti, E. (2020), The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory, in Izre'el S., Mello H., Panunzi A., Raso T. (eds.) «In search of basic units of spoken language: A corpus-driven approach», Amsterdam, John Benjamins, 181-219.

- Cresti, E. (2021). The appendix of comment according to language into act theory: corpus-based research. *CHIMERA* 8, 46–69.
- Cresti, E., Moneglia, M. (2018). The illocutionary basis of information structure. *Language into Act Theory (L-Act)*. In E. Adamou, K. Haude, M. Vanhove (eds.) «Information structure in lesser-described languages: Studies in prosody and syntax». 359-401. Amsterdam: John Benjamins.
- Danieli, M., Garrido, J.M., Moneglia, M., Panizza, A., Quazza, S., Swerts, M. (2004) Evaluation of Consensus on the Annotation of Prosodic Breaks in the Romance Corpus of Spontaneous Speech C-ORAL-ROM. In: MT. Lino, MF. Xavier, F. Ferreira, R. Costa, R. Silva (eds) *Proceedings of the 4th LREC Conference*. Paris: ELRA. 1513-1516.
- de Marchena, A., Eigsti, I. (2010). Conversational Gestures in Autism Spectrum Disorders: Asynchrony but not Decreased Frequency. *Autism Research*, 3: 311-322. <https://onlinelibrary.wiley.com/doi/full/10.1002/aur.159>
- Duffy, C., Healy, O. (2011). Spontaneous Communication in Autism Spectrum Disorder: A Review of Topographies and Interventions. *Research in Autism Spectrum Disorders*, 5, 977-983.
- Eigsti, I., Schuh, J., Mencl, E., Schultz, R., Paul, R. (2011). The neural underpinnings of prosody in autism. In *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*.
- ELAN (Version 6.7) [Computer software]. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/ta/elan>
- Esteve-Gibert, N., & Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 563, 850–864. <https://doi.org/10.1044/1092-43882012/12-0049>.
- Filipe MG, Frota S, Castro SL, Vicente SG. Atypical prosody in Asperger syndrome: perceptual and acoustic measurements. *J Autism Dev Disord*. 2014 Aug;44(8):1972-81. doi: 10.1007/s10803-014-2073-2. PMID: 24590408.
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, 29(2), 320–326.
- Froiland, JM., Davison, ML. (2016) Home literacy, television viewing, fidgeting and ADHD in young children. *Educational Psychology*, 36:8, 1337-1353, DOI: 10.1080/01443410.2014.963031
- Gorman, K., Olson, L., Hill, A. P., Lunsford, R., Heeman, P. A., & van Santen, J. P. (2016). Uh and um in children with autism spectrum disorders or language impairment. *Autism Research*, 9(8), 854–865.
- Graziano M., Nicoladis E., Marentette P. (2020). How Referential Gestures Align With Speech: Evidence From Monolingual and Bilingual Speakers. *Lang. Learn.* 70 266–304. 10.1111/lang.12376
- Fusaroli, R., Lambrechts, A., Bang, D., Bowler, DM., Gaigg, SB. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, 10, 384–407. DOI:10.1002/aur.1678
- 't Hart, Johan (1981). "Differential sensitivity to pitch distance, particularly in speech". *The Journal of the Acoustical Society of America*. 69 (3): 811–821.
- t' Hart, J., Collier, R., Cohen, A. (1990). *A perceptual study on intonation. An experimental approach to speech melody*, Cambridge, Cambridge University Press.

- Irvine, C. A., Eigsti, I. M., & Fein, D. A. (2016). Uh, um, and autism: Filler disfluencies as pragmatic markers in adolescents with optimal outcomes from autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 46(3), 1061–1070.
- Izre'el S., Mello H., Panunzi A., Raso T. (eds.) 2020 «In search of basic units of spoken language: A corpus-driven approach», Amsterdam, John Benjamins.
- Izre'el, S., Mettouchi, A. 2015. Representation of Speech in CorpAfroAs. Transcriptional Strategies and Prosodic Units. In: A. Mettouchi, M. Vanhove, D. Caubet (eds.) *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Amsterdam: Benjamins. 13–41.
- Janke, V., & Perovic, A. (2017). Advanced syntax and primary pragmatics in children with ASD. In L. R. Naigles (Ed.), *Innovative investigations of language in autism spectrum disorder* (pp. 141–161). Walter de Gruyter GmbH; American Psychological Association. <https://doi.org/10.1037/15964-008>
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2(3), 217-250.
- Kendon, A. (2004). *Gesture*. Cambridge.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162-185). Cambridge Cambridge University Press
- Kita, S., van Gijn I., van der Hulst H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In Wachsmuth I., Fröhlich M. (eds), «*Gesture and Sign Language in Human-Computer Interaction*» Berlin: Springer, 23–35.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16–32.
- Ladewig SH, Bressemer J. A linguistic perspective on the notation of gesture phases. In Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Teßendorf S. (eds), «*Body - Language - Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38)*» Vol. 1, Berlin: De Gruyter Mouton, 2013, 1060–1079.
- Lausberg, H. 2013. *Understanding Body Movement. A Guide to Empirical Research on Nonverbal Behaviour, With an Introduction to the NEUROGES Coding System*. Frankfurt am Main: Peter Lang,
- Lin, W., Orton, I., Li, Q., Pavarini, G., & Mahmoud, M. (2021). Looking at the body: Automatic analysis of body gestures and self-adaptors in psychological distress. *IEEE Transactions on Affective Computing*.
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7, 179–214
- Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 31., 71–89. <https://doi.org/10.1515/lp-2012-0006>
- Loehr, D. (2014). Gesture and prosody. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, J. Bressemer (eds.), «*Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2*» 1381-1391. de Gruyter. doi.org/10.1515/9783110302028.1381
- Lord, C., Paul, R. (1997). Language and communication in autism. In DJ. Cohen, FR. Volkmar (eds.) «*Handbook of autism and pervasive developmental disorders (2nd ed.)*» 195–225. New York, NY: Wiley.

