Studi Italiani di Linguistica Teorica e Applicata, anno LII, 2023, numero 2

Gloria Gagliardi, Fabio Tamburini — *Università di Bologna*
Milvia Innocenti — *AUSL Toscana Centro*

CAN SPEECH-BASED MEASURES SUPPORT DEVELOPMENTAL LANGUAGE DISORDER IDENTIFICATION? AN EXPLORATIVE STUDY

Abstract

This paper aims to identify possible phonetic biomarkers of *Developmental Language Disorder* (DLD) in Italian preschool children.

Speech samples, collected during three retelling tasks, were transcribed and processed through a computational pipeline. A set of acoustic and rhythmic parameters were automatically extracted from the recordings and analysed by using descriptive and inferential statistics.

Our work demonstrates that i) language difficulties of DLD take the form of reduced fluency and speech disruptions ii) some acoustical characteristics of the voice (e.g., the mean value of the fundamental frequency) can distinguish language-impaired children from peers. These results suggest that automatic voice and speech analysis could provide new markers of the DLD - markers that are not audible to the human ear and therefore fall outside the possibilities of conventional paper-and-pencil neuropsychological tests.

Keywords: Developmental Language Disorder, Italian preschoolers, speech analysis, NLP techniques, digital linguistic biomarker

1. Introduction

1.1 Developmental Language Disorder: definition and diagnosis

The identification of developmental difficulties in the child's acquisition pathway of language is a core activity for paediatricians, neuropsychologists, and speech therapists (Reilly *et al.* 2015). Linguistic problems are quite common in the early developmental stages: by 30 to 36 months of age up to 17.5 %

of toddlers are "late talkers", namely, experience a late language emergence in absence of developmental delays in other cognitive or motor domains (Horowitz *et al.* 2003). Only a subset of these children (approximately 50% to 70%, called "late bloomer" in the scientific literature) catch up to their peers in language skills by the age of three to five without therapeutic intervention (Paul *et al.* 1996; Dale *et al.* 2003). The remaining children are at risk for developing language and/or literacy difficulties as they grow up (McKean *et al.* 2017).

In particular, at the age of four, they can be diagnosed with Developmental Language Disorder (American Psychiatric Association 2013; Bishop *et al.* 2017; Sansavini *et al.* 2021), when language skills are persistently below the level expected. In this clinical condition, language deficits occur in the absence of a known biomedical condition (e.g., autism spectrum disorder or Down syndrome), and hearing or sensory-motor impairments, interfering with the child's ability to communicate effectively with caregivers and peers.

Since DLD can negatively impact children's psychosocial outcomes and academic attainment (Redmond & Timler 2007a; 2007b; Redmond 2011), early identification and intervention are crucial to avert long-term repercussions in emotional well-being in adult life (Bortolini *et al.* 2006). The evaluation of possible DLD must be addressed by a multi-professional healthcare team (Marotta & Caselli 2014; O'Hare & Bremner 2016), and all language competencies should be investigated through both quantitative tests and qualitative analysis, at the receptive and expressive levels.

In this framework, the usage of "clinical markers", namely measurable and quantifiable parameters which can support health- and cognitive-related assessments, plays a crucial role in the timely identification of the disorder. However, as stated by Bortolini *et al.* (2006), the search of clinical markers for this neurodevelopmental condition is extremely challenging for at least three reasons: (i) most salient symptoms, behavioural in nature, usually differ from typical language trajectories in degree, not kind; (ii) signs may vary over time, therefore some symptoms suggesting impairment at an older age might be quite unremarkable at a younger age; (iii) through across-linguistic comparison appears that despite the markers of DLD have some common denominators (Leonard 2014) they are largely language-specific. Moreover, a considerable variability characterizes the clinical presentation of the disorder, and a huge disagreement amongst professionals arises concerning how to interpret it, complicating both understanding the nature of this condition and the adoption of the best clinical framework for the diagnosis and treatment.

To date, linguistic markers for the identification of DLD in Italian preschoolers are non-word repetition and the production of direct-object clitic pronouns, with 90% sensitivity and 100% specificity (Bortolini *et al.* 2006; Arosio *et al.* 2014). These linguistic elements are usually investigated by clinicians in standardized settings, with conventional batteries for language assessment, within a broader evaluation process of communicative skills.

1.2 Quantitative linguistic methods and NLP techniques for pathology detection, monitoring and treatment

In recent years, the field of health research has shown a burgeoning increment in the use of Natural Language Processing (NLP) strategies and methods (Velupillai *et al.* 2018).

A growing body of scientific evidence proved that quantitative linguistic features, easily extractable from a patients' verbal productions through NLP techniques, can be very useful in distinguishing subjects with various cognitive impairments from healthy controls, even at a very early stage. Cognitive frailty assessment is the most intensively studied domain (Martínez-Nicolás *et al.* 2021), but this approach is gaining increasing popularity, with encouraging results also in the field of developmental pathologies (Solorio 2013). In this sense, "subtle language disruptions can be employed as digital linguistic biomarkers, namely objective, quantifiable behavioural data that can be collected and measured by means of digital devices, allowing for a low-cost pathology detection, classification, and monitoring" (Gagliardi *et al.* 2021). This approach shows several advantages compared to classical pen-and-paper neuropsychological tests: among these, its administration is completely non-intrusive and not prone to human-rater bias. Moreover, this technology can support the discovery of novel validated indexes for making prompt and more accurate diagnoses, that go beyond current signs-based criteria, also allowing for low-cost monitoring of patients over time.

Computer-based speech therapy systems (or virtual speech therapists - VSTs) are also spreading for the rehabilitation of children and adults with speech, hearing, or language disorders. These instruments usually combine NLP and AI technologies, such as automatic speech recognition (ASR) and speech synthesis algorithms, to target a specific communication deficit by adopting a predefined therapy program (Chen *et al.* 2016).

Even though more rigorous evaluation practices and large benchmarks are needed to improve reporting standards and results, NLP methods can be very impactful in health outcomes research.

