Automatic Detection of Rhythmic Features in Pathological Speech of MCI and Dementia Patients

Marica Belmonte[1], Gloria Gagliardi[1], Dimitrios Kokkinnakis[2], Fabio Tamburini[1]
[1] University of Bologna, [2] University of Gothenburg
marica.belmonte@studio.unibo.it, gloria.gagliardi@unibo.it,
dimitrios.kokkinakis@svenska.gu.se, fabio.taburini@unibo.it

Abstract

Linguistic alterations represent one of the prodromal signs of cognitive decline associated with Dementia. In recent years, a growing body of work has been devoted to the development of algorithms for the automatic linguistic analysis of both oral and written texts, for diagnostic purposes. The extraction of Digital Linguistic Biomarkers from patients' verbal productions can indeed provide a rapid, ecological, and cost-effective system for large-scale screening of the pathology. This article contributes to the ongoing research in the field by exploring a traditionally less studied aspect of language in Dementia, namely the rhythmic characteristics of speech. In particular, the paper focuses on the automatic detection of rhythmic features in Italian-connected speech. A landmark-based system was developed and evaluated to segment the speech flow into vocalic and consonantal intervals and to calculate several rhythmic metrics. Additionally, the reliability of these metrics in identifying Mild Cognitive Impairment and Dementia patients was tested.

Keywords: Dementia, MCI, Digital Linguistic Biomarkers, rhythm

1. Introduction

Dementia is a syndrome that causes the disturbance of multiple higher cortical functions, leading to the loss of functional autonomy (Altieri et al., 2021). It represents a major public health concern due to the high number of people affected in the world. Moreover, it is estimated that the number of cases will increase up to 139 million by 2050 (Long et al., 2023). This syndrome can be caused by many pathologies (e.g., cerebral atrophies due to protein misfolding diseases, brain damage linked to vascular issues, and metabolic disorders) making the clinical manifestations varied. Moreover, the symptoms can be easily misinterpreted as effects of physiological ageing. This is particularly true in the very early stages of the disease, a prodromic state of cognitive decline called in the scientific literature “Mild Cognitive Impairment” (MCI, Petersen et al., 1999). This timeframe holds special interest for researchers focused on early intervention tools.

A large body of evidence demonstrates that language is one of the cognitive domains affected by Dementia (Boschi et al. 2017; Gagliardi, 2024). More importantly, since the linguistic alterations manifest much earlier than other clinical symptoms (Eyigoz et al., 2020), a substantial amount of research explored the use of linguistic analysis as a screening tool (König et al., 2015; Gagliardi and Tamburini, 2021; 2022; Themistocleous et al., 2018; 2020). Therefore, language appears to be a promising and valuable source of biomarkers. Furthermore, with the emergence of sophisticated technologies for Natural Language Processing (henceforth: NLP), much work has been done in the past decade to develop automatic tools for linguistic analysis (Martínez-Nicolás et al., 2021; Calzà et al., 2021). The advantages of using NLP instruments as a screening tool are noteworthy: they are non-invasive, fast, easy to employ, and significantly less expensive than other diagnostic techniques (Gagliardi et al., 2021; Duñabeitia et al., 2024).

This work specifically focuses on the automatic detection of rhythmic features in Italian-connected speech, a level of analysis that has received less attention in the literature. A computational tool was developed and evaluated for their automatic extraction. Furthermore, their relationship with the pathological conditions of MCI and early Dementia (eD) was investigated.

The paper is structured as follows. Section 2 is devoted to the discussion of the role played by rhythmic parameters in the study of pathological speech, as well as the task of their automatic detection. In section 3, a solution based on 'acoustic landmarks' is presented. Section 4 describes and discusses the evaluation of the system's performance. Section 5 illustrates the application of the algorithm on connected speech from Italian patients diagnosed with MCI or Dementia. Additionally, the relationship between the features and the pathologies is investigated through statistical analysis. In section 6, the main limitations of the study are outlined, along with some conclusions.

2. The Analysis of Rhythm and its Application to Pathological Speech

2.1 Automatic Detection of Rhythmic Features Using Landmarks-based Acoustic Analysis

Although rhythmic linguistic analysis is a powerful tool for discriminating various pathological conditions (Keshavarzi et al., 2024; Lowit et al., 2018), it comes with some downsides. It often requires manual (time-aligned) transcription and

Proceedings of the RaPID-5 @LREC-COLING 2024, pages 34–44

annotation of the recorded speech. This procedure is not only extremely time-consuming but also demands a trained specialist for accurate execution. Furthermore, the results can be challenging to replicate due to the subjective element of human judgment. As a result, conducting large-scale studies is hardly feasible. Taken together, these obstacles make the actual use of linguistic analysis in the clinical setting very unlikely. In this respect, the development of algorithms for the automation of this task would be highly beneficial.

One promising tool for this purpose is Speechmark® (Boyce et al., 2012), a software for landmark-based acoustic analysis. The notion of 'landmarks' was first introduced by the Speech Communication Group at MIT (Stevens et al. 1992), and it can be defined as timestamps, denoting sharp changes in speech articulation, corresponding to specific transitions between different classes of sounds in the signal (Stevens, 2002). Thus, landmarks represent the acoustic correlate of distinctive articulatory features.

Utilising landmarks in acoustic analysis appears particularly suitable for automatically computing rhythmic features: from the patterns of acoustic landmarks, vocalic and consonantal intervals can be derived, facilitating the calculation of many rhythmic metrics.

