VOWEL FORMANT VARIABILITY IN THE SPEECH OF PATIENTS WITH DEMENTIA: A PRELIMINARY ANALYSIS

Gloria Gagliardi[1], *Chiara Meluzzi*[2]

1. INTRODUCTION

In the last decade, language disruption has been increasingly recognized as an early hallmark of neurodegenerative diseases (Boschi *et al.*, 2018). This evidence should not come as a surprise given that language is a complex cognitive function that leans upon widely distributed neural substrates (Catani, Bambini, 2014). Therefore, even minor brain lesions may trigger subtle language changes.

Thus, a growing interest surrounded possibility of exploiting speech analysis for screening and diagnostic purposes: in this context, dementia represents a very fruitful ground, with a a larger amount of evidence supporting the presence of phonetic alterations both in prodromal stages and in full-blow pathology (for Italian, cf. Beltrami *et al.*, 2018; Calzà *et al.*, 2021; Gagliardi, Tamburini, 2021). As a matter of fact, a large body of studies discriminate subjects with dementia (henceforth, SWD) and healthy control subjects (CS) in different languages with good accuracy through AI and Natural Language Processing technologies (a technique referred to as "Digital Linguistic Biomarkers" – DLBs, cf. Gagliardi, 2023).

However, the effect of dementia on voice and speech has been mostly considered i) quantitatively and ii) in terms of durational and prosodic cues, almost neglecting the segmental level. Nevertheless, the neural underpinnings of these voice abnormalities remain almost entirely unexplored, unlike lexical, syntactic, and pragmatic deficits that can be explained by the "typical" localization of brain atrophies (e.g., medial temporal lobe and hippocampus for Alzheimer's Disease) and interrelated with other cognitive impairments (e.g., episodic/semantic memory, executive functions).

This paper aims at exploring this grey area. In particular, we will tackle three research questions:

1. Do vowel formants dynamics show a qualitative difference in SWD versus CS?
2. Is there a quantitative difference in vowel formants between SWD and CS?
3. Is the Vowel Space of SWD sensibly reduced compared to CS?

To answer these issues, we will perform an acoustic analysis on the speech of two patients with dementia and two healthy controls (balanced by sex, age, and geographic origin), especially focusing on vowel space and formants' contour.

The paper will be structured as follows: first, it will present an overview of the segmental and suprasegmental characteristics of the language in patients with dementia

[1] Alma Mater Studiorum - Università di Bologna.
[2] Università degli Studi di Milano.
This paper has been jointly conceived and written by the two authors. However, for the requirements of the Italian Academy, Gloria Gagliardi is responsible for sections 1, 2.1, 3.1 and 6, while Chiara Meluzzi is responsible for sections 2.2, 3.2, 4 (and sub-sections) and 5.

(§ 2.1) and a brief synopsis of vowels analysis in pathological speech (§ 2.2). Second, it will illustrate data collection and analysis strategies, i.e., patients' enrollment (§ 3.1) and annotation protocol (§ 3.2). Then, the findings of this pilot study will be reported and discussed considering the previous literature (§ 4 and 5). Finally, preliminary conclusions and future perspectives will be drawn (§ 5).

2. THEORETICAL REMARKS

As mentioned in the introduction, voice and speaking problems are very common in people who suffered from neurodegenerative disease. This section aims to review some of the most remarkable aspects of the scientific literature on phonetic changes due to dementia. It will then give an overview of the current and historical landscape of vowel analysis applied to pathological speech.

2.1. *Speech in Dementia*

Over the last decades, the linguistic productions of cognitively frail patients have attracted increasing attention and have become an established research topic in the academic and medical community. Several changes in speech patterns have been observed in connection to the emergence of various types of dementia (Szatloczki *et al.*, 2015; Cummings, 2020), particularly in patients with Alzheimer's Disease (AD), Vascular Dementia (VaD), and "Mixed" Dementia, which are the object of this study.

From a linguistic perspective, AD has been the most examined pathology. It has been widely held that linguistic symptoms exclusively affect the lexical and semantic levels (i.e., ranging from word-finding problems to sentence comprehension deficits and verbal fluency decrease), and the pragmatic skills (i.e., deficits usually affect referential/temporal cohesion, coherence, discourse planning, and non-literal language). At the same time, morpho-syntactic competence is relatively unimpaired, at least until an advanced stage of disease (i.e., according to the literature, the structures produced by AD patients are usually correct but "reduced"; Kempler *et al.*, 1987, 1993). Moreover, the common view on language in dementia foresees that phonetic and phonological skills are generally spared in adults with AD (Emery, 2000). This claim is, actually, the object of a strong debate. As a matter of fact, we can mention several literature exceptions to this statement (Croot *et al.*, 2000; Luchesi Cera *et al.*, 2013; Horley *et al.*, 2010), and our previous findings demonstrated that acoustic features are the most reliable parameters for dementia early identification through AI and NLP techniques (Calzà *et al.*, 2021; Gagliardi, Tamburini, 2021).

A stream of evidence has shown that patients with VaD usually experience communication impairments too. Some scholars (e.g., Vuorinen *et al.*, 2000) stated that these language deficits resemble those observed in AD, while other investigators (e.g., Kontiola *et al.*, 1990) found different patterns of language disruptions. The main linguistic disturbance due to the disorder includes impaired fluency, syntactic comprehension, and expression deficits, anomia and lexical retrieval difficulties, and pragmatic/conversational disturbances. Conversely, morphology usually represents an area of linguistic strength (Cummings, 2020). However, these shreds of evidence are far from conclusive.

In summary: while lexical, syntactic, and pragmatic disturbances have been widely described, the nature (and the relevance) of voice abnormalities remains mostly unexplored. Furthermore, scholars have mostly tackled this domain through computational techniques, focusing on durational cues – such as pauses and speech rate

– and prosodic measures – namely Fundamental Frequency (F0) minimum, maximum, mean, standard deviation, and range – (Canning *et al.*, 2005; König *et al.*, 2015; Carozza *et al.*, 2011; Vincze *et al.*, 2020; Calzà *et al.*, 2021). On the other side, the qualitative and quantitative analysis of segmental features has been almost neglected, with only a few exceptions. To the best of our knowledge, consonantal properties have been little considered: only Baker *et al.* (2007) compared the Voice Onset Time (VOT) of six voiceless and voiced stop consonants (i.e., /p/, /t/, /k/ and /b/, /d/, /g/) produced by individuals with moderate AD and healthy age- and gender-matched peers, with no statistically relevant results. Jitter and Shimmer (i.e., respectively, frequency and amplitude fluctuations) were sporadically estimated as a proxy measure of vocal stability along the course of the disease (Meilán *et al.*, 2012, 2014; 2018; König *et al.*, 2015). Even more compelling for our study, Nishigawa *et al.* (2022) and Shamei *et al.* (2023) demonstrated a general reduction of vowel space respectively in Japanese and English patients during spontaneous speech. However, vowel distance is not reduced in Spanish read speech (Shamei *et al.*, 2023): according to the authors, the reduction of movement trajectories during vowel articulation i) results from the impairments to fine motor skills in AD, ii) may be influenced by the task (i.e., spontaneous vs. read speech), and iii) is language dependent.