1.3 Automatic Voice analysis: a promising candidate as DLD digital biomarker

The assessment of speech, at the segmental and suprasegmental level, is extremely important for the diagnosis and remediation of speech and language disorders in childhood. This is particularly true for DLD children, who were reported to show difficulties in the processing of rhythmic structure, both in music and prosody (Van Der Meulen *et al.* 1997; Goffman 1999; Wells & Peppé 2003; Caccia & Lorusso 2021) and subtle motor deficits (Diepeveen *et al.* 2018).

Many computational instruments have been developed, in recent years, to assess verbal productions of children (van Santen *et al.* 2009; Dudy *et al.* 2018), but a lot of relevant issues still need to be answered.

In particular, most of the works focused exclusively on Autism Spectrum Disorders - ASD, both describing the acoustical and perceptive characteristics of the voice (McCann & Peppé 2003; Broome *et al.* 2017; Fusaroli *et al.* 2017), such as pitch and speech spectrum, and trying to detect this neurodiversity from vocal cues (Van Santen *et al.* 2010; Tanaka *et al.* 2014; Bone *et al.* 2015; Parish-Morris *et al.* 2016; Baird *et al.* 2017; Deng *et al.* 2017; Pokorny *et al.* 2017; Lin *et al.* 2018; Kumar *et al.* 2018; Cho *et al.* 2019; Hauser *et al.* 2019; Li *et al.* 2019; Saturday & Kabari 2020). The literature does not report conclusive evidence for a single acoustic marker of ASD (Fusaroli *et al.* 2017), nevertheless, the classification process reaches fair results.

Less attention has been paid to neurodevelopmental disorders other than ASD, including Developmental Language Disorder, and most of the study put in comparison the two clinical populations. For example, Kiss *et al.* (2012) analysed the prosody of both children with ASD and DLD compared to typically developing peers, founding several significant differences in the pitch characteristics of these diagnostic groups, and achieving performances of classification well above chance level. Very interestingly, they did not find significant differences between the ASD and DLD subgroups, suggesting that some of the atypical speech alterations may not be specific to autism but instead may be part of the phenotype of several disorders or may indicate comorbidity.

In a similar way, a sample of DLD children was included in the Autism Sub-Challenge of the INTERSPEECH 2013 COMputational PARalinguistics ChallengE (ComParE), the most relevant shared task on the topic that engaged the NLP community over the last decade (Schuller *et al.* 2013; 2019). The participants were requested to determine the type of pathology of a speaker by using a suited classification algorithm and several acoustic features. The task was conduct on the corpus “Child Pathological Speech Database” (CPSD) (Ringeval *et al.* 2011), which contains 2542 speech recordings uttered by 99 French children (age range 6-18), of which 35 with Developmental Disorder (Autism, DLD, or Pervasive Development Disorder non-otherwise specified-NOS) and 64 controls. Even though the ComParE Initiative allowed to overcome comparability issues regarding data sets, evaluation measures, baseline systems, and testbeds, the systems generally achieved mediocre/fairly good results (best Unweighted Average Recall - UAR: 69.4%) (Asgari *et al.* 2013; Bone *et al.* 2013; Gosztolya *et al.* 2013; Kirchhoff *et al.* 2013; Lee *et al.* 2013; Martínez *et al.* 2013; Räsänen 2013). Moreover, the employed speech data is not spontaneous, since the corpus contains prompted imitation of 26 sentences with different modalities (e.g., declarative, interrogative) and four types of intonation (i.e., descending, falling, floating, and rising): therefore, these applications cannot be readily exported to actual clinical settings.

To the best of our knowledge, only a few recent studies specifically tackled the identification of DLD. Ramarao *et al.* (2018) applied Extreme Learning Machine (ELM) to discriminate typically developing from DLD children, using Gaussian Posteriograms learned on Mel-frequency cepstral coefficients (MFCCs). The system reached an accuracy of 99.41%. Reddy *et al.* (2020) proposed an approach based on time- and frequency- domain glottal parameters, Mel-frequency cepstral coefficients, and acoustic parameters extracted by openSMILE toolkit (Eyben *et al.* 2013), getting 98.82% of accuracy by employing neural network classifiers. Finally, Sharmaand Singh (2020) evaluated the distinction proficiency of pitch-based features (extracted from children's phonation of the vowel /a/) to classify DLD and typical peers, exerting a k-NN classifier with Neighborhood Component Analysis (NCA) approach to perform the task. The model pledges an accuracy of 97.93%.

These last results, while isolated, are very encouraging in view of designing a computer-aided diagnostic system for DLD and other Neurodevelopmental Disorders.

2. Rationale of the study

This work aims to identify subtle voice alterations in the speech of Italian children, which can support the diagnosis of DLD. To this purpose, the verbal productions of preschoolers diagnosed with language impairment and aged-matched peers were analysed through NLP techniques, extracting acoustic and rhythmical features.

This explorative study represents the first step in the wake of discovering digital linguistic biomarkers supporting the diagnosis of DLD among Italian children: as a matter of fact, a large amount of scientific work has been devoted to the description and automatic analysis of atypical prosody in the Autism Spectrum Disorders, while other Neurodevelopmental Impairments remain largely unexplored.

Moreover, as far as we know, this is the first attempt to perform such kind of analysis on Italian-speaking children: this is a quite relevant issue since prosodic, rhythmical, and typological peculiarities may strongly affect the discriminative power of the speech parameters in different languages, strongly limiting the transferability of the findings.

3. Material and Methods

3.1 Participants

We enrolled sixteen monolingual infants (13 M; 3 F) ranging in age from 4;2 to 5;4 (μ=4;7). The sample is composed of a Control Group (CG) and a DLD Group, matched by age. The CG includes eight children (5 M; 3 F) without speech, language, hearing, or cognitive impairments. The DLD group includes eight male peers who met the criteria for DLD with expressive deficits (American Psychiatric Association 2013). DLD children were recruited at the Local Health Centre Azienda USL Toscana Centro and, at the time of the study, were all receiving clinical services. CG was recruited from kindergartens in the same residential areas as the DLD group (i.e., the Metropolitan City of Florence, in the Tuscany Region).

