Proceeding Paper

Acoustic and Temporal Analysis of Speech for Schizophrenia Management [†]

Alexantrina Mouratai [1], Nikolaos Dimopoulos [2], Athanasios Dimitriadis [2], Pantelis Koudounas [2], Dimitris Glotsos [1] and Luis Pinto-Coelho [3,*]

[1] Department of Biomedical Engineering, University of West Attica, 12243 Athens, Greece; alexandrina.mour@gmail.com (A.M.); dimglo@uwa.gr (D.G.)
[2] Dromokaiteio Hospital, 12461 Athens, Greece; dmpnikos@yahoo.gr (N.D.); th.dimitriadis@gmail.com (A.D.); koudounaspan@yahoo.gr (P.K.)
[3] Physics Department, Instituto Superior de Engenharia do Porto, INESCTEC, 4249-015 Porto, Portugal
* Correspondence: lfc@isep.ipp.pt
[†] Presented at the Advances in Biomedical Sciences, Engineering and Technology (ABSET) Conference, Athens, Greece, 10–11 June 2023.

Abstract: Currently, there are no established objective biomarkers for diagnosing or monitoring schizophrenia. Studies have shown that there are noteworthy differences in the speech of schizophrenics. The primary goal of the current study is to examine possible acoustic differences in vowel production between Greek speakers with schizophrenia and healthy controls. Eleven Greek speakers with schizophrenia and twelve healthy controls participated in the study. The results showed significant differences between the two groups in F1 and F2 frequencies, in jitter and shimmer as well as in the total length of pauses in spontaneous speech. These can pave the way for future developments toward the detection of disease patterns using inexpensive and non-invasive methods.

Keywords: schizophrenia; speech; acoustic features; phonetic features; pause; jitter; shimmer

1. Introduction

Speech analysis, encompassing its several dimensions (acoustic, linguistic, or suprasegmental), stands as an invaluable tool in the domain of pathology detection, revolutionizing the landscape of healthcare research and practice. Its profound capacity to discern subtle markers and patterns within speech has rendered it an indispensable asset in early disease diagnosis. The advent of speech analysis has engendered a paradigm wherein the human voice serves as a gateway to unraveling latent patterns pertaining to our physiological and psychological states. Acoustic speech analysis probes the intricate facets of tone, pitch, and vocal attributes, unveiling a trove of valuable information. Speech-related impairments, such as dysarthria [1] or dysphonia, as well as neurological diseases like Parkinson's disease [2], Alzheimer's [3], or amyotrophic lateral sclerosis [4], can be detected and monitored using voice biomarkers. This empowers healthcare professionals to intervene promptly, enabling expedited treatment and ultimately ameliorating the quality of life for affected individuals.

Conversely, linguistic speech analysis delves into the subtler realms of language usage, syntax, and verbal expression, unmasking significant insights into individuals' mental well-being. Linguistic cues, ranging from lexical choices to the structural organization of utterances, proffer profound glimpses into the emotional and psychological states of individuals. Detecting signs of depression [5], anxiety, or cognitive decline through meticulous linguistic analysis facilitates early identification, which, in turn, paves the way for timely interventions and ensures the provision of requisite support and care along the path to recovery.

