Analysis of Prosodic Features During Cognitive Load in Patients with Depression

Carmen Martínez, Spyridon Kontaxis, Mar Posadas-de Miguel, Esther García, Sara Siddi, Jordi Aguiló, Josep Maria Haro, Concepción de la Cámara, Raquel Bailón, and Alfonso Ortega

Abstract Major Depressive Disorder (MDD) is a largely extended mental health disorder commonly associated with a hesitant and monotonous speech. This study analyses a speech corpus from a database acquired on 40 MDD patients and 40 matched controls (CT). During the recordings, individuals experienced different levels of cognitive stress when performing Stroop color test that includes three tasks with increasingly level of difficulty. Speech features based on the fundamental frequency (F0), and the speech ratio (SR), which measures the speech to silence ratio, are used for characterising depressive mood and stress responsiveness. Results show that SR is significantly lower in MDD subjects compared to healthy controls for all the tasks, decreasing as the difficulty of the cognitive tasks, and thus the stress level, increases. Moreover F0 related parameters (median and interquartile range) show higher values within the same subject in tasks with increased difficulty level for both groups. It can be concluded that speech features could be used for characterising depressive mood and assessing different levels of stress.

C. Martínez (✉)
University of Zaragoza, Campus Río Ebro, C/María de Luna 1, 50018 Zaragoza, Spain
e-mail: martineza@unizar.es

S. Kontaxis · R. Bailón · A. Ortega
University of Zaragoza, Campus Río Ebro, C/María de Luna 1, 50018 Zaragoza, Spain
e-mail: sikontax@unizar.es; rbailon@unizar.es; ortega@unizar.es

M. Posadas-de Miguel · C. de la Cámara
Hospital Clínico de Zaragoza, Zaragoza, Spain
e-mail: marposadasdemi@yahoo.es; conchidlc@hotmail.com

E. García · J. Aguiló
Autonomous University of Barcelona, Barcelona, Spain
e-mail: esther.garcia@ciber-bbn.es; jordi.aguilo@uab.es

S. Siddi · J. M. Haro
Parc Sanitari Sant Joan de Déu, Barcelona, Spain
e-mail: sara.siddi@pssjd.org; jmharo@pssjd.org

193

1 Introduction

Major Depressive Disorder (MDD) is the most prevalent mental illness and the fifth leading cause of global disability in the world [1]. Depression is associated with loss of interest, tiredness, and sleep disruptions [2]. Its diagnosis is currently based on clinical interviews and questionnaires according to Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [3], such as Hamilton Depression Rating Scale (HDRS) or Beck Depression Inventory (BDI). However, these methods may be dependent on the predisposition of the patients to talk about their symptoms.

MDD has been linked to an autonomic nervous system (ANS) dysfunction, which in turn leads to changes in psychomotor activity [4]. As a result, speech in MDD patients has been described as slower, monotonous, paused or even hesitant [5], which can guide to the idea of using speech analysis to characterise depressive mood in a non-invasive and comfortable way.

In fact, speech signals have already been used for diagnosing depression, but different results were obtained [6]. In [7, 8], a significant negative trend between fundamental frequency (F0) and depression severity (measured through HDRS or BDI) was reported. However in [9–12], F0 yielded to no significant results for discriminating depression state. Besides frequency parameters, speech ratio-related features have also been studied in depressed speech. In [8], higher depression severity scores were associated with slower speech, while in [10, 13], longer pauses were found in depression patients.

In the present work, a set of prosodic features are used for discriminating between MDD patients and control (CT) subjects. These parameters will be analysed while subjects perform a Stroop test [14], which implies an homogeneous condition for speech but a cognitive load and stressful situation for ANS that can highlight its dysfunction in MDD.

2 Materials and Methods

2.1 Experimental Protocol

A database of 40 MDD patients (24 women, age 47.33 ± 13.07 years, Body Mass Index (BMI) 27.85 ± 5.63 Kg/m^2), was recorded at the Hospital Clínico Lozano Blesa (Zaragoza) and the Parc Sanitari Sant Joan de Déu network of mental health services (Barcelona). MDD group consists of subjects with Major Depression Disorder according with Diagnostic and Statistical Manual of Mental Disorders (DSM-5) criteria [3] and the HDRS scale (HDRS > 9). Recordings of 40 CT subjects without clinical history of mental disorders, matched by sex, BMI and age, were also acquired in order to ensure that group differences are attributable to differences in depression status and not due to unbalanced demographic data.

The experimental protocol consists in an exposure to a cognitive stress named Stroop test [14]. The test was adapted to the mother tongue of the subjects, i.e.

Spanish, and it is formed of three parts lasting about 45 sec each. In the first task (T1) the subjects should read the words 'red', 'green', and 'blue' printed in black ink. In the second one (T2), the words are substituted by a set of 'XXX's written in the previous colours, and the participants should say out loud the colour of the ink. Finally, in the third test (T3), the words 'red', 'green', and 'blue' are coloured in inks that do not match the true meaning of the text, so the subject should say the colour of the ink and not the written word. Note that the difficulty of the tasks increases gradually from T1 to T3, being the colours and the words in the first level (T1) congruent, and incongruent in the second and third (T2, T3).