The cohorts are homogeneous for age and geographical provenience, but not for gender: the rationale behind this choice is due to the prevalence of neurodevelopmental disorders – and, in particular, DLD – among gender groups. There is growing evidence that being male double the risk of language impairment (Tomblin *et al.* 1997). However, little effort has been devoted to the detailed account of the epidemiology of the disorder and, to date, the reason for the sex difference is not well understood. Given this complex picture, and also considering that standardized linguistic tests are normed on the general population, we included several females in the CG.

It is important to notice that gender difference does not affect the validity of our results. Our sample is composed by preschoolers, with a maximum age of 5;4: the maturation of the phonatory structures and the development of adequate neuromuscular control starts later, around the age of six, in both sex (Cappellari & Cielo 2008). A large amount of scientific literature supports that the acoustic voice profile of children was uniform across all girls and boys younger than 10 years (Campisi *et al.* 2002; Pribuisiene *et al.* 2011). Moreover, as demonstrated by Hacki & Heitmüller (1999), the only gender-specific voice parameter for the considered age range is intensity (namely, the habitual intensity and the maximum speaking voice intensity, ±2 dB).

The diagnosis of DLD has been established according to national and international guidelines by expert clinicians, by considering anamnestic data, clinical observation, and standardized testing. Participants underwent a complete language evaluation, but particular attention has been paid to the assessment of children's comprehension profile: all subjects performed within the normal range on the test of receptive vocabulary (TNL, Test Neuropsicologico Lessicale per l'età evolutiva (Cossu 2013)), morpho-syntactic comprehension (TCGB, Test di Comprensione Grammaticale per Bambini (Chilosi & Cipriani 2006) and PVCL, Prove di Valutazione della Comprensione Linguistica (Rustioni & Lancaster 2007)) and listening comprehension (TOR, Test di Comprensione del

Testo Orale 3-8 anni (Levorato & Roch 2007)); therefore, expressive language problems occur essentially in isolation.

As already stated, all the children of the DLD group underwent extensive speech-language treatment before the study. A huge number of papers have been published on the linguistic profiles of preschoolers with DLD, but less attention has been paid to their outcomes after rehabilitation. In our opinion, this particular circumstance can represent a fruitful window on distinguishing and persistent difficulties of the disorder. *Table 1* outlines the main characteristics of our cohorts.

All ethical principles of the Helsinki Declaration were followed.

	controls	DLD
Age	μ = 4;7 σ = 0;4	μ= 4;7 σ = 0;4
Sex	5 M; 3 F	8 M; 0 F
geographical provenience	Toscana (Florence Area), Italy	Toscana (Florence Area), Italy

Table 1: Main characteristics of the cohorts.

3.2 Corpus design and speech recording

Verbal production was elicited using three different tasks: a norm-referenced linguistic task, namely the Italian version of the Bus Story Test (I-BST) (Cipriani *et al.* 2012; Renfrew 2015; Mozzanica *et al.* 2016), and two semi-spontaneous retelling, exploiting the renowned story Three Little Pigs (3LP), and a brand-new short film called Little Polar Bear (LPB). While the BST examines story retelling with coloured picture support, the unnormed tests elicit children's verbalizations through a paper book and a tablet respectively. During the 3LP task, children were asked to retell the renowned story using the pictures as prompts while flipping through the pages; in contrast, the LPB task was administered showing the video (around 100 seconds) to the child who was then requested to recount the plot while following the scrolling images without sound. None of the children knew the three stories, including the generally well-known 3LP.

The trials were administered in a single test session of varying duration. The tasks have been videotaped using a tablet placed in front of the subject.

3.3 Corpus transcription and annotation

Data was orthographically transcribed using ELAN (Wittenburg *et al.* 2006) (*fig. 1*). All parents gave their consent to data recording and processing. Orthographical transcription is compliant with the L-AcT format (Cresti & Moneglia 2018), a version of the standardized CHAT format (MacWhinney 2000) enriched with the tagging of prosodic parsing. For a detailed account of transcription methods and conventions, please refer to Beraldi *et al.* (2021). The resulting corpus consists of 1 h 57'41" of recorded speech; the verbal productions amount to a total of 4551 words, 889 utterances. *Table 2* shows the number of utterances and words for each task produced by controls and DLD children.

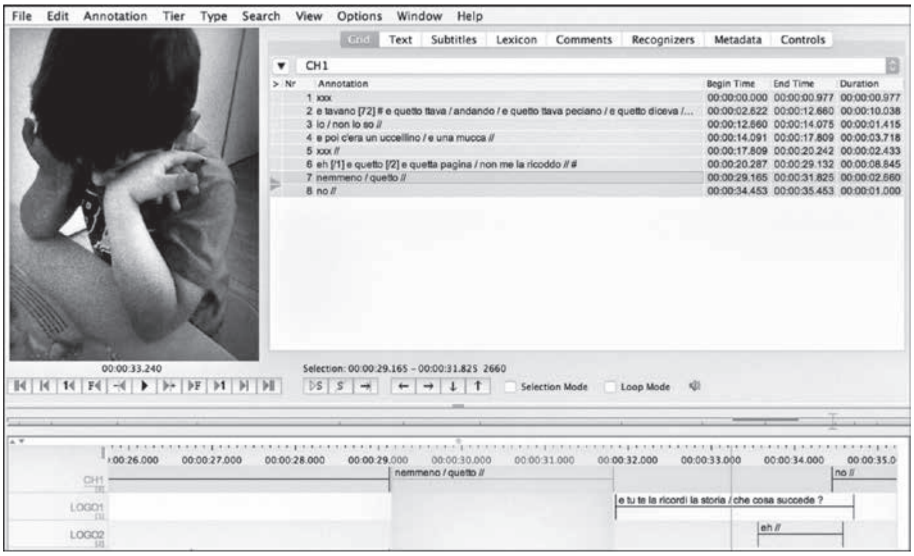

Figure 1: A screenshot of the ELAN annotation tool.

Task	CG		DLD	
	# utterances	# words	# utterances	# words
I-BST	133	833	170	802
3LP	186	978	183	915
LPB	99	567	118	456
TOTAL	418	2378	471	2173

Table 2: Number of utterances and words produced on the three tasks by controls and DLD children.