1.1. Schizophrenia and Speech

Schizophrenia represents a severe mental disorder characterized by a significant disruption in perception, thought processes, and communication [6]. Symptoms associated with schizophrenia are typically categorized into two groups: (a) positive symptoms, such as delusions and hallucinations, which are those that appear to involve an excess or distortion of normal cognitive functions, and (b) negative symptoms, such as a diminished interest in activities, which reflect a reduction in, or loss of, typical functions. An illustration of a positive symptom is disorganized speech, which is characterized by incoherent and intense speech to the extent that it impairs effective communication [7]. On the other hand, a negative symptom, like alogia or "poverty of speech", is marked by a lack of speech output and evidences thought process disruptions. Another negative symptom is "Flat Affect", which entails a lack of emotional expression in individuals with schizophrenia, affecting facial expressions, voice tone, eye contact, and body language [8]. Negative symptoms are associated with even more unfavorable functional outcomes and a lower quality of life compared to the positive symptoms. They are also linked to impaired social functioning. Studies have indicated that a lifetime prevalence of prominent primary negative symptoms ranges from 15% to 20%, and this prevalence tends to increase with age [9]. Despite the debilitating nature and high prevalence of negative symptoms, treatments for them remain exceedingly limited [10]. Furthermore, the heterogeneity of "negative symptoms" presents significant challenges in treatment planning and research efforts. Due to the diverse manifestations of these symptoms, establishing a unified classification would greatly assist clinicians in monitoring fluctuations in severity over time and enhance their comprehension of the fundamental components underlying psychotic disorders. To address these pressing needs within our field, it is imperative to develop highly reliable and efficient measures for assessing specific negative symptoms, unhampered by the inherent limitations of qualitative clinical ratings that rely on subjective categorizations such as "mild", "moderate", or "severe". A key advancement in this domain would be the ability to objectively correlate negative symptoms with vocal/acoustic parameters. This correlation would allow us to adopt a dimensional approach to describing the disease, transcending traditional categorical diagnoses. Furthermore, leveraging speech as a potential biomarker holds promise in facilitating a deeper understanding of schizophrenia, offering insights that surpass the confines of conventional diagnostic frameworks. By embracing this innovative approach, we aim to pave the way toward a more nuanced and accurate diagnosis of schizophrenia, ultimately improving patient care and treatment outcomes. An examination of spontaneous speech in patients during interviews demonstrated a strong correlation between the duration of pauses and the assessments of flat affect and alogia by clinicians [11]. Furthermore, a noteworthy association was identified between the severity of negative symptoms and the variability in tongue position from front to back (measured as formant F2) [12]. In our study, we will examine the correlation between main frequencies, shimmer, jitter, and HNR rates as well as the pauses in spontaneous speech among schizophrenic patients.

1.2. Features Derived from Acoustic–Phonetic Analysis

Acoustic analysis is a precise and reliable scientific approach used to make more accurate evaluations of vocal traits. It can also be helpful in identifying vocal disorders and tracking alterations in vocal performance over time. Nonetheless, the outcomes of acoustic voice measurements are often affected by several factors, including ambient noise, data collection and analysis equipment, microphone type, and variations among the individuals being studied. Investigations have revealed that fundamental frequency measurements can be impacted by factors such as gender, variations within the same individual, and the type of microphone employed [13]. It is also reported that perturbation values are severely influenced by variations in estimation algorithms but also by gender, thus necessitating subgroup analyses to address these variables. The result of this study is discussed without reference to these parameters or the severity of symptoms among patients.

Fundamental frequency (F0) is defined as the lowest frequency of a periodic waveform. It is perceived as the loudest, and the ear identifies it as a specific pitch of the tone [14]. The F0 of a speaker's voice is a product of the vocal folds' length, tension, and mass/thickness during the production of sound. It is detectable when the vocal folds vibrate while articulating voiced sounds, namely vowels. If F0 values fall outside specific ranges that are well established for healthy voices, this can suggest the presence of a potential pathological condition.

Jitter and shimmer are the two common perturbation measures in acoustic analysis. Jitter is a measure of frequency instability, while shimmer is a measure of amplitude instability [15]. Perturbation refers to a disturbance in the regularity of a waveform and correlates to perceived roughness or harshness of the voice.

Another acoustic measure that may be a more sensitive index of vocal function is the harmonics-to-noise ratio (HNR). The HNR quantifies the relative amount of additive noise in the voice signal. The ratio reflects the dominance of harmonic content over noise levels in the voice, and it is quantified in terms of dB. In [16] it is reported that adult speakers with normal voice quality obtained HNRs of 7.4 dB and above when producing isolated vowels. In [17] it is also suggested that values between 11 and 13 dB are normal for healthy adults. The HNR seems to be one of the parameters that can be used to relate physiological aspects of voice production to a perceptual impression of the voice because the degree of spectral noise is related to the quality of the vocal output [18,19]. It was also reported that the correlation between reduced variability in the first and second formant frequencies (F1 and F2) and schizophrenia's negative symptom severity could be identified in English speakers [20]. A formant is a concentration of acoustic energy around a particular frequency in the speech wave. The shape of our vocal tract, and the position of the tongue in particular, affects the frequencies at which formants occur. The two lowest frequency formants F1 and F2 show the greatest variation based on the tongue position. The acoustic identity of vowels, defined by critical resonances at F1 and F2, is linked to the vocal tract morphology and biodynamics during vowel articulation. Specifically, F1 is indicative of the opening of the jaw and, consequently, the height of the tongue (as the jaw drops, the tongue moves downward), while F2 is related to the forward/backward positioning of the tongue and/or the rounding of the lips [21]. It is important to note that speech production follows a complex temporal pattern and rhythm. Speech units are sequentially arranged with brief intervals in between them, which can be bridged by the use of fillers (such as "erm" or "ah") or acoustic pauses [22].