- Lord, C., Rutter, M., & LeCouteur, A. (1994). 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*, 24, 659–685.
- Lord, C., Rutter, M., DiLavore, P.C., Risi, S. (2002). *Autism diagnostic observation schedule (ADOS)*. Los Angeles: Western Psychological Services.
- Mahmoud, M., Morency, L. P., & Robinson, P. (2013, December). Automatic multimodal descriptors of rhythmic body movement. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 429-436).
- Maffia, M., De Micco, R., Pettorino, M., Siciliano, M., Tessitore, A., De Meo, A. (2021). Speech Rhythm Variation in Early-Stage Parkinson's Disease: A Study on Different Speaking Tasks. *Front. Psychol., Psychology of Language*, Volume 12. <https://doi.org/10.3389/fpsyg.2021.668291>
- Martin, P. (2004). WinPitch Corpus: a text to speech alignment tool for multimodal corpora. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisboa, Portugal, European Language Resources Association (ELRA), 537–540.
- Martin, P. 2009. *Intonation du Français*. Paris: Armand Colin.
- Mehrabian, A., & Friedman, S. L. (1986). An analysis of fidgeting and associated individual differences. *Journal of Personality*, 54(2), 406-429.
- McCann et al., 2007;
- McGregor, K. K., & Hadden, R. R. (2020). Brief Report: “Um” fillers distinguish children with and without ASD. *Journal of Autism and Developmental Disorders*, 50(5), 1816–1821.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago
- Mehrabian, A., & Friedman, S. L. (1986). An analysis of fidgeting and associated individual differences. *Journal of Personality*, 54(2), 406-429.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
- Miller, J., Weinert, R. 1998. *Spontaneous Spoken Language. Syntax and Discourse*. Oxford: Clarendon Press.
- Moneglia, M. (2005). The C-ORAL-ROM Resource. In: Cresti, E., Moneglia, M. (eds).. C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages, pp. 1-70, AMSTERDAM: Benjamins. DOI: <https://doi.org/10.1075/scl.15.03mon>
- Moneglia, M., Raso, T., Malvessi-Mittmann, M., Mello, H. 2010. Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL / C-ORAL-ROM. In: *Speech Prosody 2010*, W1.09, Satellite workshop on Prosodic Prominence: Perceptual, Automatic Identification. Chicago. https://www.isca-speech.org/archive/sp2010/sp10_2010.html, 2010
- Moneglia, M., Raso, T. (2014). Notes on Language into Act Theory (L-AcT). In T. Raso, H. Mello (eds.), «Spoken Corpora and Linguistic Studies» 468–495. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.61.15mon>
- Mundy Peter, Marian Sigman, Judy Ungerer, Tracy Sherman (1986) DEFINING THE SOCIAL DEFICITS OF AUTISM: THE CONTRIBUTION OF NON-VERBAL COMMUNICATION MEASURES 27, 5, 657-669
- Neff, M., Toothman, N., Bowmani, R., Fox Tree, J. E., & Walker, M. A. (211). Don't scratch! Self-adaptors reflect emotional stability. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings* 11 (pp. 398-411). Springer Berlin Heidelberg.