3.4 Features extraction and data processing

We applied the set of algorithms developed in the wake of the OPLON project (Beltrami *et al.* 2018; Calzà *et al.* 2021; Gagliardi & Tamburini 2022) to extract acoustic and rhythmic parameters from each session. *Tables 3* outlines the list and the description of the linguistic parameters considered in this study. We selected these cues because they showed to be relevant in supporting the identification of speech disruptions in adult Italian speakers (Calzà *et al.* 2021; Gagliardi & Tamburini 2021, Gagliardi *in press*). Feature names ending with '_M' refer to the mean value, those ending with '_MD' refer to the median and those ending with '_ SD' to the standard deviation of the feature calculated on the whole speech production. For a comprehensive description, the reader is referred to Calzà *et al.* (2021).

Features	Description and References
Silence segments duration (SPE_SIL)	Silence segments of the signal (cf. Satt *et al.* 2013).
Speech segments duration (SPE_SPE).	Speech segments of the signal (cf. Satt *et al.* 2013).
Temporal regularity of voiced segments (SPE_TRVS)	It captures the temporal structure of the voiced segments, providing information on the rate of change in the different spectrum bands (cf. Satt *et al.* 2013).
Verbal Rate (SPE_VR)	The number of tokens in the sample divided by the Total Locution Time (i.e., speech time including pauses) (cf. Singh *et al.* 2001; Roark *et al.* 2011).
Transformed Phonation Rate (SPE_TPR)	It measures the ratio between the total phonation time (i.e., speech time without pauses) and the total locution time (i.e., speech time including pauses) (cf. Singh *et al.* 2001; Roark *et al.* 2011).
Standardized Phonation Time (SPE_SPT)	The number of tokens in the sample divided by the total phonation time (i.e., speech time excluding pauses) (cf. Singh *et al.* 2001; Roark *et al.* 2011).
Standardized Pause Rate (SPE_SPR)	The number of tokens in the sample divided by the number of pauses (cf. Singh *et al.* 2001; Roark *et al.* 2011).
Root Mean Square energy (SPE_RMSE)	The energy of a signal at a specific time, calculated by windowing the signal, squaring the samples and taking the average. The square root of this result is the engineering quantity known as RMS (cf. Lòpez-de-Ipiña *et al.* 2013).
Pitch (SPE_PITCH)	Fundamental frequency (F0)
Spectral Centroid (SPE_SPCENTR)	It is calculated by evaluating the “centre of gravity” of the spectrum using the Fourier transform’s frequency and magnitude information. The feature captures the perceptual “brightness” of a sound (cf. Lòpez-de-Ipiña *et al.* 2013).
Higuchi Fractal Dimension (SPE_HFractD)	The algorithm measures fractal dimension (i.e., “self-similarity”, identical/similar structures repeating over a pattern) of discrete time sequences directly from time series. The feature describes the complexity of the signal at a specific time (cf. Lòpez-de-Ipiña *et al.* 2013).
Percentage of vocalic intervals (RHY%V)	The proportion of vocalic intervals within the utterance, namely the sum of vocalic intervals divided by the total duration of the utterance (cf. Ramus *et al.* 1999).
Std. deviation of vocalic and consonantal interval durations, ΔV and ΔC (RHYDeltaV, RHYDeltaC).	The standard deviation of the duration of vocalic and consonantal intervals within each utterance (cf. Ramus *et al.* 1999)
Pairwise Variability Index (RHY_VnPVI, RHY_CrPVI)	This rhythm metric takes into account the temporal succession of the vocalic and consonantal intervals. It is based on a pairwise comparison of the duration of either two vocalic or consonantal intervals, expressing the level of variability in consecutive measurements (cf. Grabe & Low 2002).
Variation coefficient for ΔV and ΔC (RHY_VarcoC, RHY_VarcoV)	A variation coefficient (“varco”) is a value describing relative variation. In particular, VarcoΔC is calculated as the percentage of the ΔC of the average duration of consonantal intervals (Delwo 2006).

Table 3: Acoustic and rhythmic features considered in the study.

Results were analysed by using descriptive and inferential statistics, employing R statistical software version 3.4.4. Because of the small sample size, we applied the non-parametric Wilcoxon-Mann-Whitney test (two-sided) with the Bonferroni correction to compare the groups (CG and DLD). Threshold (p-) values are statistically significant at <0.05.

4. Results

Table 4 outlines the different levels of significance for the considered linguistic features. Given the small size of the corpus, we did not distinguish among the three tasks (I-BST, 3LP, LPB).

Feature	CG $\mu \pm \sigma$	DLD $\mu \pm \sigma$	MW-test p-value	
Spe_SilMean	1.56 ± 0.81	1.69 ± 0.69	0.193	
Spe_SilMedian	1.15 ± 0.53	1.30 ± 0.54	0.05523	
Spe _SilSd	1.23 ± 1.07	1.33 ± 0.82	0.3185	
Spe _SpeMean	1.18 ± 0.44	1.17 ± 0.39	0.9533	
Spe _SpeMedian	0.86 ± 0.38	0.88 ± 0.35	0.8595	
Spe_SpeSd	1.15 ± 0.55	1.10 ± 0.47	0.8606	
Spe_TrvSd	0.42 ± 0.19	0.48 ± 0.19	0.3104	
Spe_VR	0.98 ± 0.35	0.66 ± 0.40	0.004652	**
Spe_TPR	0.85 ± 0.22	0.80 ± 0.19	0.2719	
Spe_SPT	1.90 ± 0.88	1.40 ± 0.87	0.01959	*
Spe_SPR	1.97 ± 0.54	1.53 ± 0.97	0.01417	*
Spe_RmseM	557.41 ± 353.59	625.01 ± 472.87	0.7443	
Spe _RmseSd	382.06 ± 166.98	450.62 ± 267.00	0.3437	
Spe_PitchM	269.02 ± 21.76	246.761 ± 33.00	0.006131	**
Spe_PitchSd	74.08 ± 22.44	69.719 ± 25.92	0.2793	
Spe_SpCentM	2751.26 ± 507.29	2545.32 ± 487.12	0.1762	
Spe_SpCentSd	1295.62 ± 275.42	1214.48 ± 304.60	0.4978	
Spe_HFractDM	1.87 ± 0.03	1.86 ± 0.03	0.2172	
Spe_HFractDSd	0.08 ± 0.02	0.08 ± 0.02	0.7955	
Rhy_%V	0.45 ± 0.07	0.45 ± 0.08	0.9658	
Rhy_DeltaV	0.21 ± 0.04	0.20 ± 0.06	0.9648	
Rhy_DeltaC	0.26 ± 0.07	0,27 ± 0.06	0.7044	
Rhy_VnPVI	0.83 ± 0.09	0.84 ± 0.08	0.6123	
Rhy_CrPVI	0.21 ± 0.07	0.20 ± 0.06	0.8167	
Rhy_VarcoV	142.20 ± 17.06	142.68 ± 16.76	0.9087	
Rhy_VarcoC	143.68 ± 14.72	146.35 ± 14.59	0.5192	