To date, various algorithms allow a reliable extraction of all these acoustic features from speech signals (e.g., openSMILE, cf. Eyben *et al.*, 2010, 2015). However, we believe that a deeper understanding of phonetic changes due to dementia can only pass through accurate manual analysis.

2.2. *Vowel analysis and Pathological speech*

Clinical phoneticians have frequently analyzed vowels as a means to highlight different patterns between pathological and non-pathological speakers. As in general phonetic research, three main analyses are usually offered on pathological subjects, with a preference for one or the other according to the research protocol and the scholars: target-analysis (or static analysis), dynamic analysis, and vowel space analysis (see Van der Harst, 2011, among others, for a critical overview). Regarding elicitation methods, word or sentence reading tasks are usually preferred, along with sustained vowels realizations, depending on the language pathology being investigated. A comparison between pathological and non-pathological (or control) speakers is always supposed in this field of study.

With respect to vowel analysis, among the most-studied pathological populations are speakers with Parkinson's disease or other types of dysarthria (e.g., Whitfield, Metha, 2019; Le *et al.*, 2017), apraxia of speech (den Ouden *et al.*, 2018), but, mostly, speech by individuals with deafness or severe hearing impairment, typically before and after cochlear implants (see also Meluzzi, 2021: 7-8, for a detailed overview).

Despite the large (and growing) body of literature on vowel analysis in pathological speech, there is an astonishing lack of interest in vowel analysis in speakers with Dementia. Indeed, scholars have primarily focused on durational or phonological cues, such as speech rate and pauses (cf. Canning *et al.*, 2004; Carozza *et al.*, 2011; König *et al.*, 2015). As for segmental cues, only a few recent studies on Japanese speakers with Dementia have specifically considered vowel variability as an important feature in the discrimination of this pathology, also from an automatic machine-learning-oriented approach. Two works by Nishikawa *et al.* (2022a and 2022b) are based on a corpus of 6 pathological elderly speakers confronted with 10 control speakers. So far, their results have shown a lowering

of formant values, particularly F1, in speakers with Dementia (SWD) compared to their controls (CS).

3. MATERIALS AND METHODS

In the present work, an attempt was made to provide a preliminary understanding of formant dynamics and vowel space reduction brought on by dementia. Under this assumption, we analyze the verbal production of two elderly patients diagnosed with dementia and two control subjects extracted from the linguistic resource DemCorpus-Basilicata (Martinelli *et al.*, 2022).

The research protocol was approved by the ethical committee of the University of Bologna (n. 0072032/2022). The study follows the principles of the Declaration of Helsinki.

3.1. *Data Collection*

The corpus consists of 8 hours and 50 minutes of semi-spontaneous speech by 40 Italian individuals living in Basilicata (i.e., an administrative region in Southern Italy) resident in the nursing home RSA Universo Salute – Opera Don Uva (PZ). The speakers are divided into two groups balanced by sex (24 F, 16 M) and age (SWD age = 81 ± 6.9, range: 63-91; CS age = 81 ± 6.3, range: 63-92): 20 healthy subjects and 20 subjects affected by different types of dementia (i.e., Alzheimer's disease, mixed dementia, frontotemporal dementia, and vascular dementia). Both cohorts signed informed consent to take part in the study and privacy consent for the treatment of their personal data.

After the neuropsychological evaluation (i.e., *Mini-Mental State Examination* – MMSE, Magni *et al.*, 1996; *MOntreal Cognitive Assessment* – MoCA, Conti *et al.*, 2015; phonemic and semantic fluency, Caltagirone *et al.*, 1995; Carlesimo *et al.*, 1995; Spinnler, Tognoni, 1987), the speech was elicited by administering three elicitation tasks:

- Task 1**:** Narrative – Tale of a journey.
- Task 2: Narrative – Christmas tradition.
- Task 3: Picture description ("Esame del Linguaggio II"; Ciurli *et al.*, 1996).

Table 1 summarizes the enrollment criteria for the cohorts.

Table 1. *Enrollment criteria*

	SWD	**CS**
Inclusion criteria	Age > 60 Monolingual linguistic exposure L1 = Italian Mild/moderate cognitive decline: - MMSE < 22 - MoCA ≤ 19,262 - phonemic fluency < 17,35 - semantic fluency < 7,25	Age > 60 Monolingual linguistic exposure L1 = Italian No cognitive decline - MMSE ≥ 22 - MoCA > 19,262 - phonemic fluency ≥ 17,35 - semantic fluency ≥ 7,25

Exclusion criteria	- Blindness, deafness, or other serious sensory impairments. - Acute neurological diseases (e.g., recent stroke, traumatic brain injury). - Severe psychiatric diseases (e.g., major depressive disorder).	- Blindness, deafness, or other serious sensory impairments. - Acute neurological diseases (e.g., recent stroke, traumatic brain injury). - Severe psychiatric diseases (e.g., major depressive disorder).

For this pilot study, we focused on two patient-control couples: SOG1 – SOG21 and SOG10 – SOG33. Table 2 outlines the main demographic characteristics of these subjects.

Table 2. *Main characteristics of the considered subjects*

	SWD	**CS**
1	SOG1 Female; Age: 63 Diagnosis: Alzheimer's Disease	SOG21 Female; Age: 63 No cognitive decline
2	SOG10 Male; Age: 73 Diagnosis: Mixed dementia	SOG33 Male; Age: 73 No cognitive decline

3.2. *Annotation Protocol*

We develop an annotation protocol to allow for both segmental and supra-segmental analysis of pathological speech, and, in particular, for speakers with Dementia. The annotation consists of three tiers, namely sentence, word, and vowel. The sentence tier provides an orthographic transcription of what the speaker actually said, with a few symbols similar to those already used in Conversation Analysis. In particular, we transcribed vowel or consonant lengthening (with ":"), self-interruptions and reformulations (with "/"), truncated words (with "-"), and overlaps (with "#"). It should be noted that overlaps are usually minimal in the recordings, as the researcher's main aim was to allow the patient to speak freely. Other labels included unintelligible speech (labeled as "XXX") and portions where the speaker switched to his/her Italo-Romance dialects (labeled as "<DIALECT>").