- Panunzi A., Gregori L. (2012), *DB-IPIC. An XML database for the representation of information structure in spoken language*, in Panunzi A., Raso T., Mello H. (a cura di) «Pragmatics and prosody. Illocution, modality, attitude, information patterning and speech annotation», Firenze, Firenze University Press.
- Panunzi, A., Gregori, L., Rocha, B. 2020. Comparing annotations for the prosodic segmentation of spontaneous speech: Focus on reference units. In: S. Izre'el, H. Mello, A. Panunzi, T. Raso (eds.) *In Search of Basic Units of Spoken Language. A corpus-driven approach*. Amsterdam: Benjamins. 403-431
- Rescorla, Leslie A., and Paige Safyer. "Lexical Composition in Children with Autism Spectrum Disorder (ASD)." *Journal of Child Language* 40, no. 1 (2013): 47-68, doi: 110.1017/S0305000912000232.
- Robins, D.L., Fein, D., Barton, M.L., & Green, J.A. (2001). The Modified Checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 31, 131–144.
- Rohrer, P. L., Prieto, P. & Delais-Roussarie, E. 2019. Beat gestures and prosodic domain marking in French. In: S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds) *Proceedings of the 19th International Congress of Phonetic Sciences*. Australasian Speech Science and Technology Association Inc. 1500-1504.
- Rohrer, P. L. 2022. A temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel MultiModal MultiDimensional M3D labeling system. Ph.D. thesis Barcelona: University Pompeu Fabra
- 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., Panunzi, A., (2020), Le unità di comment multiplo. Analisi secondo la Teoria della Lingua in Atto, in De Meo/Dovetto (a cura di) «La comunicazione parlata. Atti del Congresso SLI–GSCP – Università degli Studi di Napoli “L’Orientale” (Napoli, 12-14 dicembre 2018)», Napoli, Aracne; 263-286.
- Saccone, V., Panunzi, A. (2023), Strutture parentetiche nel parlato italiano: classificazione funzionale e identificazione prosodica, in M. Castagneto, M. Ravetto (eds.) «La comunicazione parlata. Atti del Congresso SLI–GSCP – Università del Piemonte Orientale (Vercelli, 5-7 maggio 2021)», Napoli, Aracne.
- Saccone, V., Trillocco, S. (2022) Segmentation of the Speech Flow for the Evaluation of Spontaneous Productions in Pathologies Affecting the Language Capacity. A Case Study of Schizophrenia, in *Proceedings of the RAPID-4 @LREC 2022, Marseille, 94–99*, © European Language Resources Association (ELRA), licensed under CC-BY-NC 4.0.
- Saccone, V., Trillocco, S., Moneglia, M. (2023). Markers of Schizophrenia at the Prosody/Pragmatics interface. Evidence from corpora of spontaneous speech interactions, in *Front. Psychol.*, Sec. Psychology of Language, Volume 14. doi.org/10.3389/fpsyg.2023.1233176
- Saccone, V. Trombetta, C. 2021. Parenthetical Units and Structures in Italian and German spoken language Prosodic and textual analysis. *CHIMERA. Romance Corpora and Linguistic Studies*. V.8, 1-23. <https://doi.org/10.15366/chimera2021.8.001>
- Saccone, V., Vieira, M., Panunzi, A. (2018), Complex Illocutive Units in Language into Act Theory: an analysis of non-terminal prosodic breaks of Bound Comments and Lists in *JoSS Special Issue: Spoken Corpora advances: prosody as the crux of speech segmentation, annotation and multilevel linguist*, State University of Campinas; 7(2); 51-64.

- Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology, 9*, Article 1514. <https://doi.org/10.3389/fpsyg.2018.01514>
- Shattuck-Hufnagel, S., Ren, P. L., & Tauscher, E. 2010. Are torso movements during speech timed with intonational phrases? In: *Proceedings of the International Conference on Speech Prosody*. ISCA Archive. 1–4.
- So, W., Wong, M.K., Lui, M. & Yip, V. (2015). The development of co-speech gesture and its semantic integration with speech in 6- to 12-year-old children with autism spectrum disorders. *Autism, 19*(8); 956-968. DOI: 10.1177/1362361314556783
- Sparaci, L. (2008). Embodying gestures: The Social Orienting Model and the study of early gestures in autism, *Phenom Cogn Sci 7*, 203–223 DOI 10.1007/s11097-007-9084-9
- Sparaci, L., Lasorsa, FR., Capirci, O. (2019). More Than Words: Gestures in Typically Developing Children and in Children with Autism. In Grove, N., Launonen, K. (eds.) «Manual Sign Acquisition in Children with Developmental Disabilities», , NOVA Science Publishers.
- Swerts, M., & Kraemer, E. 2010. Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics, V. 38*, 197–206. <https://doi.org/10.1016/j.wocn.2009.10.002>
- Wetherby, AM., Prutting, CA. (1984). Profiles of communicative and cognitive-social abilities in autistic children. *Journal of Speech and Hearing Research, 27*, 364–377.
- Wing, L. (1981) Asperger's syndrome: A clinical account. *Psychological Medicine 11*:115–29.