Table 4: Linguistic features considered in this study (refer to Table3 for the descriptions and abbreviations): results are expressed as means ± standard deviations, and statistical significance at the Wilcoxon-Mann-Whitney test is also reported (* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001).

Our results show that DLD children emerged as less fluent in comparison to their peers. As a matter of fact, the significant parameters of our study are the Verbal Rate, the Standardized Phonation Time, and the Standardized Pause Rate. In other words, from a qualitative point of view, their spoken texts are richer with hesitation phenomena, e.g., filled and empty pauses. In line with the scientific literature on the topic (and confirming the findings of Beraldi *et al.* 2021, obtained by traditional phonetic analysis), our work demonstrates that i) even after therapeutic remediation, language difficulties remain, taking the form of reduced fluency or speech disruptions during utterance formulation (Hall 1996; Finneran *et al.* 2009) and ii) these subtle alterations can be easily detected through NLP algorithms. Therefore, we can easily assume that these cues can represent a distinctive trait of the disorder also at the earlier stages of linguistic development, before the logopedic intervention.

But the more interesting point is the statistical relevance of the pitch, namely the mean value of the Fundamental Frequency of the voice: as attested by Ramarao *et al.* (2018), Reddy *et al.* (2020) and Sharma & Singh (2020), acoustic features of the speech can represent new markers of neurodevelopmental disorders - markers that are not quantifiable by the therapist without instrumental support, and therefore fall outside the possibilities of conventional paper-and-pencil neuropsychological tests. On the contrary, none of the rhythmic features emerges as statistically significant: this finding deserves further investigation, to ascertain if the result is due to the logopedic treatment, or the actual sparing of the articulation mechanism (e.g., vs. Childhood Apraxia of Speech and/or Speech Sound Disorder).

5. Concluding Remarks

This paper presented an explorative study on the identification of speech parameters able to support the diagnosis of DLD among Italian preschoolers. To this aim, we applied a computational pipeline developed by Gagliardi & Tamburini (2022) for the automatic extraction of acoustic and rhythmic features from children's verbal productions. We identified, through inferential statistics, a set of acoustic parameters which can distinguish DLD children from typically developing peers. As far as we know, this is the first study on Italian applying NLP techniques to the analysis of DLD speech, for screening and diagnosis purposes.

Our findings, though preliminary, suggest that voice and speech analysis could be exploit for the automatic detection of DLD in Italian preschoolers, since a set of acoustic indexes resulted statistically significant in distinguishing language-impaired children from peers: among the relevant features, some are related to fluency (e.g., Verbal Rate, Standardized Phonation Time, Standardized Pause Rate), while the fundamental frequency of the voice perceptually correlates with the tone.

However, further investigations are needed to validate this hypothesis. In particular, the acoustical properties of the voice of DLD children should be better investigated by repeating the analysis on a larger corpus. With this respect, a comparison with both other clinical targets (e.g., Late Talkers, DLD before and after treatment) and other Neurodevelopmental Disorders (e.g., Autism Spectrum Disorder, Speech Sound Disorder, and Childhood Apraxia of Speech) would be extremely useful. Therefore, the next steps of this project will involve i) the widenings of the sample, ii) the development and testing of a fully automatic pipeline, which will include, besides the features extraction, the transcription of the verbal productions employing an Automatic Speech Recognition system (ASR) specifically trained on children's voice (Beckman *et al.* 2017; Gale *et al.* 2019), and iii) the application of Machine Learning algorithms for subjects' automatic classification.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research did not receive any specific grant from funding caragencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

Due to the Italian privacy policy, supporting raw data (i.e., speech recordings, transcriptions, and clinical information) is not available. Processed data (i.e., tables of acoustic/rhythmic values, with the name of the speakers masked through an alphanumeric acronym to ensure anonymity) are available from the corresponding author (GG), upon reasonable request.

References

American Psychiatric Association, 2013, *Diagnostic and Statistical Manual of Mental Disorders, Fifth edition, DSM-5*. Washington D.C.-London, American Psychiatric Publishing.

Arosio F., Branchini C., Barbieri L. and Guasti M.T., 2014, "Failure to produce direct object clitic pronouns as a clinical marker of SLI in school-aged Italian speaking children". *Clinical Linguistics & Phonetics* 28, 639–663.

Asgari M., Bayestehtashk A. and Shafran I., 2013, "Robust and accurate features for detecting and diagnosing autism spectrum disorders", in *Proceedings of INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 191–194.

Baird A., Amiriparian S., Cummins N. Alcorn A., Batliner A., Pugachevskiy S., Freitag M., Gerczuk M. and Schuller B., 2017, "Automatic classification of autistic child vocalisations: A novel database and results", in *Proceedings of INTERSPEECH 2017*, 849–853.

Beckman M., Plummer A., Munson B. and Reidy P., 2017, "Methods for eliciting, annotating, and analyzing databases for child speech development". *Computer Speech & Language* 45, 278–299.

Beltrami D., Gagliardi G., Rossini Favretti R., Ghidoni E., Tamburini F. and Calzà L., 2018, "Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline?". *Frontiers in Aging Neuroscience* 10, 369.

Beraldi F., Gagliardi G. and Innocenti M., 2021, "Narrative abilities of Italian preschool children with developmental language disorder". *Studi e Saggi Linguistici* LIX (1), 119–150.