2.2 Rhythmic Features in the Study of Pathological Speech

Various kinds of linguistic rhythm metrics have been employed in the study of pathological speech, yielding robust results. For instance, rhythmic alterations have been found to be strongly linked to Dysarthria resulting from Parkinson's disease (Pettorino et al., 2016; Lowit et al., 2018). Nevertheless, Ivanova et al. (2024) highlighted that rhythmic alterations in cognitive decline due to Dementia are less clear, given the largely inconsistent results available in the literature. Cera et al. (2018), among others, analysed several rhythmic features, such as vowel duration and the ratio between pauses and phonation time, in Dementia of the Alzheimer type. Their patients exhibited significantly longer vowel percentages and longer pauses compared to healthy controls matched by age. In Meilán et al. (2020), various acoustic and rhythmic parameters were detected, comparing subjects with non-amnesic MCI and subjects with prodromal Dementia. Regarding the rhythmic features, they effectively discriminate between the two groups. Contrary to expectations, in Beltrami et al. (2018) and Calzà et al. (2021), the computed rhythmic parameters do not significantly differ between healthy control subjects and patients, nor between MCI subjects and eD subjects.

Therefore, it is even more complex to identify the physiological correlates of linguistic rhythm and their alterations due to pathological conditions. Likely the interplay of numerous physiological factors overall accounts for linguistic rhythm (Poeppel and Assaneo, 2020). As stated by Lowit (2014), anything that disturbs the natural flow of speech could essentially cause deviations in rhythmic structure. It is known that, since many rhythmic metrics are influenced by speech rate, rhythm is intertwingled with speech rate. In terms of physiology, it is reported that the overall speech rate declines with healthy aging (Pellegrino et al., 2018; Linville, 1996). Specifically, the temporal properties of speech, such as articulation rate, articulation rate stability, and movement time (i.e., the time from movement initiation to completion), are disrupted in normal aging, most likely reflecting central difficulties at the level of speech motor planning or execution (Tremblay et al., 2019) and muscular atrophy at the level of articulatory organs (Scholtz, 2007). Those difficulties in healthy older people may be exacerbated in people affected by a disease. In neuropathological conditions, specific and additional damages are present in the cortical areas affected by the disorder. For instance, Parkinson's disease is characterised by a disruption in the cortical sensorimotor system (Chen et al., 2022) leading to neuromuscular control impairment that is reflected in the rhythmic alterations consistently associated with this disease (Lowit et al., 2018). With regard to Dementia, the cortical areas involved may vary considerably and the effects on linguistic rhythm depend on the localisation and the extension of the neural disruption which is described as atrophy. While in Alzheimer's disease the temporoparietal regions are the most affected by the atrophy, in Frontotemporal Dementia it is the frontotemporal area to be mainly involved (Nicastro et al., 2020). According to Meilàn et al. (2020), the disordered rhythm in eD subjects is the result of alterations comparable to the ones found in neurogenic speech disorder patients: such as changes in speech timing and poor coordination in articulatory systems. Similarly, Cera et al. (2018) argue that these disorders are related to phonetic-motor planning, which leads to poor pronunciation and an alteration in phonological planning and rhythm. Overall, the evidence from the neurophysiology of Dementia seems to lead to the hypothesis of a speech impairment characterised by rhythmic problems. Nevertheless, more research is needed to identify the exact physiological mechanisms underlying the linguistic rhythm phenomena both in healthy and pathological subjects.

3. A Landmark-based Algorithm

In the present work, a landmark-based system was developed to automatically segment speech into vocalic and consonantal intervals and to calculate several rhythmic metrics. The algorithm comprises the software Speechmark (Boyce et al., 2012) and a custom-designed Python script.

A two-step procedure is foreseen:

1. Landmarks are identified by Speechmark (SM), which provides a time-aligned annotation (i.e., each landmark is associated with a timestamp) (§ 3.1).
2. The script extracts consonantal and vocalic intervals from the SM's annotation, from which, in turn, rhythmic features are computed (§ 3.2).

3.1 Speechmark

Speechmark (Boyce et al., 2012) is a MATLAB® toolbox that automatically detects landmarks directly from the audio files. It was developed based on the work of Stevens (2002), Howitt (2000), and Liu (1996). The software (Ishikawa et al., 2017) has been largely employed in the clinical linguistics field to study numerous different pathologies: Dysarthria (Liu and Chen, 2021), Dysphonia (Ishikawa et al., 2023), Autism Spectrum Disorder (Lau et al., 2023), and Speech Sound Disorder (Valentine et al., 2023), to mention a few.

In the present study, the vowel_segs_full function from the 1.3 version of the SM MATLAB toolbox was employed. The SM algorithm distinguishes among several types of landmarks based on whether they signal laryngeal or vocal tract events, as well as abrupt or peak events (MacAuslan, 2016). The peak events are detected when there's a peak in the energy of the signal. For instance, a vowel peak landmark (V-lm) is found when there is «a local peak of harmonic power. Articulatorily, vowel landmarks often correspond to the maximum opening of the mouth within a syllabic unit» (MacAuslan and Boyce, 2016). The abrupt ones are named as such because they are identified by a rapid rise or fall of energy across several frequency bands. For this reason, the abrupt landmarks come in pairs of positive and negative: positive (+) for energy rising and negative (-) for energy declining. For instance, one of the main abrupt landmarks detected by SM is the (+/-) g-landmark (g-lm). It is particularly significant since it signals the start and the end of vocal folds' activation. For a more comprehensive description of the landmarks, please refer to Appendix A.