It has been well established that individuals with schizophrenia exhibit distinctive patterns of linguistic organization in their spontaneous speech. These patterns often involve a reduction in syntactic complexity and an increase in syntactic errors [23]. This observation is not surprising, because there is a strong connection between thought and language. Speech production essentially involves the translation of thoughts into a sequential arrangement of spoken elements. Consequently, disruptions in thought processes are likely to manifest as disruptions in speech. In typical speech, pauses lasting anywhere from 250 to 3000 milliseconds are a natural and significant aspect of the cognitive and linguistic processes, constituting a substantial portion of total speech time. However, individuals with schizophrenia deviate from neurotypical controls in terms of the frequency of pauses, the proportion of silence, the overall duration of pauses, and the average duration of each pause, particularly during tasks like reading [24].

Considering this, the methods used will now be covered in the next section.

2. Methods

2.1. Data Collection

A cohort of 11 volunteer schizophrenic patients aged between 42 and 63 years (7 male and 4 female) agreed to participate in this study. In parallel, a complementary control group, consisting of 12 participants, was constituted. This control group was characterized by 7 male and 5 female subjects, their ages ranging from 21 to 81 years. It is crucial to underline

that all participants underwent rigorous health assessments, revealing sound physical well-being, and they lacked any pertinent personal or familial psychiatric antecedents. Furthermore, a shared trait among both the schizophrenia-afflicted and control subjects was their literacy. Within the scope of ethical considerations, a comprehensive exposition of the study's objectives and modus operandi was conveyed to each participant, culminating in the solicitation and documentation of written informed consent.

The recordings took place at the Psychiatric Hospital of Attica-Greece "Dromokaitio", and throughout each session, a standardized configuration was employed, consisting of a laptop and a cell phone, both with integrated microphones. The recordings were conducted simultaneously, with Praat [25] on the laptop and the Parrot voice recorder [26] on the mobile phone, at a sampling rate of 44,100 Hz. Audio files were recorded in a quiet room, but background noise was not particularly controlled. Our recording script for each session consist of three different tasks:

- Task 1: speaking the five Greek vowels (/a/, /o/, /u/, /i/, /e/) in a sustained manner for at least five seconds.
- Task 2: reading a standardized list of thirty words from a predefined script (constructed with the purpose of achieving a high phonetic diversity).
- Task 3: participating in a non-instructed interview where the participants were recorded while having a spontaneous talk. These recordings were used separately to extract the acoustic features that characterize the speech signal (all mentioned in Section 2.2).

2.2. Feature Extraction

Praat was used to extract the values of the phonetic linguistic parameters. The F0(Hz), HNR (dB), shimmer (%), and jitter (%) values were extracted for each one of the five Greek vowels that were recorded (Task 1) in the following way: The recordings were loaded in Praat, then all the vowels were selected separately and extracted into a new file. Next, Praat's toolkit was used (new trimmed sound selection/analyze periodicity/pitch (cc)/to point process (cc)/voice report). As soon as all the necessary values were collected, the mean and standard deviation (STD) were calculated for all the features.

Formant values F1 (in Hz) and F2 (in Hz) were extracted for each of the five Greek vowels recorded during Task 1. The vowel recordings were loaded into Praat, and each vowel was selected individually and extracted into a new file. The central time point of the spectrogram for each vowel was chosen to extract the F1 and F2 values. For each speaker, we calculated the mean value and standard deviation (STD) of F1 and F2.

A vocal toolkit was installed in Praat where the pauses were cut from the recordings and then calculated. For each recording of the schizophrenics and the control subjects, the total, mean, and STD were calculated for the recording time and for the length of pauses. After the description of the experimental setup, we can now explore the results.

3. Results and Discussion

Tables 1 and 2 summarize the mean values and SD for the calculated acoustic voice measures for schizophrenia patients (SCH) and healthy control (HCG) groups, respectively: (1) fundamental frequency (F0); (2) jitter percent; (3) shimmer percent; (4) harmonics-to-noise ratio (HNR).

The mean value of every voice measure was calculated for all the mean and STD values of every Greek vowel. In Table 3, we can observe a summary of the obtained average values for these measures, regardless of the gender and the ages of the subjects.