Bishop D., Snowling M., Thompson P., Greenhalgh T. and the CATALISE-2 consortium, 2017, "Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology". *Journal of Child Psychology and Psychiatry* 58, 1068–1080.

Bone D., Black M.P., Ramakrishna A., Grossman R.B. and Narayanan S.S., 2015, "Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism", in *Proceedings of INTERSPEECH 2015*, 1616–1620.

Bone D., Chaspari T., Audkhasi K., Gibson J., Tsiartas A., Segbroeck M.V., Li M., Lee S. and Narayanan S., 2013, "Classifying language-related developmental disorders from speech cues: the promise and the potential confounds", in *Proceedings of INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 182–186.

Bortolini U., Arfé B., Caselli M., Degasperi L., Deevy P. and Leonard L., 2006, "Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition". *International Journal of Language Communication Disorders* 41, 695–712.

Broome K., McCabe P., Docking K. and Doble M., 2017, “A systematic review of speech assessments for children with autism spectrum disorder: Recommendations for best practice”. *American Journal of Speech-Language Pathology* 26, 1011–1029.

Caccia M. and Lorusso M.L., 2021, “The processing of rhythmic structures in music and prosody by children with developmental dyslexia and developmental language disorder”. *Developmental science* 24, e12981.

Calzà L., Gagliardi G., Rossini Favretti R., Tamburini F., 2021, “Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia”. *Computer Speech & Language* 65, 101113.

Campisi P., Tewfik T., Manoukian J., Schloss M., Pelland-Blais E. and Sadeghi N., 2002, “Computer-assisted voice analysis: establishing a pediatric database”. *Archives of otolaryngology - Head & neck surgery* 128, 156–160.

Cappellari V. and Cielo A., 2008, “Vocal acoustic characteristics in pre-school aged children”. *Brazilian journal of otorhinolaryngology* 74, 265–272.

Chen Y.P.P., Johnson C., Lalbakhsh P., Caelli T., Deng G., Tay D., Erickson S., Broadbridge P., El Refaie A., Doube W. and Morris M.E., 2016, “Systematic review of virtual speech therapists for speech disorders”. *Computer, Speech & Language* 37, 98–128.

Chilosi A. and Cipriani P., 2006, *TCGB. Test di comprensione grammaticale per bambini*. Pisa, Edizioni Del Cerro.

Cho S., Liberman M., Ryant N., Cola M., Schultz R. T. and Parish-Morris J., 2019, “Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations”, in *Proceedings of INTRSPEECH 2019*, 2513–2517.

Cipriani P., Salvadorini R. and Zarmati G., 2012, *Bus Story Test. Test di valutazione delle abilità narrative*. Milano, Edizioni La Favelliana.

Cossu G., 2013, *TNL. Test neuropsicologico lessicale per l’età evolutiva*. Firenze, Hogrefe Editore.

Cresti E. and Moneglia M., 2018, “The illocutionary basis of information structure: The Language into Act Theory (L-AcT)”, in Adamou E., Haude K., Vanhove M. (eds.), *Information Structure in Lesser-described Languages. Studies in prosody and syntax*, 360–402.

Dale P., Price T., Bishop D. and Plomin R., 2003, “Outcomes of early language delay. I. predicting persistent and transient language difficulties at 3 and 4 years”. *Journal of Speech, Language, and Hearing Research* 46, 544–560.

Dellwo V., 2006, “Rhythm and speech rate: a variation coefficient for deltaC’”, in Karnowski P., Szigeti I. (eds.), *Language and Language-Processing*, 231–241.

Deng J., Cummins N., Schmitt M., Qian K., Ringeval F. and Schuller B., 2017, “Speech-based diagnosis of autism spectrum condition by generative adversarial network representations”, in *Proceedings of the 2017 International Conference on Digital Health*.

Diepeveen F.B., van Dommelen P., Oudesluys-Murphy A.M. and Verkerk P.H., 2018, “Children with specific language impairment are more likely to reach motor milestones late”. *Child: Care, Health and Development* 44, 857–862.

Dudy S., Bedrick S., Asgari M. and Kain A., 2018, "Automatic analysis of pronunciations for children with speech sound disorders". *Computer Speech & Language* 50, 62–84.

Eyben F., Weninger F., Groß F. and Schuller B., 2013, "Recent developments in open smile, the Munich open-source multimedia feature extractor", in *Proceeding of the 21th ACM Conference on Multimedia*, 835–838.

Finneran D.A., Leonard L. and Miller C., 2009, "Speech disruptions in the sentence formulation of school-age children with specific language impairment". *International Journal of Language Communication Disorders* 44, 271–286.

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

Gagliardi G., *in press*, "Natural Language Processing techniques for studying language in pathological aging: a scoping review". *International Journal of Language & Communication Disorders.*

Gagliardi G., Kokkinakis, D., Duñabeitia, J., 2021, "Editorial: Digital linguistic biomarkers: beyond paper and pencil tests". *Frontiers in Psychology* 12, 752238.

Gagliardi G. and Tamburini F., 2021, "Linguistic biomarkers for the detection of Mild Cognitive Impairment". *Lingue e linguaggio* XX, 3–31.

Gagliardi G. and Tamburini F., 2022, "The automatic extraction of Linguistic Biomarkers as a viable solution for the early diagnosis of Mental Disorders", in *Proceedings of the Thirteenth Language Resources and Evaluation Conference - LREC 2022*, 5234–5242.

Gale R., Chen L., Dolata J., van Santen J. and Asgari M., 2019, "Improving ASR Systems for Children with Autism and Language Impairment Using Domain-Focused DNN Transfer Techniques", in *Proceedings of INTERSPEECH 2019*, 11–15.

Goffman L., 1999, "Prosodic influences on speech production in children with specific language impairment and speech deficits. kinematic, acoustic, and transcription evidence". *Journal of Speech, Language, and Hearing Research* 42: 1499–1517.

Gosztolya G., Busa-Fekete R. and Tòth L., 2013, "Detecting autism, emotions and social signals using adaboost", in *INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 220–224.