The pairs of abrupt landmarks serve as the starting point for our script to detect vowel and consonant segments.

3.2 From Speechmark's Annotation to the Rhythmic Features

The script takes the landmark annotation as input and produces a list of vocalic and consonantal intervals as output. Rhythmic features are estimated from these intervals.

First, the system locates the g-lms and defines the intervals between pairs of + g-lm and - g-lm. To identify vowels, it searches for intervals opened by a + g-lm, which indicates the activation of the vocal folds. Then, it checks if a V-lm exists within the same time interval. If one is found, the segment is labelled as vocalic. If there is no matching V-lm, the system looks for landmarks that correlate with voiced consonants (cf. Appendix A). If those are found, the segment is labelled as consonantal. If they are not found, the segment is labelled as vocalic. Thus, the primary criterion used to identify vocalic intervals is finding an interval opened by a + g-lm and a correspondent V-lm within the same time span. Conversely, if the interval starts with a - g-lm, it indicates that the speech segment is unvoiced. It is therefore labelled either as silence or as consonantal. Silence is identified if no other landmark is present between the - g-lm and the successive + g-lm, and the interval is at least 200 ms long. In all other cases, the interval is labelled as consonantal.

These intervals are utilised to compute the rhythmic features described in § 5.2.

4. Algorithm Evaluation

4.1 Materials and Methods

The system was then evaluated for performance testing. The material selected for the evaluation was composed of 100 audio recordings extracted from the CLIPS corpus (Albano Leoni, 2007; 2004), balanced by speaker gender and elicitation task. This linguistic resource provides different levels of manual annotation, including time-aligned phonetic transcription, which was exploited as a starting benchmark for performance assessment, to carry out the automatic evaluation. Moreover, moving forward in the next stages of the system's development, this baseline will be essential for tracking the evolution of performance.

The evaluation was conducted by measuring the alignment between the system's annotation and the target annotation. The fair evaluation approach (*FairEval*), as described in Ortmann (2022), was adopted to make the metrics both insightful and suitable for comparison with other systems. According to the scholar, traditional metrics, (i.e., precision, recall, and F1-score) can result in double penalties when applied naively to segmentation alignment measures. Consequently, the following types of errors were examined:

- *Deletion*: the target span is missed. It counts as a false negative.
- *Insertion*: the span is present in the output but doesn't correspond (not even partially) to any of the ones in the target annotation. It counts as a false positive.
- *Labelling error (L_E)*: the output span matches with the target span but the label is incorrect.
- *Boundary error (B_E)*: the output span partially overlaps with the target span and the label is correct.

- *Labelling and boundary error (L_BE)*: the output span partially overlaps with the target span and the label is incorrect.

A threshold of 20 ms was adopted.

4.2 Results

The Figure 1 displays algorithm errors across the five different types.

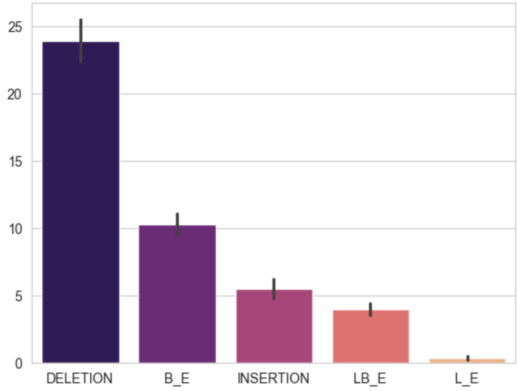

Figure 1: Errors made by the landmark-based system.

Precision, recall, and F1-score values were obtained by converting errors into false positives and false negatives (with true positives being annotations that had both matching boundaries and labels). According to the equation proposed by Ortmann (2022), different weights were assigned to different errors. The results of the evaluation are listed below:

PRECISION = 0.576
RECALL =0.325
F1-score =0.415

We can observe a trade-off between precision and recall. The system lacks in sensitivity (i.e., recall) what it gains in confidence (precision). In our data, this is due to the considerably higher number of false negatives compared to false positives. In other terms, these results can be explained by the disproportion between the number of deletions and the number of insertions (cf. Figure 1), with deletions accounting for 54% of the total errors. The proposed fine-grained error taxonomy allows us to separately analyse the performance of the system on both the segmentation and labelling tasks. Although the two stages of the model are not completely independent, since finding a span is preliminary to tagging it. Generally speaking, the overall unsatisfactory performance of the system is mainly due to the limited ability of the model to accurately predict the span's boundaries. In addition to deletions, a considerable number of boundary errors are reported, i.e., cases where the system correctly predicts the label but only partially predicts the boundaries of the span. Thus, most of the errors can be ascribed to segmentation.

4.3 Discussion

It is possible to make some hypotheses about the causes of the algorithm's low performance. One potential source of errors can be identified in the clusters of vowels and sonorant consonants, especially approximants, which are classified as consonants. As mentioned earlier, landmarks are detected based on an abrupt rise or fall of energy in the spectrum. In the case of a sequence of sounds that share many acoustic characteristics, such as heavy voicing, it is expected that there will be no abrupt transitions and therefore no landmarks. This issue is exacerbated by the effect of coarticulation.

Moreover, often the landmark only appears to mark one side of the transition: for example, there may be a (+) sign landmark but not the respective (-) sign landmark closing the interval, because the fall in energy was not abrupt enough for the Speechmark system to detect it. This partly explains the missing spans (i.e., deletions).