From the results, we can see that there are significant differences between the measures obtained from schizophrenics and the control subjects. The study showed that the range for mean F0 in schizophrenics was around 178.1 Hz, slightly higher than the mean F0 in control subjects of 162.8 Hz. The distribution of the fundamental frequency (F0) was found not to be statistically significant and was not a useful measure in the differentiation between

the two groups, considering that no significant difference was found between the speech parameters of the two groups.

Table 1. Acoustic voice measures obtained during sustained vowel phonation from schizophrenia patients (SC). Standard deviation is present inside parentheses.

Vowels	F0 (STD) Hz	Jitter (STD) %	Shimmer (STD) %	HNR (STD) dB
a	164.7 (36.7)	2.3 (0.9)	15.9 (4.2)	7.2 (3.4)
e	181.5 (53.0)	2.7 (1.4)	16.1 (5.1)	9.1 (2.9)
i	184.2 (50.0)	2.4 (1.0)	14.0 (5.0)	13.0 (2.2)
o	180.3 (38.1)	2.7 (1.5)	15.1 (4.9)	10.2 (3.6)
u	181.0 (48.0)	2.8 (1.8)	15.3 (4.3)	11.1 (4.0)

Table 2. Acoustic voice measures obtained during sustained vowel phonation from the control group (CG). Standard deviation is present inside parentheses.

Vowels	F0 (STD) Hz	Jitter (STD) %	Shimmer (STD) %	HNR (STD) dB
a	151.6 (47.2)	1.0 (0.4)	12.2 (6.0)	14.9 (4.7)
e	162.3 (49.9)	1.0 (0.4)	6.0 (2.6)	16.1 (4.9)
i	157.7 (53.8)	1.0 (0.3)	5.3 (1.9)	18.2 (3.8)
o	170.4 (55.6)	0.9 (0.5)	5.2 (3.0)	18.6 (4.7)
u	172.2 (69.2)	2.3 (1.6)	6.2 (3.6)	20.3 (6.1)

Table 3. Statistics for acoustic voice measures obtained during sustained vowel phonation (mean values and standard deviations of all vowels combined). SCH: schizophrenia; HCG: control group.

	SCH	HCG	*t*-Test (SCH, HCG)
Parameter	**Mean (STD)**	**Mean (STD)**	***p*-Value**
F0 (Hz)	178.1 (44.4)	162.8 (54.3)	0.4410
Jitter (%)	2.5 (1.3)	1.2 (0.9)	0.0049
Shimmer (%)	15.3 (4.5)	6.9 (4.4)	0.0028
HNR (dB)	10.0 (3.7)	17.6 (5.1)	0.0023

However, we see significant differences in jitter and shimmer values between healthy and schizophrenia subjects. Schizophrenics exhibited a jitter of 2.5% as well as a shimmer at 15.3%, while controls generated a lower jitter at 1.2% and a shimmer of 6.9%. Our results for shimmer and jitter mean values appear to be consistent with the average values of other studies such as the study conducted by [27]. The way jitter was calculated puts a focus on the physiological ability to maintain a constant period and suggests a deficiency in this ability in schizophrenia subjects. Also, the shimmer results may suggest some problems in the spontaneous control of the glottal production mechanism.

The HNR rates also exhibit a significant decline between the two groups, with the patients' recordings having a mean value of 10 ± 0.5 dB and the control subjects' recordings a value of 17.6 dB. Moreover, the mean HNR value cannot be considered a reliable component of speech that could serve as a basis for objective analysis in schizophrenia, owing to the fact that the values for the control subjects in this research are not equivalent to the standard values that other researchers suggest as the average HNR values for healthy voices. Factors such as speaking distance or room acoustics may be the cause of these differences. Also, intensity differences among phonemes can influence the results given that the loudness of the subject's voice, adjusted for a conversation with a nearby listener in specific room conditions, depends on their feelings.