Grabe E. and Low E.L., 2002, "Durational variability in speech and the rhythm class hypothesis", in Gussenhoven C. and Warner N. (eds.), *Papers in Laboratory Phonology 7*, 515–546.

Hacki T. and Heitm ̈uller S., 1999, "Development of the child's voice: premutation, mutation". *International Journal of Pediatric Otorhinolaryngology* 49, S141–S144.

Hall N., 1996, "Language and fluency in child language disorders: changes over time". *Journal of Fluency Disorders* 21, 1–32.

Hauser M., Sariyanidi E., Tunc B., Zampella C., Brodkin E., Schultz R. and Parish-Morris J., 2019, "Using natural conversations to classify autism with limited data:

Age matters”, in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 45–54.

Horowitz S., Irwin J., Briggs-Gowan M., Bosson Heenana J., Medoza J. and Carter A., 2003, “Language delay in a community cohort of young children”. *Journal of the American Academy of Child and Adolescent Psychiatry* 42, 932–940.

Kirchhoff K., Liu Y. and Bilmes J., 2013, “Classification of developmental disorders from speech signals using submodular feature selection”, in *INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 187–190.

Kiss G., van Santen J.P.H., Prud’hommeaux E. and Black L.M., 2012, “Quantitative analysis of pitch in speech of children with neurodevelopmental disorders”, in *Proceedings of INTERSPEECH 2012*, 1343–1346.

Kumar M., Chebolu P., Hyun Kim S., Martinez K., Lord C. and Narayanan S., 2018, “A knowledge driven structural segmentation approach for play-talk classification during autism assessment”, in *Proceedings of INTERSPEECH 2018*, 2763–2767.

Lee H., Hu T., Jing H., Chang Y.F., Tsao Y., Kao Y.C. and Pao T.L., 2013, “Ensemble of machine learning and acoustic segment model techniques for speech emotion and autism spectrum disorders recognition”, in *INTERSPEECH 2013,14th annual conference of the International Speech Communication Association*, 215–219.

Leonard L., 2014, “Specific language impairment across languages”. *Child development perspectives* 8, 1–5.

Levorato M. and Roch M., 2007, *TOR. Test di Comprensione del Testo Orale - 3-8 anni*. Firenze, Giunti O.S.

Li M., Tang D., Zeng J., Zhou T., Zhu H., Chen B. and Zou X., 2019, “An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder”. *Computer Speech Language* 56: 80–94.

Lin Y.S., Shur-Fen Gau S. and Lee C.C., 2018, “An interlocutor-modulated attentional LSTM for differentiating between subgroups of autism spectrum disorder”, in *Proceedings of INTERSPEECH 2018*, 2329–2333.

Lòpez-de-Ipiña K., Alonso J.B., Travieso C.M., Solé-Casals J., Egiraun H., Faundez-Zanuy M., Ezeiza A., Barroso N., Ecay-Torres M., Martinez-Lage P. and Martinez de Lizardui U., 2013, “On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis”. *Sensors* 13, 6730–6745.

MacWhinney B., 2000, *The CHILDES Project: Tools for analyzing talk*. Hillsdale, Lowrence Erlbaum Associates.

Marotta L. and Caselli M.C., 2014, *I disturbi del linguaggio. Caratteristiche, valutazione, trattamento*. Erickson.

Martìnez D., Ribas D., Lleida E., Ortega A. and Miguel A., 2013, “Suprasegmental information modelling for autism disorder spectrum and specific language impairment classification”, in *INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 195–199.

Martìnez-Nicolàs I., Llorente T., Martìnez-Sànchez F. and Garcìa Meilàn J., 2021, “Ten years of research on automatic voice and speech analysis of people with

Alzheimer's disease and Mild Cognitive Impairment: A systematic review article". *Frontiers in Psychology* 12, 620251.

McCann J. and Peppé S., 2003, "Prosody in autism spectrum disorders: a critical review". *International Journal of Language & Communication Disorders* 38, 325–350.

McKean C., Wraith D., Eadie P., Cook F., Mensah F. and Reilly S., 2017, "Subgroups in language trajectories from 4 to11 years: the nature and predictors of stable, improving and decreasing language trajectory groups". *The Journal of Child Psychology and Psychiatry* 58, 1081–1091.

Mozzanica F., Salvadorini R., Sai E., Pozzoli U., Maruzzi P., Scarponi L., Barillari M., Spada E., Ambrogi F. and Schindler A., 2016, "Reliability, validity and normative data of the Italian version of the Bus Story Test". *International Journal of Pediatric Otorhinolaryngology* 89, 17–24.

O'Hare A. and Bremner L., 2016, "Management of developmental speech and language disorders: Part 1". *Archives of disease in childhood* 101, 272–277.

Parish-Morris J., Liberman M., Ryant N., Cieri C., Bateman L., Ferguson E. and Schultz R., 2016, "Exploring autism spectrum disorders using HLT", in *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 74–84.

Paul R., Hernandez R., Taylor L. and Johnson K., 1996, "Narrative development in late talkers: Early school age. Journal of speech and hearing research". *Journal of Speech, Language, and Hearing Research* 39, 1295–1303.

Pokorny F., Schuller B., Marschik P., Brueckner R., Nystr ̈om P., Cummins N., B ̈olte S., Einspieler C. and Falck-Ytter T., 2017, "Earlier identification of children with autism spectrum disorder: An automatic vocalisation-based approach", in *Proceedings of INTERSPEECH 2017*, 309–313.

Pribuisiene R., Uloza V. and Kardisiene V., 2011, "Voice characteristics of children aged between 6 and 13 years: Impact of age, gender, and vocal training". *Logopedics Phoniatrics Vocology* 36, 150–155.

Ramarao D., Singh C., Shahnawazuddin S., Adiga N. and Pradhan G., 2018, "Detecting developmental dysphasia in children using speech data", in *2018 International Conference on Signal Processing and Communications - SPCOM*, 100–104.

Ramus F., Nespor M. and Mehler J., 1999, "Correlates of linguistic rhythm in the speech signal". *Cognition* 73, 265–292.