On the second tier, single words were isolated, and any pauses between one word and another were annotated with a "P" symbol. On both the first and second tiers, personal names or place name that could identify the speakers were anonymized by labelling those portions as "[NAME]". These elements were not annotated in the third tier.

In the third tier, each vowel was isolated based on the F2 transition at the beginning and at the end (cf. Di Paolo, Yaeger-Dror, 2011). For vowel quality annotation, X-Sampa symbols were preferred to facilitate the export of acoustic data into Excel or SPSS software[3]. The central vowel [ə] was labelled as "SCHWA". Diphthongs were annotated with the corresponding X-SAMPA symbols, as previously explained (e.g. "Ej" for [ɛj]).

[3] For the sake of clarity, we remind the reader that, in the case of vowels, the only differences in the X-Sampa symbols compared to IPA are for the mid-low vowels [ɛ] and[ɔ]. For these vowels, we have used the labels "E" and "O", respectively.

All vowels were tagged within the same label for three additional features:

- Accent: "_UN" for unstressed, and "_ST" for stressed vowels.
- Nasalization: "_NAS" for nasalized portions of the vowel.
- Rhoticization: "_ROT" for vowels that were slightly or heavily rhoticized.

The accent tag was added to all vowels in the corpus, to analyze differently stressed and unstressed ones. As for nasalization and rhoticization, these labels were added only in case the vowels showed a more or less extended portion of both phenomena, as evident mainly from the spectrogram and waveform, but also acoustically.

On the annotated files, a PRAAT script for extracting vowel formants was applied to semi-automatically extract those values. We preliminary checked the main intervals of the first three formants to better set the script for each vowel (cf. Barreda, 2021). The script worked with a window-length of 0.025 seconds and a pre-emphasis of 50 Hz. The resulting matrix includes values for the first three formants (F1, F2, F3) and intensity, measured both at the midpoint of the vowel interval and at 7-points along the vowel trajectory. The final matrix was then cleaned by checking for outliers (see 4.1) and for any remaining mismatching values,–particularly for /u/ sounds (see also D'Aco, Meluzzi, 2020).

4. RESULTS

We divide the analysis into three main sessions. Firstly, we explore the dataset descriptively, also highlighting potential issues in data distribution across different speakers or subsets (namely, male vs. female pathological speakers and their control speakers). A quantitative analysis is then performed on the vowels' formants as extracted at the midpoint; at this stage, we have decided to exclude diphthongs from the analysis. Finally, a qualitative explorative analysis considers vowels trajectories across time: the purpose was to compare the steadiness of pathological speakers vs. control speakers by means of formant values extracted at 7-time points across the vowels.

In the different sections, statistical analysis was performed with IBM SPSS 21, which was also used to create boxplots. Graphs showing vowel spaces and vowel trajectories were created using the open-source tool Visible Vowels (Heering, Van De Velde, 2018).

4.1. *Descriptive analysis*

After the annotation, the number of tokens available for the analysis is summarized in Table 3, distinguishing among speakers and vowel stress patterns.

Table 3. *Vowel tokens available for the considered subjects*

	SWD		**CS**	
	MALE	**FEMALE**	**MALE**	**FEMALE**
Stressed Vowels	1030	107	415	431
Unstressed Vowels	2623	211	1077	1111
Diphthongs	331	0	104	149
Total	3984	318	1596	1691

Diphthongs have been grouped into one general category, as this issue will not be addressed in the present work. As is evident, the corpus is highly unbalanced: the male SWD, in particular, produces the largest number of tokens in his recordings, whereas the female SWD utters only 336 vowels. This obvious difference is linked to their different pathological conditions and underscores the difficulty of building a balanced corpus of pathological speech for phonetic analysis (cf. Mildner, 2017).

As for nasalization, it occurred in a small portion of the dataset: 152 tokens for the male SWD (4.2X%), 41 tokens for the female SWD (12.3%), 110 tokens for the male CS (7.4%), and 136 tokens for the female CS (8.8%). Rhoticity was annotated in an even smaller number of tokens: 55 tokens for the male SWD (1.5%), 4 tokens for the female SWD (1.2%), 13 tokens for the male CS (0.9%), and 10 tokens for the female CS (0.6%). However, since both these phenomena could modify formant behavior, especially in a dynamic analysis, tokens labeled for either nasalization or rhoticity will be excluded from the present analysis.

4.2. *Quantitative analysis*

Descriptive statistics were first performed on formant values calculated at the midpoint. Each dataset (i.e., each speaker's productions) was divided by vowel quality and accent (stressed vs. unstressed). After an initial correction of the dataset for outliers, we run ANOVA analysis to compare pathological speakers with their respective controls. Firstly, we considered stressed vowels only separated for male speakers (Table 4) and female ones (Table 5).

Table 4. *Formants' Mean (Hz), Standard Deviation and ANOVA for stressed vowels in male speakers (n = 1445); * indicates not significant values*

		SWD	**CS**	**ANOVA (F)**	**p-value**
/a/	F1	636.248 ± 15.2	676.248 ± 15.2	14.631	.0001
	F2	1423.2 ± 12.63	1538.89 ± 17.28	58.328	.0001
	F3	2641.72 ± 18.25	2466.94 ± 13.85	89.861	.0001
/e/	F1	461.604 ± 51.42	394.48 ± 82.96	33.018	.001
	F2	1879.885 ± 204.05	1969.26 ± 194.55	5.228	.024
	F3	2746.747 ± 126.45	2668.312 ± 140.4	9.728	.002
/ɛ/	F1	515.081 ± 77.98	476.118 ± 71.06	12.729	.0001
	F2	1778.867 ± 175.22	1905.346 ± 133.65	28.682	.0001
	F3	2739.401 ± 128.01	2656.112 ± 119.51	21.305	.0001
/i/	F1	359.801 ± 39.86	344.268 ± 23.86	.520	.472*
	F2	2088.459 ± 158.44	2188.179 ± 142.98	17.874	.0001
	F3	2855.03 ± 141.17	2784.648 ± 20.94	7.557	.007
/o/	F1	490.841 ± 52.96	509.509 ± 84.01	1.453	.232*
	F2	1236.721 ± 20.43	1391.37 ± 287.46	5.312	.024
	F3	2783.837 ± 138.3	2449.251 ± 20.64	73.572	.0001

/ɔ/	F1	542.701 ± 64.85	527.7 ± 64.85	1.324	.251*
	F2	1282.796 ± 287.09	1234.959 ± 375.49	.737	.392*
	F3	2743.399 ± 157.7	2401.209 ± 153.52	129.306	.0001
/u/	F1	414.283 ± 58.37	367.831 ± 27	8.749	.004
	F2	1291.757 ± 201.66	1429.379 ± 364.15	.951	.333*
	F3	2801.637 ± 157.929	2506.084 ± 149.239	40.671	.0001

The results show generally high standard deviation values for back vowels. Formants values are also slightly higher but remain consistent across individual speakers. Regarding formant differences, F2 and F3 are generally statistically significant, except for /ɔ/ and /u/. F1 usually distinguishes between SWD and CS, except for /i/ and /o/.