For the speech formant frequencies analysis, we started by representing the F1 and F2 values in two scatter plots, as shown in Figure 1. We can observe the F1 × F2 frequency

space, showcasing the values for all five Greek vowels, extracted from recordings from both the individuals diagnosed with schizophrenia and the healthy control subjects. An ellipse, representing a confidence interval of two standard deviations for each vowel-frequency cluster, is also represented for easy comparison of frequency distributions between the two groups being analyzed.
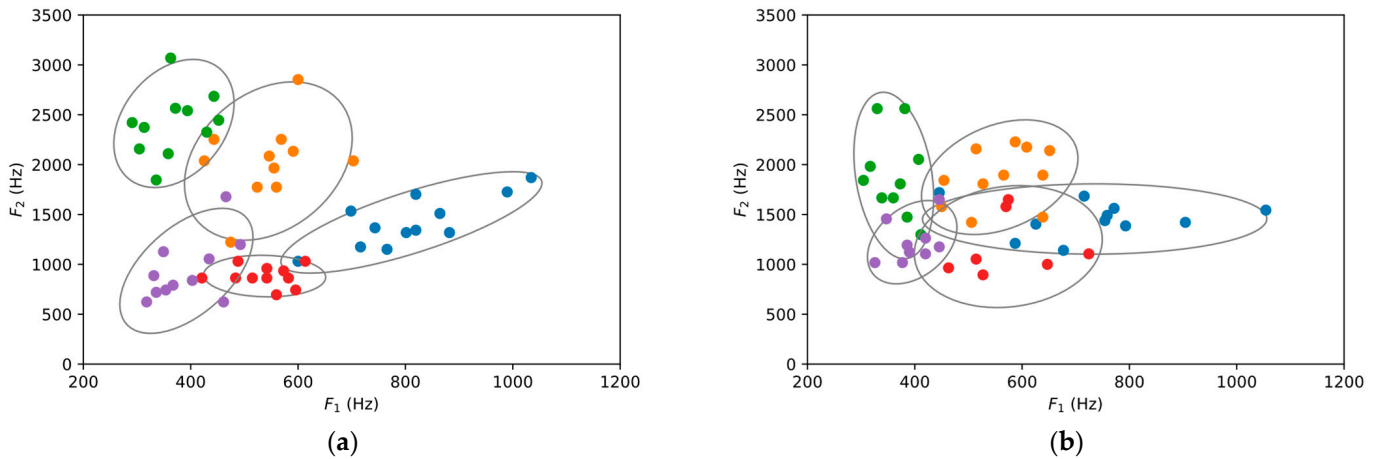

(**a**)

(**b**)

Figure 1. Phoneme formant frequencies for Greek vowels during spontaneous speech by (**a**) healthy control group (HCG) and (**b**) schizophrenia patients (SCH). A two standard deviation confidence ellipse is represented for each vowel group. Vowel identity is color-coded as /a/—blue, /e/—orange, /i/—green, /o/—red, and /u/—purple.

Frequency regions for vowel groups show some internal variability with overlap-ping zones, but in Figure 1b, which concerns the patients, these phenomena are considerably more intense. There is an extensive overlap in the center of the vowel space and a general merging of vowel formant distributions, which results in less differentiated vowels. While for the control subjects, there is clear distinction among the phones /i/, /a/, and /u/, variability is evident both in F1 and F2 resulting in a considerable spread of the frequency values for all vowel categories and considerable overlap especially between the phone /u/ and the phones /e/ and /o/ at the center of the frequency space.

In Figure 2 we can observe box plots for the formant frequencies for each phone group associated to the vowels, with the results of a paired *t*-test on top.
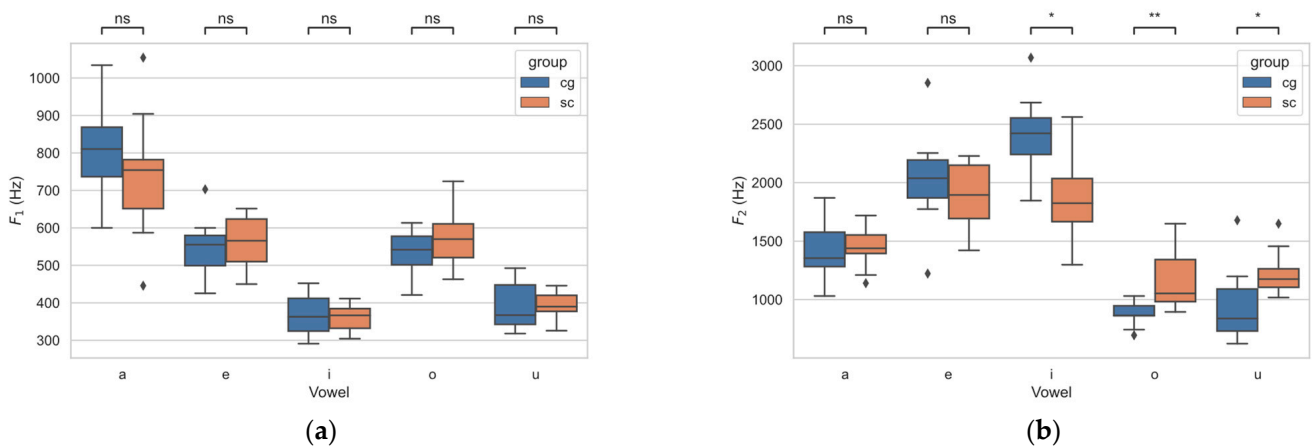