Räsänen O. and Pohjalainen J., 2013, "Random Subset Feature Selection in Automatic Recognition of Developmental Disorders, Affective States, and Level of Conflict from Speech", in *INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 210–214.

Reddy M., Alku P. and Rao K., 2020, "Detection of specific language impairment in children using glottal source features". *IEEE Access* 8, 15273–15279.

Redmond S., 2011, "Peer victimization among students with specific language impairment, attention deficit hyperactivity disorder, and typical development". *Language, Speech, and Hearing Services in Schools* 42, 520–535.

Redmond S.M. and Timler G., 2007a, "Addressing the social concomitants of developmental language impairments", in Apel, K., Kamhi, A. and Masterson, J. (eds.), *Clinical decision making in developmental language disorders,* , 185–202.

Redmond S.M. and Timler G., 2007b, "Don't talk to him; he's weird. A social consequences account of language and social interactions", in Rice M. (ed.), *Communication and language intervention series: Vol 2. Enhancing children's communication. Research foundations for intervention*, 139–158.

Reilly S., McKean C., Morgan A. and Wake M., 2015, "Identifying and managing common childhood language and speech impairments". *BMJ* 350, h2851.

Renfrew C., 2015, *Bus Story Test. A test of narrative speech*. Speechmark.

Ringeval F., Demouy J., Szaszak G., Chetouani M., Robel L., Xavier J., Cohen D. and Plaza M., 2011, "Automatic intonation recognition for the prosodic assessment of language impaired children". *IEEE Transactions on Audio, Speech Language Processing* 19, 1328–1342.

Roark B., Mitchell M., Hosom J.P., Hollingshead K. and Kaye J.A., 2011, "Spoken Language Derived Measures for Detecting Mild Cognitive Impairment". *IEEE Transactions on Audio Speech, and Language Processing* 19, 2081–2090.

Rustioni D. and Lancaster M., 2007, *PVCL. Prove di Valutazione della Comprensione Linguistica*. Firenze, Giunti O.S.

van Santen J., Prud'hommeaux E. and Black L., 2009, "Automated assessment of prosody production". *Speech Communication* 51, 1082–1097.

Sansavini A., Favilla M. E., Guasti M. T., Marini A., Millepiedi S., Di Martino M. V., Vecchi S., Battajon N., Bertolo L., Capirci O., Carretti B., Colatei M. P., Frioni C., Marotta L., Massa S., Michelazzo L., Pecini C., Piazzalunga S., Pieretti M., Rinaldi P., Salvadorini R., Termine C., Zuccarini M., D'amico S., De Cagno A. G., Levorato M. C., Rossetto T. and Lorusso M. L., 2021, "Developmental language disorder: Early predictors, age for the diagnosis, and diagnostic tools. A scoping review". *Brain Sciences* 11/5, 24–56.

Satt A., Sorin A., Toledo-Ronen O., Barkan O., Kompatsiaris I., Kokonozi A. and Tsolaki M., 2013, "Evaluation of Speech-Based Protocol for Detection of Early-Stage Dementia", in *INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*, 1692–1696.

Saturday N. and Kabari L., 2020, "Voice parameter analysis for detection of autism spectrum disorder". *Journal of Environmental Science, Computer Science and Engineering & Technology* 9, 92–99.

Schuller B., Steidl S., Batliner A., Vinciarelli A., Scherer K., Ringeval F., Chetouani M., Weninger F., Eyben F., Marchi E., Mortillaro M., Salamin H., Polychroniou A., Valente F. and Kim S., 2013, "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism", in *INTERSPEECH 2013, 14thannual conference of the International Speech Communication Association*, 148–152.

Schuller B., Weninger F., Zhang Y., Ringeval F., Batliner A., Steidl S., Eyben F., Marchi E., Vinciarelli A., Scherer K., Chetouani M. and Mortillaro M., 2019, "Affective and behavioural computing: Lessons learnt from the first computational paralinguistics challenge". *Computer Speech Language* 53, 156–180.

Sharma Y. and Singh B., 2020, "Classification of children with specific language impairment using pitch-based parameters", in *2020 IEEE Recent Advances in Intelligent Computational Systems* - *RAICS*, 42–46.

Singh S., Bucks R.S. and Cuerden J.M., 2001, "An evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech". *Aphasiology* 15, 571–583.

Solorio T., 2013, "Survey on emerging research on the use of natural language processing in clinical language assessment of children". *Language and Linguistics Compass* 7, 633–646.

Tanaka H., Sakti S., Neubig G., Toda T. and Nakamura S., 2014, "Linguistic and acoustic features for automatic identification of autism spectrum disorders in children's narrative" in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 88–96.

Tomblin J., Records N., Buckwalter P., Zhang X., Smith E. and O'Brien M., 1997, "Prevalence of specific language impairment in kindergarten children". *Journal of Speech, Language, and Hearing Research* 40, 1245–1260.

Van Der Meulen S., Janssen P. and Den Os E., 1997, "Prosodic abilities in children with specific language impairment". *Journal of Communication Disorders* 30: 155–170.

Van Santen J., Prud'hommeaux E., Black L. and Mitchell, M., 2010, "Computational prosodic markers for autism". *Autism* 14, 215–236.

Velupillai S., Suominen H., Liakata M., Roberts A., Shah A.D., Morley K., Osborn D., Hayes J., Stewart R., Downs J., Chapman W. and Dutta R., 2018, "Using clinical natural language processing for health outcomes research: Overview and actionable suggestions for future advances". *Journal of Biomedical Informatics* 88, 11–19.

Wells B. and Peppé S., 2003, "Intonation abilities of children with speech and language impairments". *Journal of Speech, Language, and Hearing Research* 46, 5–20.

Wittenburg P., Brugman H., Russel A., Klassmann A. and Sloetjes H., 2006, "ELAN: a professional framework for multimodality research", in Calzolari N., Choukri K., Gangemi A., Maegaard B., Mariani J., Odijk J. and Tapias D. (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, 1556–1559.

Yu H. and Mak M., 2011, "Comparison of voice activity detectors for interview speech in NIST speaker recognition evaluation", in *Proceedings of INTERSPEECH 2011*, 2353–2356.