Table 5. *Formants' Mean (Hz), Standard Deviation and ANOVA for stressed vowels in female speakers (n = 538); * indicates not significant values*

		SWD	**CS**	**ANOVA (F)**	**p-value**
/a/	F1	587.411 ± 90.98	776.546 ± 114.23	61.239	.0001
	F2	1710.817 ± 22.54	1595.826 ± 120.99	12.570	.001
	F3	2759.776 ± 42.54	2594.504 ± 255.96	6.4	.014
/ɛ/	F1	497.864 ± 115.98	615.653 ± 93.59	20.589	.0001
	F2	2061.142 ± 37.36	1970,012 ± 141.346	2.753	.101*
	F3	2761.914 ± 249.91	2733.554 ± 125.347	4.711	.033
/i/	F1	381.71 ± 66.57	375.113 ± 49.05	.222	.639*
	F2	2317.587 ± 333.74	2289.78 ± 26.72	.144	.706*
	F3	3084.757 ± 324.38	3084.75 ± 32.38	9.354	.003
/o/	F1	437.695 ± 78.67	503.73 ± 70.13	5.88	.02
	F2	1086.962 ± 55.65	1102.45 ± 221.42	.016	.9*
	F3	2432.719 ± 34.7	2862.77 ± 30.51	13.047	.001
/ɔ/	F1	459.026 ± 124.96	595.199 ± 96.2	14.874	.0001
	F2	1072.049 ± 35.58	1255.384 ± 159.29	7.427	.008
	F3	2284.529 ± 214.32	2728.52 ± 263.37	23.439	.0001

For female speakers, the reduced sample made it impossible to conduct reliable analyses (e.g., for /u/ or /e/), since there were few or no instances for the SWD. The results reported in Table 5 indicate that the vowel /i/ is produced similarly by the two speakers, whereas the major (statistically significant) differences were found for /a/ and /ɔ/. SWD's formants are usually lower than those of her CS counterpart, except for F3, which displays higher values.

We also examined whether unstressed vowels showed a similar pattern. Table 6 reports the mean values and ANOVA results for the male speakers, while the corresponding values for female speakers are shown in Table 7.

Table 6. *Formants' mean (Hz) and ANOVA for unstressed vowels in male speakers (n = 3700)*

		SWD	**CS**	**ANOVA (F)**	**p-value**
/a/	F1	583.727 ± 72.15	553.674 ± 15.7	16.272	.0001
	F2	1432.93 ± 159.67	1608.509 ± 22.17	183.769	.0001
	F3	2686.806 ± 18.91	2532.385 ± 16.15	146.769	.0001
/e/	F1	481.546 ± 54.74	418.311 ± 186.13	55.997	.0001
	F2	1697.28 ± 225.786	1876.47 ± 20.86	104.097	.0001
	F3	2749.528 ± 159.33	2620.711 ± 18.69	94.464	.0001
/ɛ/	F1	497.387 ± 42.65	382.217 ± 24.92	25.458	.0001
	F2	1689.144 ± 20.69	1949.286 ± 188.58	5.539	.034
	F3	2671.332 ± 154.79	2610.031 ± 84.26	.554	.469*
/i/	F1	399.734 ± 53.32	438.791 ± 39.32	4.624	.032
	F2	1942.071 ± 19.87	2102.571 ± 19.67	95.185	.0001
	F3	2797.268 ± 139.02	2767.56 ± 266.32	3.694	.055*
/o/	F1	509.521 ± 69.65	465.537 ± 177.61	20.453	.0001
	F2	1375.052 ± 22.35	1453.445 ± 35.26	11.838	.001
	F3	2784.388 ± 176.67	2442.125 ± 26.32	426.689	.0001
/ɔ/	F1	528.719 ± 85.15	533.355 ± 85.15	.007	.35*
	F2	1265.872 ± 22.12	1333.722 ± 341.79	.180	.679*
	F3	2800.7 ± 18.33	2295.56 ± 132.7	19.427	.001
/u/	F1	460.965 ± 55.19	369.49 ± 71.15	.003	.955
	F2	1321.887 ± 23.55	1467.397 ± 44.65	9.623	.003
	F3	2798.21 ± 154.41	2500.95 ± 155.37	25.824	.0001
/ə/	F1	540.826 ± 75.64	447.05 ± 22.84	23.813	.0001
	F2	1640.975 ± 27.22	1808.44 ± 33.63	16.370	.0001
	F3	2861.154 ± 15.85	2698.551 ± 21.66	42.323	.0001

The differences between SWD and CS are usually statistically significant, albeit with some random exceptions (F3 for /ɛ/ and /i/, F1 and F2 for /ɔ/). SWD's values are also usually higher than CS for all vowels, especially in the case of F2.

Table 7. *Formants' mean (Hz) and ANOVA for unstressed vowels in female speakers (n = 1322)*

		SWD	CS	ANOVA (F)	p value
/a/	F1	537.926 ± 138.5	706.729 ± 103.39	105.545	.0001
	F2	1582.20 ± 352.6	1632.25 ± 141.97	2.991	.085*
	F3	2734.86 ± 339.07	2634.06 ± 275.143	5.541	.019
/e/	F1	450.288 ± 156.91	499.601 ± 86.4	8.764	.003
	F2	1943.7 ± 491.75	1973.7 ± 192.18	.481	.489*
	F3	2902.6 ± 358.61	2902.6 ± 385.61	8.294	.004
/i/	F1	439.106 ± 195.45	394.501 ± 148.88	2.502	.115*
	F2	2164.65 ± 496.1	2239.247 ± 243.04	1.995	.159*
	F3	3031.086 ± 344.23	2944.95 ± 268.52	2.898	.09*
/o/	F1	398.319 ± 12.71	515.273 ± 97.05	31.212	.0001
	F2	1493.73 ± 43.99	1302.08 ± 25.01	11.053	.001
	F3	2598.541 ± 38.94	2782.543 ± 28.51	8.614	.004

Despite having more tokens available for unstressed vowels, differences for female speakers are rarely significant. Notably, apart from /o/, F2 does not help differentiate between the pathological and control speaker. The vowel /i/ shows no statistically significant differences.