On the other hand, this highlights a more general issue related to the interface between phonology and acoustic phonetics. While landmarks are inherently acoustic in nature, a phonological criterion is adopted to distinguish between vowels and consonants. Thus, even the most outperforming landmark annotation system would present discrepancies with the theoretical classification required by a phonological category, such as vowels and consonants. More importantly, the actual realization of speech is susceptible to great variability (i.e., the *lack of invariance* problem, Klatt, 1986; Liberman et al., 1967). As an example, it is not rare for an occlusive to be uttered as if it were an approximant. Therefore, the patterns of landmarks are considerably more varied than Stevens' model allows us to predict.

For future improvements, instead of defining the algorithm solely based on the rules from Stevens' paradigm, an algorithm for automatic phoneme-landmark mapping in Italian could be implemented, as described in DiCicco and Patel (2008).

Furthermore, one substantial source of errors can be found in some unexpected SM behaviours. It was observed that the system often failed to detect voicing in the speech. Since landmarks come in pairs, the system's ability to correctly predict subsequent ones is compromised if even just one is missing.

Therefore, one prospect for future development could be integrating some formant tracking features into the system. This improvement could be achieved either by using the formant tracking function provided by SM itself or by implementing it with a custom-designed script. This would allow for a more precise identification of vowel spans and for a better distinction between vowels and consonants in heavily voiced clusters in the utterances.

5. Automatic Detection of the Rhythmic Features from the Speech of MCI and Dementia Patients

5.1 Materials and Methods

The landmark-based system was ultimately employed to detect rhythmic features in Italian-connected speech. We used a subset of the speech corpus described in Gagliardi et al. (2016), thus replicating the results of Beltrami et al. (2018) by means of a novel landmark-based automatic detection system and extracting additional rhythmic features.

The final dataset consisted of 198 audio recordings from 66 subjects, comprising 33 healthy control subjects and 33 pathological subjects. The groups were balanced for age, gender, and years of education. The pathological group comprised 11 subjects with amnesic Mild Cognitive Impairment (aMCI), 11 subjects with multidomain Mild Cognitive Impairment (mdMCI), and 11 subjects with early Dementia (eD). All the subjects underwent a neuropsychological screening (Velayudhan et al., 2014) composed by MMSE – Mini-Mental State Examination (Folstein et al., 1975; Measso et al., 1993), MoCA – Montreal Cognitive Assessment (Nasreddine et al., 2005; Conti et al., 2015), GPCog – General Practitioner Assessment of Cognition (Brodaty et al., 2002; Pirani et al., 2017), CDT – Clock Drawing Test (Critchley et al., 1953; Lee et al., 2011), and verbal fluency tests (phonemic and semantic, Carlesimo et al., 1996; Novelli et al., 1986).

Their semi-spontaneous monological speech was recorded in a clinical setting using off-the-shelf equipment. Each subject completed three elicitation tasks, resulting in three audio recordings per subject: describing a picture, describing a typical workday, and recounting the last dream they could remember.

Following the requirements of SM, the audio files were subsampled to 16kHz. Thus, using SM, landmark annotations were obtained for each audio file. As described in Section 3, these landmark annotations were then converted into time-aligned segmentations of vocalic and consonantal intervals, and the rhythmic metrics were computed.

5.2 The Features

The following parameters have been computed based on landmark-derived intervals:

- V%: Percentage of vocalic intervals within the utterance. It represents the sum of the duration of vocalic intervals over the total duration of the utterance (Ramus et al., 2000).
- Std_V and std_C: Standard deviation of both vocalic and consonantal interval durations (Ramus et al., 2000).
- Varco_V and Varco_C: Variation coefficient of the standard deviation of vocalic and consonantal intervals (Dellwo, 2006).
- nPVI and rPVI: Pairwise Variability Index (PVI), both raw and normalized. The index quantifies the level of variability in successive measurements of vowel intervals (Grabe and Low, 2002).
- VtoV_mean and VtoV_std: Vowel onset point interval durations, including both mean and standard deviation (Pettorino et al., 2013).
- Varco_VC: Coefficient of variation of interval duration between a vowel and the successive consonant. It approximates the duration of a syllable (Liss et al., 2009).

5.3 Statistical Analysis

All the statistical analysis was carried out in Python. Table 1 summarizes the descriptive statistics of rhythmic metrics computed on our cohort.

	CON	MCIa	MCImd	eD
V_%	17.35 (15.38)	14.45 (11.20)	20.94 (14.93)	15.94 (12.58)
Std_V	0.09 (0.05)	0.08 (0.04)	0.10 (0.04)	0.10 (0.05)
Std_C	0.34 (0.71)	0.26 (0.39)	0.30 (0.44)	0.21 (0.34)
Varco_V	0.93 (0.22)	0.86 (0.12)	0.94 (0.16)	0.93 (0.24)
Varco_C	1.30 (0.86)	1.17 (0.63)	1.40 (1.11)	1.08 (0.75)
rPVI	0.08 (0.04)	0.08 (0.04)	0.10 (0.04)	0.10 (0.05)
nPVI	0.73 (0.15)	0.74 (0.11)	0.81 (0.12)	0.76 (0.13)
VtoV_mean	1.11 (0.99)	1.10 (0.79)	0.90 (0.91)	1.22 (0.89)
VtoV_std	1.26 (0.90)	1.47 (1.02)	1.19 (1.39)	1.46 (0.95)
Varco_VC	1.20 (0.29)	1.35 (0.36)	1.19 (0.32)	1.26 (0.30)

Table 1. Rhythmic features across the cohorts. Values are expressed as means and (standard deviations).