(**a**)

(**b**)

Figure 2. Frequency variability for each Greek vowel for (**a**) formant 1 and (**b**) second formant; *t*-test statistics are represented on top (ns: not significant; *: $0.01 < p \leq 0.05$; **: $1.00 \times 10^{-3} < p \leq 0.01$).

Moreover, Figure 2 shows the mean F1 and F2 formant frequencies produced by the two groups for a better and more obvious comparison of the difference in the mean values.

The overall impression from all the graphs is that schizophrenics show a tendency for a more reduced vowel space with extra overlapping vowels, a characteristic that is not present in healthy controls.

To better observe the vowel space, we have represented cluster centers (not the exact ellipsis centers) in another representation in Figure 3, where we can better observe the differences in the vowel-frequency space between the two groups.
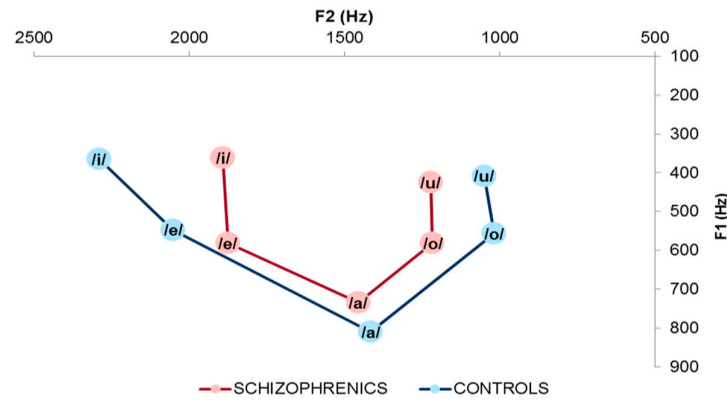

Figure 3. Mean F1, F2 formant frequencies for all five Greek vowels.

For the purpose of comparing groups based on time-related measurements, we utilized the recordings from Task 2, selected due to their shared predefined script. In Figure 4a, a noticeable distinction emerges in speaking time between schizophrenia patients and the control group (verified by a *t*-test yielding a *p*-value of <0.01). Regarding the cumulative duration of pauses, in Figure 4b, a discernible increase is evident within the patient group (with a mean increase of 4.5 s). This disparity is highly significant in comparison to the control group (validated by a *t*-test resulting in a *p*-value of $<1 \times 10^{-5}$). On average, individuals diagnosed with schizophrenia tend to incorporate a greater number of pauses into their speech patterns, contributing to an elevated total pause duration relative to the control subjects. Notably, the pauses observed in healthy participants exhibit a more consistent duration, as indicated by the narrower standard deviation. This phenomenon is indicative of the structured nature of their speech/dialogue interactions.
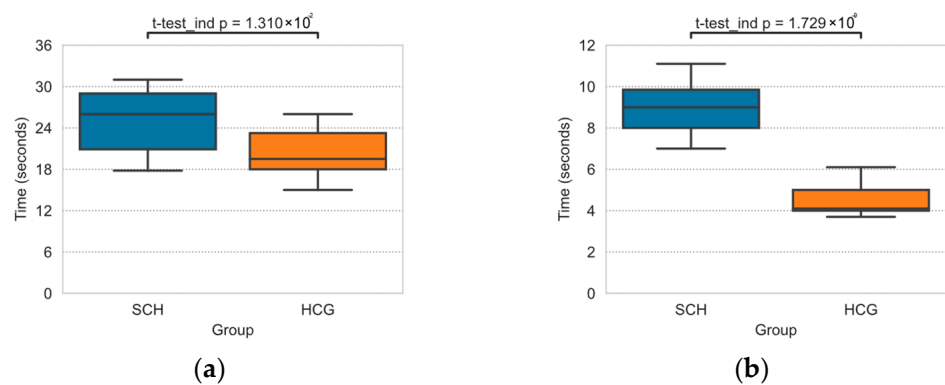

Figure 4. Most relevant differences between schizophrenia patients (SCH) and health control group (HCG) were observed for (**a**) recording time (for the word reading task) and for (**b**) length of pauses (for the spontaneous speech task).