4.3. *Qualitative analysis*

Quantitative results are also supported by a qualitative inspection of the vowel spaces of the different subjects in our corpus. For male speakers (Figure 1), it appears that the pathological speaker has a reduced vowel space along both the F2 and, primarily, the F1 dimension. A centralization of the front mid vowels /e/ and /ɛ/ is also visible, as well as a lowering along the F1-axis for the back mid vowels /o/ and /ɔ/.

Figure 1. *Vowel space in male speakers*

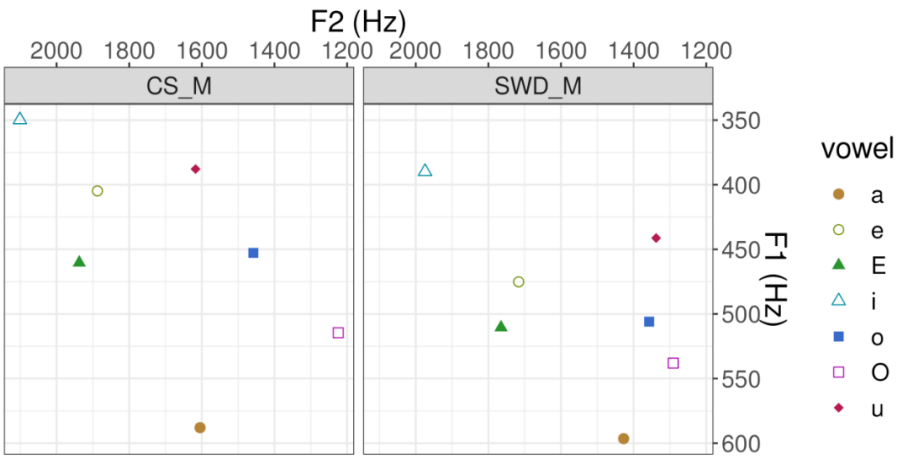
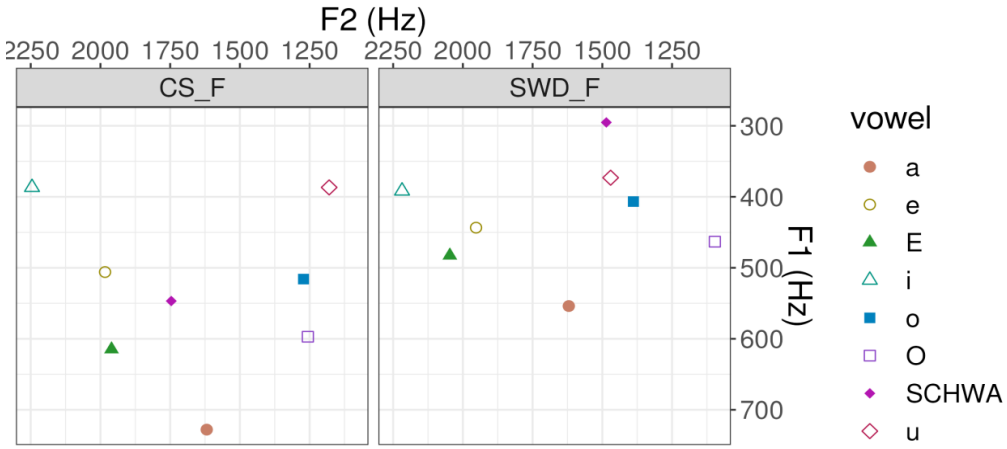


Figure 2. *Vowel space in female speakers*

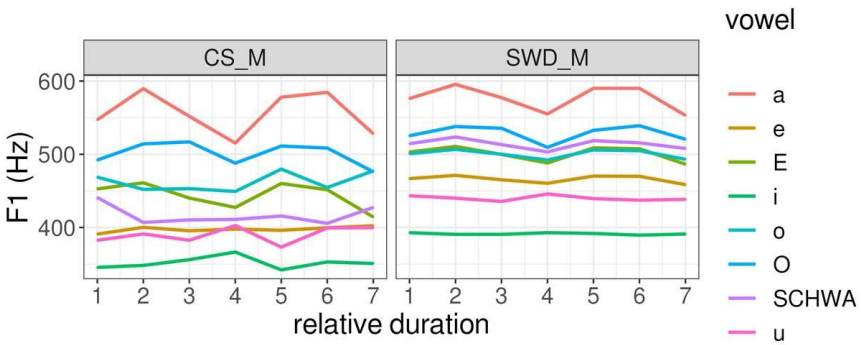


Female speakers (Figure 2) also exhibit ~~a~~ differences between SWD and CS, primarily along the F1 dimension, though in the opposite direction compared to male speakers. Specifically, the female SWD shows lower F1 values for all vowels. In this case, there is no evidence of centralization of the front mid vowels. Lastly, the few realizations of /ə/ by the female SWD display an apparent anomaly, which could be attributed to an interaction with F0, as this speaker has a peculiar voice quality consistent with a creaky voice.

We also investigated whether vowel formants reach a steady state during phonation or, conversely, if the variability across time is greater in pathological speakers compared to their controls. The analysis conducted at seven different time points throughout the vowel duration revealed variability, particularly for female speakers more than male ones, and on F2 more than F1 (see Figures 3 and 4).

Regarding F1 variability, Figure 1 highlights greater instability for the female SWD compared to the male SWD. The realization of schwa is, once again, highly variable over time for this speaker, supporting our earlier claim of a possible link with voice quality features. In contrast, the male SWD exhibits a pattern similar to his CS, but with smoother and flatter dynamics. This may suggest a low degree of co-articulation with preceding and following sounds as there is little to no variation between the start and end of the vowels (points 1 and 7 on the graph, respectively) and the middle points. A partial exception is observed for the /a/ and /ɔ/ sounds. Conversely, for the female SWD, vowel dynamics show greater variability at the midpoint, particularly for the vowels /ɛ/ and /u/. It worth noting that in both SWD speakers, the temporal pattern of /a/ mirrors that of their respective CSs, albeit in a smoother manner.