A non-parametric Kruskal-Wallis test was conducted on the data ($\alpha = 0.05$). As shown in Table 2, the inferential analysis did not reveal any significant difference in the metrics across the different cohorts (i.e., CON, MCIa, MCImd, eD).

	statistics	p-value	statistical significance
V_%	3.96	0.26	/
Std_V	3.51	0.31	/
Std_C	0.81	0.84	/

Varco_V	5.94	0.11	/
Varco_C	2.11	0.54	/
rPVI	4.08	0.25	/
nPVI	6.16	0.10	/
VtoV_mean	6.02	0.11	/
VtoV_std	7.05	0.07	/
Varco_VC	6.55	0.08	/

Table 2. Results of the inferential test of Kruskal-Wallis.

5.4 Discussion

In the previous sections, the experimental procedure adopted to investigate the relation between the rhythmic features and the pathological conditions of MCI and Dementia was described. The statistical analysis of the rhythmic parameters did not reveal any difference between the patients' group and the healthy control group. In fact, none of the parameters were found to be significantly divergent among the four sampled cohorts (healthy control, aMCI, mdMCI, and eD), (p-value > 0.05 at the Kruskal-Wallis test). Thus, it appears that linguistic rhythmic metrics are not able to discriminate between healthy controls and pathological subjects, nor between MCI and Dementia patients.
Considering the inconsistency of the results obtained through this class of linguistic biomarkers (Ivanova et al., 2024) across different languages, further work is needed to determine the reason behind the negative results, whether it is the poor accuracy of the algorithm or the irrelevance of the rhythmic metrics.

6. Concluding Remarks

This work aimed to investigate the relationship between the pathological conditions of MCI, and early Dementia, and the rhythmic features extracted from semi-spontaneous speech. It also proposed the prototype of a landmark-based system for the automatic detection of these features from Italian-connected speech. The results from the system evaluation and metrics extraction were presented and discussed.
To summarise, an unsatisfactory performance level of the algorithm was reported. The low evaluation metrics are mainly due to the system's limited ability to accurately predict the span's boundaries. Accordingly, several options for future improvements were discussed, including an algorithm implementation for automatic phoneme-landmark mapping and the integration of some formant tracking features.
Moreover, in line with the results of Beltrami et al. (2018) and Calzà et al. (2021) on Italian, the analysis of rhythmic parameters did not reveal any difference between patients and healthy controls.
Although the former is a clearly negative result, it remains to be clarified whether the lack of significance of the rhythmic features is due to the insensitivity of these indices or the poor reliability of the algorithm, given the variety of findings in languages other than Italian.
It is also worth noticing that this study has several limitations that need to be addressed. Firstly, the syllable-based metrics are currently not included among the ones analysed. It would be interesting in future work to analyse those features as well, given the results reported by Meilán et al. (2020) on Spanish. Furthermore, the effect of the elicitation task employed should be considered. Several studies (Maffia et al., 2021) suggest that reading tasks are more sensitive in capturing rhythm alterations. Thus, they could be the subject of future investigations.
Finally, the main limitation of the present work is the small dataset used for testing. A bigger sample size would enhance the accuracy of the results.

7. Acknowledgments

The authors would like to thank Joel MacAuslan, Suzanne Boyce, and Liu Chin-Ting for their invaluable support with Speechmark.

8. Funding

This study was partially funded by the European Union – NextGenerationEU programme through the Italian National Recovery and Resilience Plan – NRRP (Mission 4 – Education and research), as a part of the project *ReMind: an ecological, cost-effective AI platform for early detection of prodromal stages of cognitive impairment* (PRIN 2022, 2022YKJ8FP – CUP J53D23008380006).
In addition, the work was made possible by the funding received by MB for her mobility at the University of Gothenburg under the Erasmus+ (*Mobility for Traineeships*) programme 2023/24 (University of Bologna, Managerial Decree 1553/2023 Prot. No. 0064732).

9. CRediT Author Statement

MB: Data Curation, Software, Formal analysis, Writing - Original Draft. **GG**: Conceptualization, Funding acquisition, Supervision, Writing - Review & Editing. **DK**: Supervision. **FT**: Methodology, Resources, Supervision, Software.

10. Bibliographical References

Albano Leoni, F., Sobrero, A.A., Paoloni, A. (2007). Corpora e lessici di italiano parlato e scritto (CLIPS). *Bollettino di italianistica, Rivista di critica, storia letteraria, filologia e linguistica*, 2007(2): 121–148. doi: 10.7367/71826

Altieri, M., Garramone, F. and Santangelo, G. (2021). Functional autonomy in dementia of the Alzheimer's type, mild cognitive impairment, and healthy aging: a meta-analysis. *Neurological sciences*, 42(5):1773–1783. doi: 10.1007/s10072-021-05142-0

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. doi: 10.3389/fnagi.2018.00369

Boyce, S., Fell, H., and MacAuslan, J. (2012). Speechmark: Landmark detection tool for speech analysis. In *Proceedings of Interspeech 2012*, pp. 1894–1897. Portland (OR), USA, 9-13 September 2012. doi: 10.21437/Interspeech.2012-513

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A. and Cappa, S.F., (2017). Connected speech in neurodegenerative language disorders: a review. Frontiers in Psychology, 8:1–21.