4. Conclusions

This study investigated the differences in acoustic and temporal features of speech that have been found to be potentially linked to the cognitive impairment experienced by individuals with schizophrenia. Cognitive impairment, often accompanied by psychomotor retardation, significantly impacts speech production. Previous research has explored the

feasibility of speech as a biomarker for managing schizophrenia, with a primary focus on the English language. Our study specifically examined the speech characteristics of individuals with schizophrenia in the context of the Greek language. The results obtained from our analysis revealed statistically significant differences between the control group and the group of patients. Notably, variables such as the total length of pauses in spontaneous speech and F2 frequency, and measures like shimmer and jitter, emerged as the most significant factors for distinguishing between the two groups. These findings indicate that the acoustic and temporal analysis of speech holds promise as a potential tool for objectively analyzing schizophrenia. Leveraging speech as an assessment tool boasts several advantages, including its accessibility, the minimal prerequisites, and patient comfort. Moreover, the feasibility of conducting repetitive speech evaluations adds to its practicality. Beyond that, the inclusion of speech analysis as an integral component of diagnostic and therapeutic approaches holds substantial promise, augmenting disease management strategies. Moving forward, further investigation into speech analysis in schizophrenia, encompassing larger sample sizes and diverse linguistic contexts, will contribute to a more comprehensive understanding of its potential as an objective assessment tool. Such advancements can aid in refining diagnostic procedures and facilitating tailored therapeutic interventions for individuals with schizophrenia.

Author Contributions: Conceptualization, L.P.-C. and D.G.; methodology, L.P.-C. and N.D.; software, A.M.; validation, A.M., A.D. and P.K.; formal analysis, L.P.-C.; investigation, A.M.; resources, L.P.-C.; data curation, D.G.; writing—original draft preparation, A.M.; writing—review and editing, L.P.-C. and D.G.; visualization, L.P.-C.; supervision, L.P.-C. and D.G.; project administration, D.G.; funding acquisition, L.P.-C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data can be provided for research purposes by contacting the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Joshy, A.A.; Rajan, R. Automated Dysarthria Severity Classification Using Deep Learning Frameworks. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–21 January 2021; pp. 116–120.
2. Braga, D.; Madureira, A.M.; Coelho, L.; Abraham, A. Neurodegenerative Diseases Detection Through Voice Analysis. In *Hybrid Intelligent Systems*; Abraham, A., Muhuri, P.K., Muda, A.K., Gandhi, N., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 734, pp. 213–223.
3. Vigo, I.; Coelho, L.; Reis, S. Speech- and Language-Based Classification of Alzheimer's Disease: A Systematic Review. *Bioengineering* **2022**, *9*, 27. [CrossRef] [PubMed]
4. Vieira, H.; Costa, N.; Sousa, T.; Reis, S.; Coelho, L. Voice-Based Classification of Amyotrophic Lateral Sclerosis: Where Are We and Where Are We Going? A Systematic Review. *NDD* **2019**, *19*, 163–170. [CrossRef] [PubMed]
5. Yin, F.; Du, J.; Xu, X.; Zhao, L. Depression Detection in Speech Using Transformer and Parallel Convolutional Neural Networks. *Electronics* **2023**, *12*, 328. [CrossRef]
6. Arieti, S. *Interpretation of Schizophrenia*; Robert Brunner: New York, NY, USA, 1955.
7. Ueda, N.; Maruo, K.; Sumiyoshi, T. Positive symptoms and time perception in schizophrenia: A meta-analysis. *Schizophr. Res. Cogn.* **2018**, *13*, 3–6. [CrossRef] [PubMed]
8. Cohen, A.S.; Mitchell, K.R.; Elvevåg, B. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. *Schizophr. Res.* **2014**, *159*, 533–538. [CrossRef] [PubMed]
9. Buchanan, R.W. Persistent Negative Symptoms in Schizophrenia: An Overview. *Schizophr. Bull.* **2007**, *33*, 1013–1022. [CrossRef] [PubMed]
10. Murphy, B.P.; Chung, Y.-C.; Park, T.-W.; McGorry, P.D. Pharmacological treatment of primary negative symptoms in schizophrenia: A systematic review. *Schizophr. Res.* **2006**, *88*, 5–25. [CrossRef] [PubMed]
11. Alpert, M.; Kotsaftis, A.; Pouget, E.R. At issue: Speech fluency and schizophrenic negative signs. *Schizophr. Bull.* **1997**, *23*, 171–177. [CrossRef] [PubMed]