Figure 3. *Seven-points dynamic analysis for F1 (above: male speakers, below: female speakers)*



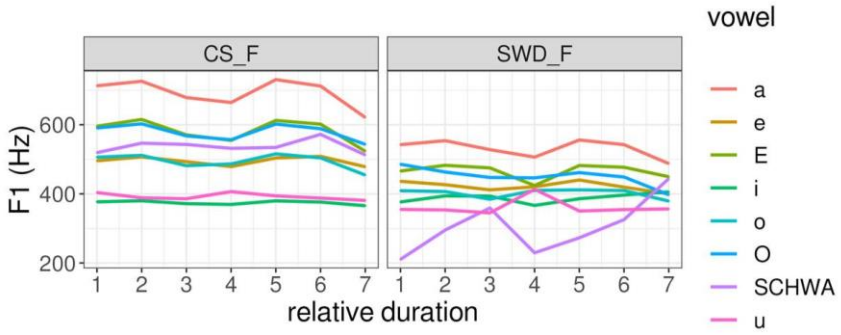
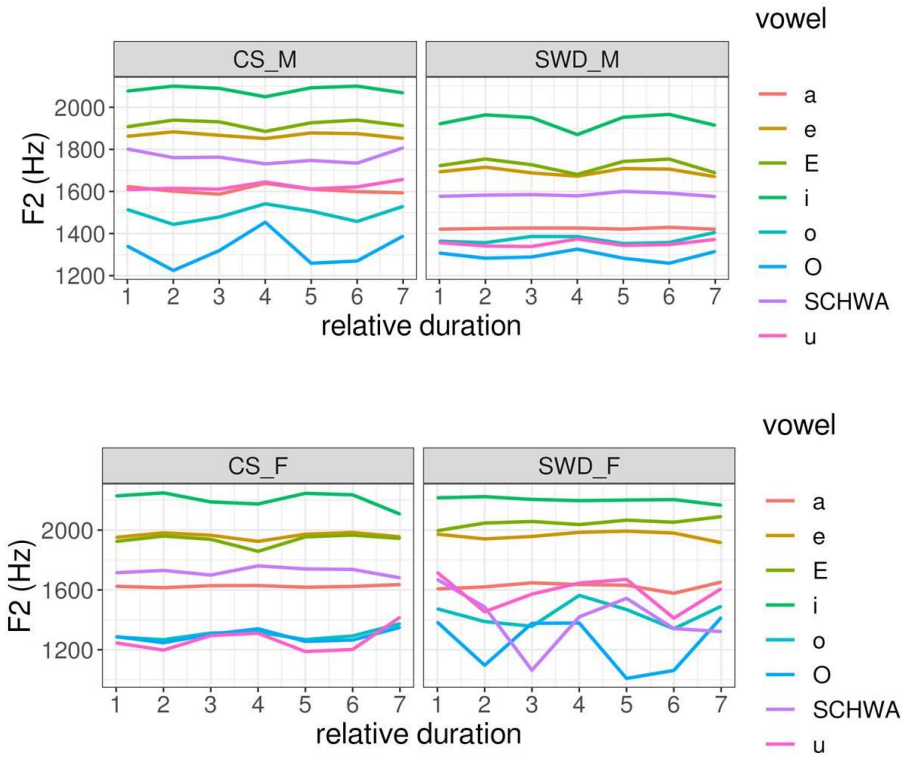


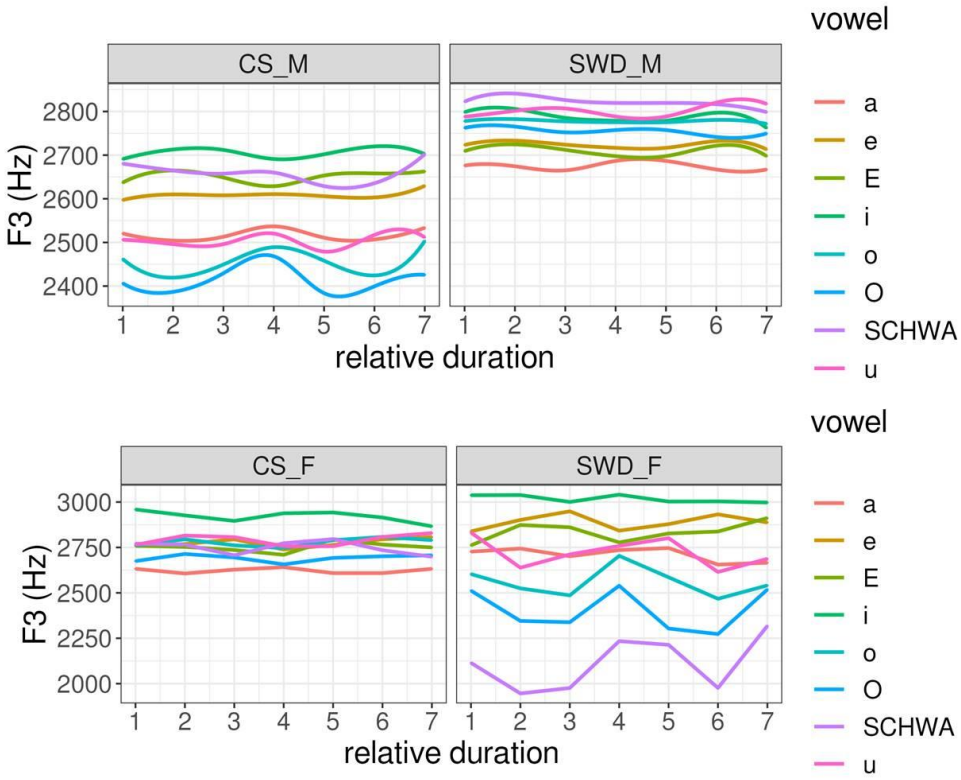
Figure 4. *Seven-points dynamic analysis for F2 (above: male speakers, below: female speakers)*



The temporal patterns observed for F1 in the male SWD are also confirmed by the dynamic analysis of F2 (Figure 4). Once again, the variability seen in the CS is attenuated in the SWD, showing little to no variability over time, except for the vowels /a/, /ɔ/ and, in the case of F2, also /ɛ/. In contrast, the female SWD displays extreme variability in the back vowels, including /ə/, which, as previously noted (Figure 2), is realized as a back sound.

The F3 contour analysis also revealed some interesting peculiarities between pathological and non-pathological speakers, as shown in Figure 5. The male SWD appears to have higher F3 values compared to his CS, but with smoother contours. Conversely, the female SWD is once again highly variable, with her vowel dynamics showing greater temporal instability, particularly for the back vowels.

Figure 5. *Seven-points dynamic analysis for F3 (above: male speakers, below: female speakers)*



5. Discussion

As stated at the beginning of this paper and in the result section, this work has, for now, two major limitations. On the one hand, we have offered a comparison between only two speakers with dementia and their respective controls. On the other hand, the variability in the corpus size and the poor number of tokens available for the female SWD constitutes a constraint for a strictly quantitative analysis. However, we believe these limitations could be addressed by considering both the intrinsic variability of speakers' pathological histories and the fact that we have compared pathological speakers with precisely matching controls, rather than comparing the speakers with each other. Furthermore, unlike the vast majority of studies in clinical phonetics, our data came from spontaneous (albeit guided) conversations. Finally, given the general lack of phonetic studies on vowels in speech in dementia, this work has yielded some important, albeit preliminary, results, which could be expanded upon in future research with a larger dataset.