Brodaty, H., Pond, D., Kemp, N.M., Luscombe, G., Harding, L., Berman, K. and Huppert, F.A. (2002). The GPCOG: a new screening test for dementia designed for general practice. *Journal of the American Geriatrics Society*, 50(3):530–534. doi: 10.1046/j.1532-5415.2002.50122.x

Calzà, L., Gagliardi, G., Rossini Favretti, R. and Tamburini, F. (2021). Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer, Speech & Language*, 65: 101113. doi: 10.1016/j.csl.2020.101113

Carlesimo, G.A., Caltagirone, C., Gainotti, G. and the Group for the Standardization of the Mental Deterioration Battery (1996). The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. *European Neurology*, 36: 379–384.

Cera, M.L., Ortiz, K.Z., Bertolucci, P.H.F., and Minett, T. (2018). Phonetic and phonological aspects of speech in Alzheimer's disease. *Aphasiology*, 32(1), 88–102. doi: 10.1080/02687038.2017.1362687

Chen R, Berardelli A, Bhattacharya A, Bologna M, Chen KS, Fasano A, Helmich RC, Hutchison WD, Kamble N, Kühn AA, Macerollo A, Neumann WJ, Pal PK, Paparella G, Suppa A, Udupa K. Clinical neurophysiology of Parkinson's disease and parkinsonism. Clin Neurophysiol Pract. 2022 Jun 30;7:201-227. doi: 10.1016/j.cnp.2022.06.002. PMID: 35899019; PMCID: PMC9309229.

Conti, S., Bonazzi, S., Laiacona, M., Masina, M. and Vanelli Coralli, M. (2015). Montreal Cognitive Assessment (MoCA) – Italian version: regression-based norms and equivalent scores. *Neurological sciences*, 36(2): 209–214. doi: 10.1007/s10072-014-1921-3

Critchley, M. (1953). *The parietal lobes*. New York: Hafner Publishing Company.

Dellwo, V. (2006). *Rhythm and Speech Rate: A Variation Coefficient for DeltaC*. In P. Karnowski & I. Szigeti (Eds.) *Language and language-processing*. Frankfurt/Main: Peter Lang, pp. 231–241. doi: 10.5167/UZH-111789

DiCicco, T.M. and Patel, R. (2008). Automatic landmark analysis of dysarthric speech. *Journal of Medical Speech-Language Pathology*, 16(4): 213–219.

Duñabeitia, J.A., Kokkinakis, D. and Gagliardi, G. (2024). Editorial: Digital linguistic biomarkers: beyond paper and pencil tests, volume II. *Frontiers in Psychology*, 14:1358852. doi: 10.3389/fpsyg.2023.1358852

Eyigoz, E., Mathur, S., Santamaria, M., Cecchi, G., and Naylor, M. (2020). Linguistic markers predict onset of Alzheimer's disease. *EClinicalMedicine*, 28: 100583. doi: 10.1016/j.eclinm.2020.100583

Folstein, M.F., Folstein, S.E. and McHugh P.R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3):189–198. doi: 10.1016/0022-3956(75)90026-6

Gagliardi, G. (2024). Natural language processing techniques for studying language in pathological ageing: A scoping review. *International Journal of Language & Communication Disorders*, 59(1): 110–122. doi: 10.1111/1460-6984.12870

Gagliardi, G., Kokkinakis, D. and Duñabeitia, J.A. (2021). Editorial: Digital Linguistic Biomarkers: Beyond Paper and Pencil Tests. *Frontiers in Psychology*, 12:752238. doi: 10.3389/fpsyg.2021.752238

Gagliardi, G. and Tamburini, F. (2021). Linguistic biomarkers for the detection of Mild Cognitive Impairment. *Lingue e Linguaggio*, XX(1): 3–31. doi: 10.1418/101111

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 13th Conference on Language Resources and Evaluation (LREC 2022)*, pp. 5234–5242, Marseille, France, 21-23 June 2022. European Language Resource Association (ELRA).https://aclanthology.org/2022.lrec-1.561

Grabe, E. and Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7*. Berlin/New York: De Gruyter Mouton, pp. 515–546. doi: 10.1515/9783110197105.2.515

Howitt, A.W. (2000). *Automatic syllable detection for vowel landmarks*. Thesis (Sc.D.) Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science.

Ishikawa, K., MacAuslan, J., and Boyce, S. (2017). Toward clinical application of landmark-based speech analysis: Landmark expression in normal adult speech. *The Journal of the*

Acoustical Society of America, 142(5): EL441–EL447.
doi: 10.1121/1.5009687

Ishikawa, K., Pietrowicz, M., Charney, S., and Orbelo, D. (2023). Landmark-based analysis of speech differentiates conversational from clear speech in speakers with muscle tension dysphonia. *JASA Express Letters*, 3(5), 055203.
doi: 10.1121/10.0019354

Ivanova, O., Martínez-Nicolás, I. and Meilán, J.J.G. (2024). Speech changes in old age: Methodological considerations for speech-based discrimination of healthy ageing and Alzheimer's disease. *International Journal of Language & Communication Disorders*, 59(1), 13–37.
doi: 10.1111/1460-6984.12888

Keshavarzi, M., Di Liberto, G.M., Gabrielczyk, F., Wilson, A., Macfarlane, A. and Goswami, U. (2024). Atypical speech production of multisyllabic words and phrases by children with developmental dyslexia. *Developmental science*, 27(1): e13428.
doi: 10.1111/desc.13428

Klatt, D.H. (1986) The problem of variability in speech recognition and in models of speech perception. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and Variability in Speech Processes*. New York: Psychology Press, pp. 300–319.
doi: 10.4324/9781315802350

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P.H., and David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(1), 112–124.
doi: 10.1016/j.dadm.2014.11.012

Lau, J.C.Y., Losh, M. and Speights, M. (2023). Differences in speech articulatory timing and associations with pragmatic language ability in autism. *Research in Autism Spectrum Disorders*, *102*, 102118.
doi: 10.1016/j.rasd.2023.102118

Lee, J.H., Oh, E.S., Jeong, S.H., Sohn, E.H., Lee, T.Y. and Lee, A.Y. (2011). Longitudinal changes in clock drawing test (CDT) performance according to dementia subtypes and severity. Archives of gerontology and geriatrics, 53(2): e179–e182.
doi: 10.1016/j.archger.2010.08.010

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461. doi: 10.1037/h0020278

Linville, S.E. The sound of senescence. J Voice. 1996 Jun;10(2):190-200. doi: 10.1016/s0892-1997(96)80046-4. PMID: 8734394.