12. Covington, M.A.; Lunden, S.L.A.; Cristofaro, S.L.; Wan, C.R.; Bailey, C.T.; Broussard, B.; Fogarty, R.; Johnson, S.; Zhang, S.; Compton, M.T. Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. *Schizophr. Res.* **2012**, *142*, 93–95. [CrossRef]
13. Deliyski, D.D.; Shaw, H.S.; Evans, M.K.; Vesselinov, R. Regression tree approach to studying factors influencing acoustic voice analysis. *Folia. Phoniatr. Logop.* **2006**, *58*, 274–288. [CrossRef] [PubMed]
14. Benward, B.; Saker, M. *Music in Theory and Practice Volume 1*, 9th ed.; McGraw-Hill Education: Dubuque, IA, USA, 2014; ISBN 978-0-07-802515-0.
15. Teixeira, J.P.; Fernandes, P.O. Acoustic Analysis of Vocal Dysphonia. *Procedia Comput. Sci.* **2015**, *64*, 466–473. [CrossRef]
16. Yumoto, E.; Sasaki, Y.; Okamura, H. Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J. Speech Hear. Res.* **1984**, *27*, 2–6. [CrossRef]
17. Murry, T.; Brown, W.S.; Rothman, H. Judgments of voice quality and preference: Acoustic interpretations. *J. Voice* **1987**, *1*, 252–257. [CrossRef]
18. Coelho, L.; Braga, D.; Dias, M.; Garcia-Mateo, C. An Automatic Voice Pleasantness Classification System Based on Prosodic and Acoustic Patterns of Voice Preference. In Proceedings of the 12th Annual Conference of the International Speech Communication Association, Florence, Italy, 27–31 August 2011.
19. Ferrand, C.T. Harmonics-to-Noise Ratio: An Index of Vocal Aging. *J. Voice* **2002**, *16*, 480–487. [CrossRef] [PubMed]
20. Covington, M.A.; He, C.; Brown, C.; Naçi, L.; McClain, J.T.; Fjordbak, B.S.; Semple, J.; Brown, J. Schizophrenia and the structure of language: The linguist's view. *Schizophr. Res.* **2005**, *77*, 85–98. [CrossRef] [PubMed]
21. Compton, M.T.; Lunden, A.; Cleary, S.D.; Pauselli, L.; Alolayan, Y.; Halpern, B.; Broussard, B.; Crisafio, A.; Capulong, L.; Balducci, P.M.; et al. The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. *Schizophr. Res.* **2018**, *197*, 392–399. [CrossRef]
22. Çokal, D.; Zimmerer, V.; Turkington, D.; Ferrier, N.; Varley, R.; Watson, S.; Hinzen, W. Disturbing the rhythm of thought: Speech pausing patterns in schizophrenia, with and without formal thought disorder. *PLoS ONE* **2019**, *14*, e0217404. [CrossRef] [PubMed]
23. Fraser, W.I.; King, K.M.; Thomas, P.; Kendell, R.E. The diagnosis of schizophrenia by language analysis. *Br. J. Psychiatry* **1986**, *148*, 275–278. [CrossRef] [PubMed]
24. Rapcan, V.; D'Arcy, S.; Yeap, S.; Afzal, N.; Thakore, J.; Reilly, R.B. Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Med. Eng. Phys.* **2010**, *32*, 1074–1079. [CrossRef] [PubMed]
25. Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. *Ear Hearing* **2011**, *32*, 266. [CrossRef]
26. Android Application "Parrot—Voice Recorder". Available online: https://play.google.com/store/apps/details?id=com.SearingMedia.Parrot (accessed on 9 June 2023).
27. Kliper, R.; Portuguese, S.; Weinshall, D. Prosodic Analysis of Speech and the Underlying Mental State. In *Pervasive Computing Paradigms for Mental Health*; Serino, S., Matic, A., Giakoumis, D., Lopez, G., Cipresso, P., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 52–62.