In this respect, the analysis has highlighted a substantial difference in vowel formants variability between SWD and their CS, both quantitatively and qualitatively. A general reduction of the vowel space has been observed.

F1 dimension appears to be the most variable between pathological and non-pathological speakers, both quantitatively (see 4.2) and with respect to vowel space (Figures 1-2). This confirms a trend recently observed for Japanese by Nishikawa *et al.* (2022a). Front vowels tend to vary more than back vowels on the F1 dimension, while the opposite is true for F2, at least in the female SWD.

Finally, the dynamic analysis has highlighted generally smoother vowel contours for SWD compared to their controls. While the female SWD exhibits some exceptions in the F2 patterns of back vowels (Figure 4), the male SWD appears to mimic the patterns found

for his CS but with reduced variability. There are no substantial differences at the onset and offset of vowels (with the notable exception of /a/). This could be interpreted as a lack of co-articulatory patterns between the vowels and the preceding or following consonants. This is only partially in line with Nishikawa *et al.* (2002b: 24): indeed, in Japanese, a lower and more stable tongue position was found for all vowels, but particularly for /a/. This could be explained by a language-specific difference between Italian and Japanese, although it could be also linked to individual variability, as already hypothesized by Nishikawa *et al.* (2002a, 2002b).

6. Conclusions and further perspectives

This paper presented a preliminary investigation of vowel variation in speakers with dementia. To reach this goal, 2 Italian-speaking patients (1 male and 1 female) diagnosed with dementia were compared with 2 control speakers balanced for sex, age, and geographical origin. We performed a descriptive, static, and dynamic analysis of vowel formants, combining quantitative and qualitative techniques.

Our results showed a reduction of vowel space in speakers with dementia, especially along the F1 axis. A more stable vowel contour compared to control speakers has also been pinpointed. In line with previous scientific evidence, these results have been explained as a reduction of tongue movement, particularly on the vertical axis, during speech. However, further studies are needed to confirm this hypothesis.

By way of conclusion, voice analysis opens novel opportunities for health-related purposes. Vocal biomarkers can be exploited for diagnosis, risk prediction, and remote monitoring of various clinical syndromes (Fagherazzi *et al.*, 2021). However, in the near future, several challenges should be overcome for the efficient use of this technique in healthcare. Despite the high reliability of NLP techniques in detecting acoustic and temporal irregularities of speech in frail patients compared to healthy controls, a deeper understanding of the phonetic alteration due to neurodegeneration is essential. As a matter of fact, the use of AI methodologies raises a big deal known as the "black box problem", i.e., the inability of the programmer/user to see how deep learning systems make their decisions. This issue is crucial in a clinical setting since physicians should always be able to explain how a diagnosis was posed (Dashwood et al., 2021; Gagliardi, 2023). For this purpose, "traditional" acoustic analysis should parallel these novel methodologies.

References

Baker J., Ryalls J., Brice A., Whiteside J. (2007), "Voice onset time production in speakers with Alzheimer's disease", in *Clinical Linguistics & Phonetics*, 21, 11-12, pp. 859-867.

Barreda S. (2021), "Fast Track: fast (nearly) automatic formant-tracking using Praat", in *Linguistics Vanguard*, 7, 1: https://doi.org/10.1515/lingvan-2020-0051.

Beltrami D., Gagliardi G., Rossini Favretti R., Ghidoni E., Tamburini F., Calzà L. (2018), "Speech analysis by Natural Language Processing techniques: a possible tool for very early detection of cognitive decline?", in *Frontiers in Aging Neuroscience*, 10, 369, pp. 1-13: https://doi.org/10.3389/fnagi.2018.00369.

Boschi V., Catricalà E., Consonni M., Chesi C., Moro A., Cappa S. F. (2017), "Connected Speech in Neurodegenerative Language Disorders: A Review", in *Frontiers in Psychology*, 8, 269, pp. 1-21: https://doi.org/10.3389/fpsyg.2017.00269.

Caltagirone C., Gainotti G., Carlesimo G. A., Parnetti L. (1995), "Batteria per la valutazione del deterioramento mentale: I. Descrizione di uno strumento di diagnosi neuropsicologica", in *Archivio di Psicologia, Neurologia e Psichiatria*, 56, 4, pp. 461-470.

Calzà L., Gagliardi G., Rossini Favretti R., Tamburini F. (2021), "Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and Dementia", in *Computer, Speech & Language*, 65, 101113: https://doi.org/10.1016/j.csl.2020.101113.

Canning S. J., Leach L., Stuss D., Ngo L., Black S. E. (2004), "Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia", in *Neurology*, 62, 4, pp. 556-562.

Carlesimo G., Caltagirone C., Gainotti G., Nocentini U. (1995), "Batteria per la valutazione del deterioramento mentale: standardizzazione ed affidabilità diagnostica nell'identificazione di pazienti affetti da sindrome demenziale", in *Archivio di Psicologia Neurologia e Psichiatria*, 56, pp. 471-488.

Carozza L., Quinn M., Mack J., Bell-Berti F. (2011), "Temporal structure in the speech of a person with dementia: A longitudinal case study", in *Acta Neuropsychologica*, 9, 4, pp. 383-392.

Ciurli P., Marangolo P., Basso A. (1996), *Esame del Linguaggio II. Manuale e materiale d'esame*, Giunti, Firenze.

Catani M., Bambini V. (2014), "A model for Social Communication and Language Evolution and Development (SCALED)", in *Current Opinion in Neurobiology*, 28, pp. 165-171.

Conti S., Bonazzi S., Laiacona M., Masina M., Vanelli Coralli M. (2015), "Montreal Cognitive Assessment (MoCA) - Italian version: regression-based norms and and equivalent scores", in *Neurological Sciences*, 26, pp. 209-214.

Croot K., Hodges J. R., Xuereb J., Patterson K. (2000), "Phonological and articulatory impairment in Alzheimer's disease: a case series", in *Brain and language*, 75, 2, pp. 277-309.

D'Aco M., Meluzzi C. (2020), "Variabilità fonetica nelle produzioni di un adulto ipoudente con e senza feedback uditivo", in Dovetto F. (a cura di), *Lingua e Patologia*, *I sistemi* instabili, vol. 3, Aracne, Roma, pp. 401-408.

Dashwood M., Churchhouse G., Young M., Kuruvilla T. (2021), "Artificial intelligence as an aid to diagnosing dementia: An overview", in *Progress in Neurology and Psychiatry*, 25, 3, pp. 42-47.

Di Paolo M., Yaeger-Dror M. (2011), *Sociophonetics. A Student's Guide*, Routledge, London.