Liss, J.M., White, L., Mattys, S.L., Lansford, K., Lotto, A.J., Spitzer, S.M. and Caviness, J.N. (2009). Quantifying Speech Rhythm Abnormalities in the Dysarthrias. *Journal of Speech, Language, and Hearing Research*, 52(5), 1334–1352.
doi: 10.1044/1092-4388(2009/08-0208)

Liu, C.-T. and Chen, Y. (2021). Consonantal Landmarks as Predictors of Dysarthria among English-Speaking Adults with Cerebral Palsy. *Brain Sciences*, 11(12), 1550.
doi: 10.3390/brainsci11121550

Liu, S. A. (1996). Landmark detection for distinctive feature-based speech recognition. *The Journal of the Acoustical Society of America*, 100(5), 3417–3430.
doi: 10.1121/1.416983

Long, S., Benoist, C. and Weidner, W. (2023). *World Alzheimer Report 2023: Reducing dementia risk: never too early, never too late*. London, England: Alzheimer's Disease International.https://www.alzint.org/resource/world-alzheimer-report-2023/

Lowit, A., (2014) Quantification of rhythm problems in disordered speech: a re-evaluation, Philosophical Transactions of The Royal Society B, vol. 369, no. 1658, article 20130404.

Lowit, A, Marchetti, A, Corson, S, Kuschmann, A. (2018). Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks. *Journal of communication disorders*, 72:26–39.
doi: 10.1016/j.jcomdis.2018.02.005

MacAuslan, J. (2016). *What are Acoustic Landmarks, and What Do They Describe?* Retrieved from: https://speechmrk.com/category/blog/tutorials-and-user-guides/ (*Last access*: 16/03/2024)

MacAuslan, J., Boyce, S., (2016). Peak Landmarks in SpeechMark Retrieved from: https://speechmrk.com/category/blog/tutorials-and-user-guides/ (*Last access*: 05/04/2024)

Maffia, M., De Micco, R., Pettorino, M., Siciliano, M., Tessitore, A. and De Meo, A. (2021) Speech Rhythm Variation in Early-Stage Parkinson's Disease: A Study on Different Speaking Tasks. *Frontiers in Psychology*, 12:668291.
doi: 10.3389/fpsyg.2021.668291.

Martínez-Nicolás, I., Llorente, T.E., Martínez-Sánchez, F. and Meilán, J.J.G. (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.
doi : 10.3389/fpsyg.2021.620251

Measso, G., Cavarzeran, F., Zappalà, G., Lebowitz, B. D., Crook, T. H., Pirozzolo, F.J., Amaducci, L.A., Massari, D. and Grigoletto F. (1993). The mini-mental state examination: normative study of an italian random sample. *Developmental Neuropsychology*, 9(2): 77–95.
doi: 10.1080/87565649109540545

Meilán, J.J.G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T.E. and Carro, J. (2020).

Changes in the Rhythm of Speech Difference between People with Nondegenerative Mild Cognitive Impairment and with Preclinical Dementia. *Behavioural Neurology*, 2020: 4683573. doi:10.1155/2020/4683573

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L. and Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699.doi: 10.1111/j.1532-5415.2005.53221.x.

Nicastro N, Malpetti M, Cope TE, Bevan-Jones WR, Mak E, Passamonti L, Rowe JB, O'Brien JT. Cortical Complexity Analyses and Their Cognitive Correlate in Alzheimer's Disease and Frontotemporal Dementia. J Alzheimers Dis. (2020);76(1):331-340. doi: 10.3233/JAD-200246. PMID: 32444550; PMCID: PMC7338220.

Novelli, G., Papagno, C., Capitani, E., Laiacona, M., Vallar, G. and Cappa, S.F. (1986). Tre test clinici di ricerca e produzione lessicale. Taratura su soggetti normali. *Archivio di Psicologia, Neurologia e Psichiatria*, 4, 477–506.

Ortmann, K. (2022). Fine-Grained Error Analysis and Fair Evaluation of Labeled Spans. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pp. 1400–1407, Marseille, France, 21-23 June 2022. European Language Resource Association (ELRA).
https://aclanthology.org/2022.lrec-1.150.pdf

Pellegrino, E., He, L., Dellwo, V. (2018). The Effect of
Ageing on Speech Rhythm: A Study on Zurich German. In: Speech Prosody 2018, Poznan, 13 June2018 - 16 June 2018. ISCA, 133-137.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G. and Kokmen, E. (1999). Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology*, 56(3), 303.
doi: 10.1001/archneur.56.3.303

Pettorino, M., Maffia, M., Pellegrino, E., Vitale, M., & Meo, A. D. (2013). *VtoV: a perceptual cue for rhythm identification*. In P. Mertens & A.C. Simon (Eds), *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, pp. 101–106, Leuven, Belgium, 11-13 September 2013. https://www.arts.kuleuven.be/ling/cohistal/conference/idp2013/documents/proceedings_idp2013

Pettorino, M., Busà, M.G. and Pellegrino, E. (2016). Speech Rhythm in Parkinson's Disease: A Study on Italian. In *Proceedings of Interspeech 2016*, 1958–1961. San Francisco (CA), USA, 8-12 September 2016.
doi: 10.21437/Interspeech.2016-74

Pirani, A., Benini, L., Codeluppi, P.L., Ricci, C., Casatta, L., Lovascio, S., Pellegrini, M., Mazzoleni, F. and Brignoli, O. (2017). Il GPCog nel case-finding del deterioramento cognitivo in Medicina Generale: esperienze nella pratica ambulatoriale. *Rivista Società Italiana di Medicina Generale*, 6(24): 20–24.

Poeppel, D., Assaneo, M.F. Speech rhythms and their neural foundations. Nat Rev Neurosci 21, 322–334 (2020). https://doi.org/10.1038/s41583-020-0304-4

Ramus, F., Nespor, M. and Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, 75(1), AD3–AD30. doi: 10.1016/S0010-0277(00)00101-3

Scholtz, S. (2007). Acoustic Analysis of Adult Speaker Age. In C.Müller [Ed.], Speaker Classification I, LNAI 4343, Berlin, Heidelberg: Springer, pp. 88-107.

Stevens, K.N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 111(4), 1872–1891.
doi: 10.1121/1.1458026

Stevens, K.N., Manuel, S.Y., Shattuck-Hufnagel, S., Liu, S. (1992). Implementation of a model for lexical access based on features. *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP 1992)*, pp. 499-502. Banff, Alberta, Canada, 13-16 October 1992.
doi: 10.21437/ICSLP.1992-161

Themistocleous, C., Eckerström, M. and Kokkinakis, D. (2018). Identification of Mild Cognitive Impairment from Speech in Swedish Using Deep Sequential Neural Networks. *Frontiers in Neurology*, 9:975.
doi: 10.3389/fneur.2018.00975

Themistocleous, C., Eckerström, M. and Kokkinakis, D. (2020). Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls. PLoS ONE 15(7): e0236009.doi: 10.1371/journal.pone.0236009

Tremblay P, Poulin J, Martel-Sauvageau V, Denis C. Age-related deficits in speech production: From phonological planning to motor implementation. Exp Gerontol. 2019 Oct 15;126:110695. doi: 10.1016/j.exger.2019.110695. Epub 2019 Aug 22. PMID: 31445106.

Valentine, H., MacAuslan, J., Grigos, M., & Speights, M. (2023). My Vowels Matter: Formant Automation Tools for Diverse Child Speech. In *Proceedings of Interspeech 2023*, pp. 674–675. Dublin, Ireland, 20-24 August 2023.
doi: 10.21437/Interspeech.2023

Velayudhan, L., Ryu, S. H., Raczek, M., Philpot, M., Lindesay, J., Critchfield, M. and G. Livingston (2014). Review of brief cognitive tests for patients with suspected dementia. *International psychogeriatrics*, 26(8):1247–1262.

doi: 10.1017/S1041610214000416

9. Language Resource References

Albano Leoni et. al. 2007. CLIPS Corpus. http://www.clips.unina.it/it/corpus.jsp.

Gagliardi G. et al. 2016. OPLON: Opportunities for active and healthy LONgevity Corpus. http://hdl.handle.net/20.500.11752/ILC-992

Appendix A

List of Landmarks detected by Speechmark

The following table summarizes the landmark symbols, the acoustic events they represent, and the rules adopted by Speechmark for detecting them.

(source: MacAuslan, 2016)

Symbol	Mnemonic	Rule
+g	Glottal onset	Beginning of sustained laryngeal vibration, i.e., of periodicity or of power and spectral slope similar to that of a nearby segment of sustained periodicity
-g	Glottal offset	End of sustained laryngeal motion
+p	Periodicity onset	Beginning of sustained periodicity of appropriate period
-p	Periodicity offset	End of sustained periodicity of appropriate period
+j	F0 jump upward	Abrupt upward jump in F0 by at least 0.1 octave (approx.)
-j	F0 jump down	Abrupt downward jump in F0 by at least 0.1 octave (approx.)
+b	Burst onset	At least 3 of 5 frequency bands show simultaneous power increases of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in an unvoiced segment (not between +g and the next -g)
-b	Burst offset	At least 3 of 5 frequency bands show simultaneous power decreases of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in an unvoiced segment
+s	Syllabic onset	At least 3 of 5 frequency bands show simultaneous power increases of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in a voiced segment (between +g and the next -g)
-s	Syllabic offset	At least 3 of 5 frequency bands show simultaneous power decreases of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in a voiced segment
+f	Frication onset	At least 3 of 5 frequency bands show simultaneous 6-dB power increases at high frequencies and decreases at low frequencies (unvoiced segment)
-f	Frication offset	At least 3 of 5 frequency bands show simultaneous 6-dB power decreases at high frequencies and increases at low frequencies (unvoiced segment)
+v	Voiced frication onset	At least 3 of 5 frequency bands show simultaneous 6-dB power increases at high frequencies and decreases at low frequencies (voiced segment)
-v	Voiced frication offset	At least 3 of 5 frequency bands show simultaneous 6-dB power decreases at high frequencies and increases at low frequencies (voiced segment)