Eyben F., Wöllmer M. W., Schuller B. (2010), "OpenSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor", in *MM '10: Proceedings of the 18th ACM international conference on Multimedia*, *Florence, Italy*, ACM, pp. 1459-1462.

Eyben F., Scherer K. R., Schuller B. W., Sundberg J., André E., Busso C., Devillers L., Epps J., Laukka P., Narayanan S., Truong K. P. (2015), "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing", in *IEEE transactions on affective computing*, 7, 2, pp. 190-202.

Emery V. O. (2000), "Language impairment in dementia of the Alzheimer type: A hierarchical decline?", in *International Journal of Psychiatry in Medicine*, 30, 2, pp. 145-164.

Fagherazzi G., Fischer A., Ismael M., Despotovic V. (2021), "Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice", in *Digital Biomarkers*, 5, 1, pp. 78-88.

Gagliardi G., Tamburini F. (2021), "Linguistic biomarkers for the detection of Mild Cognitive Impairment", in *Lingue e Linguaggio*, XX, 1, pp. 3-31.

Gagliardi G. (2023), "Natural language processing techniques for studying language in pathological ageing: A scoping review", in *International Journal of Language Communication Disorders* (*IJLCD*), 59, 1, pp. 110-122: https://doi.org/10.1111/1460-6984.12870.

Heeringa W., Van de Velde H. (2018), "Visible Vowels: a Tool for the Visualization of Vowel Variation", in *Proceedings CLARIN Annual Conference 2018, 8-10 October, Pisa, Italy,* CLARIN ERIC, pp. 120-123.

Horley K., Reid A., Burnham D. (2010), "Emotional prosody perception and production in dementia of the Alzheimer's type", in *Journal of speech, language, and hearing research*, 53, 5, pp. 1132-1146.

Kempler D., Curtiss S., Jackson C. (1987), "Syntactic preservation in Alzheimer's disease", in *Journal of Speech and Hearing Research*, 30, 3, pp. 343-350.

Kempler S., LaBarge E., Ferraro F. R., Cheung H, Storandt M. (1993), "On the preservation of syntax in Alzheimer's disease. Evidence from written sentences", in *Archives of Neurology*, 50, 1, pp. 81-86.

Kontiola P., Laaksonen R., Sulkava R., Erkinjuntti T. (1990), "Pattern of language impairment is different in Alzheimer's disease and multi-infarct dementia", in *Brain and Language*, 38, 3, pp. 364-383.

König A., Satt A., Sorin A., Hoory R., Toledo-Ronen O., Derreumaux A., Manera V., Verhey F., Aalten P., Robert P. H., Renaud D. (2015), "Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease", in *Alzheimer's & dementia*, 1, 1, pp. 112-124: https://doi.org/10.1016/j.dadm.2014.11.012.

Luchesi Cera M., Zazo Ortiz K., Henrique Ferreira Bertolucci P., Soares Cianciarullo Minett T. (2013), "Speech and orofacial apraxias in Alzheimer's disease", in *International psychogeriatrics*, 25, 10, pp. 1679-1685.

Magni E., Binetti G., Bianchetti A., Rozzini R., Trabucchi M. (1996), "Mini-Mental State Examination: a normative study in Italian elderly population", in *European Journal of Neurology*, 3, 3, pp. 198-202.

Martinelli E., Garrammone V., Mori F., Nolè I., Cameriero F., Martino M., Di Bello G., Gagliardi G. (2022), "DemCorpus-Basilicata: Dementia Corpus, ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics 'A. Zampolli'", National Research Council, in Pisa: http://hdl.handle.net/20.500.11752/OPEN-989.

Meilán J. G., Martínez-Sánchez F., Carro J., López D., Millian-Morell L., Arana J. L (2012), "Acoustic markers associated with impairment in language processing in Alzheimer's Disease", in *The Spanish journal of psychology*, 15, 2, pp. 487-494.

Meilán J. G., Martínez-Sánchez F., Carro J., Sánchez J. A., Pérez E. (2014), "Speech in Alzheimer's Disease: Can Temporal and Acoustic Parameters Discriminate Dementia?", in *Dementia and Geriatric Cognitive Disorders*, 37, 5-6, pp. 327-334.

Meilán J. G., Martínez-Sánchez F., Carro J., Carcavilla N., Ivanova O. (2018), "Voice Markers of Lexical Access in Mild Cognitive Impairment and Alzheimer's Disease", in *Current Alzheimer Research*, 15, pp. 111-119.

Meluzzi C. (2021), "Sound Spectrography", in Ball M. J. (ed.), *Handbook of Clinical Phonetics*, Routledge, London, pp. 418-443.

Mildner V. (2017), "Experimental and quasi-experiental research in clinical linguistics and phonetics", in Müller N., Ball M. J. (eds.), *Research methods in clinical linguistics and phonetics*, Wiley-Blackwell, London, pp. 28-47.

Nishikawa K., Hirakawa R., Kawano H., Kakatoh Y. (2022a), "Analysis of Prosodic Features and Formant Dementia Speech for Machine Learning", in *Proceedings of the 5th International Conference on Information and Computer Technologies* (*ICICT*), pp. 173-176.

Nishikawa K., Kuwahara A., Hirakawa R., Kawano H., Kakatoh Y. (2022b), "Machine learning model for discrimination of mild dementia patients using acoustic features", in *Cognitive Robotics*, 2, pp. 21-29.

Shamei A., Liu Y., Gick B. (2023), "Reduction of vowel space in Alzheimer's disease", in *JASA Express Letters*, 3, 035202: DOI: 10.1121/10.0017438.

Spinnler H., Tognoni G. (1987), *Standardizzazione e taratura italiana di test neuropsicologici*, Masson Italia periodici (Supplementum 8 - *Italian journal of neurological sciences*), Milano.

Szatloczki G., Hoffmann I., Vincze V., Kalman J., Pakaski M. (2015), "Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease", in *Frontiers in Aging Neuroscience*, 7, pp. 195: DOI: 10.3389/fnagi.2015.00195.

Vincze V., Szatlóczki G., Tóth L., Gosztolya G., Pákáski M., Hoffmann I., Kálmán J. (2020), "Telltale silence: temporal speech parameters discriminate between prodromal dementia and mild Alzheimer's disease", in *Clinical Linguistics & Phonetics*, 35, 8, pp. 727-742. https://doi.org/10.1080/02699206.2020.1827043.

Vuorinen E., Laine M., Rinne J. (2000), "Common pattern of language impairment in vascular dementia and in Alzheimer disease", in *Alzheimer Disease and Associated Disorders*, 14, 2, pp. 81-86: DOI: 10.1097/00002093-200004000-00005.