Speech signals were acquired using an AKG CK-80 microphone and a Tascam us-122L recording device at a sampling frequency of 44.1 kHz.

2.2 Speech Processing

At the beginning, speech signals were preprocessed, using Audacity® and FFMPEG. Noisy segments, e.g. coughs or throat cleanings, as well as parts of the recording in which the interviewer interfered with the subject, e.g. correcting errors or providing details related to task, were removed.

Prosodic parameters are obtained in order to analyse differences between depressive and control speech. First, F0 is estimated each 10 msec using the robust algorithm for pitch tracking (RAPT) [15] included in the openSMILE toolkit [16]. Then, median and interquartile range of F0 ($F0_m$ and $F0_{iqr}$) during each task are computed as prosodic features. However, these values can be highly variant among different subjects and thus, the normalised differences $\widetilde{F}0_m$, $\widetilde{F}0_{iqr}$ for T2 and T3 with respect to T1 will be studied for measuring response to increasing stress level in both populations.

Moreover, speech ratio (SR), which measures the proportion between the time that the subject is speaking and the total duration of the recording, is computed using a voice activity detector algorithm (VAD) [17] based on Long-Term Spectral Divergence (LTSD).

2.3 Statistical Analysis

Feature set consists of previously mentioned parameters from speech analysis i.e., $\widetilde{F}0_m$, $\widetilde{F}0_{iqr}$, and SR. Unpaired tests between MDD and CT populations are carried out to study the effect of depression on speech, while paired tests are conducted to study the differences within the same subject on stress response at tasks of different difficulty. Moreover, prosodic indexes are analysed taking into account the differences in the frequency ranges of each gender [18], generating four study groups. Student t-tests or Wilcoxon tests are implemented depending on the distribution of the data (Shapiro-Wilk test), Gaussian or not, for comparing means or medians of the distribution, respectively. In this study, the level of statistical significance is set to $p = 0.05$.

3 Results

Results show that $\widetilde{F}0_m$ (Fig. 1(a–b)) and $\widetilde{F}0_{iqr}$ (Fig. 1(c–d)) do not differ significantly between CT and MDD population (unpaired tests) for both genders. Regarding paired analysis, $\widetilde{F}0_m$ and $\widetilde{F}0_{iqr}$ show increasing trends in almost every group due to the increment of difficulty at T3. Only $\widetilde{F}0_m$ in male subjects with MDD do not exhibit significant differences as the cognitive stress raises.

Moreover, SR values, shown in Fig. 2, were found to be significantly lower in every Stroop task in MDD with respect to CT groups. Note that, in both groups, SR values decrease significantly as the difficulty of the cognitive tasks, and thus the stress level, increases, i.e., from T1 to T3.

4 Discussion

In the present work, an analysis in depressive subjects and paired controls has been conducted by means of speech features for studying differences in stress responsiveness. Stroop Test has been previously used in healthy subjects for measuring cognitive load, i.e., mentally stressful situations, from spoken speech [19–21], showing that vocal frequencies can correlate with cognitive load and be used for its classification.

Results in Fig. 1 show that MDD and healthy subjects exhibit higher values in frequency-related features during T3, thereby suggesting that both groups reacted to stress induction. The values of $\widetilde{F}0_m$ and $\widetilde{F}0_{iqr}$ in females (Fig. 1 (a, c)) increase significantly in both groups as the difficulty of the tasks increases, while in males
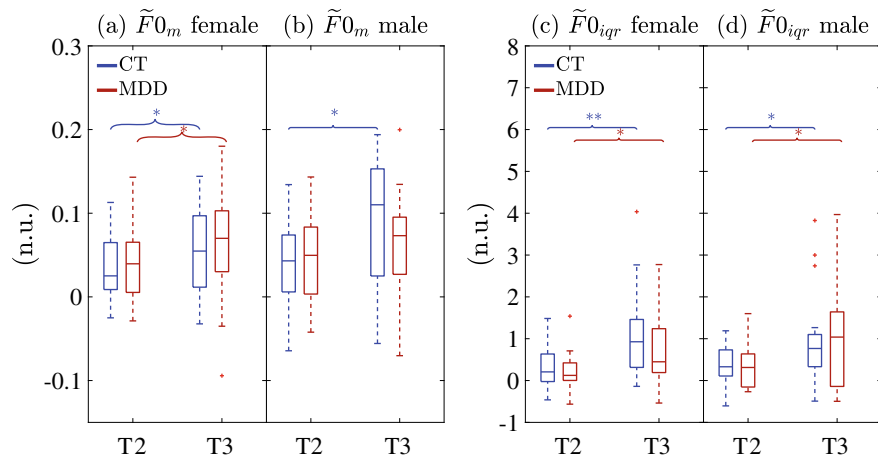

Fig. 1 Boxplot of prosodic features. (a) $\widetilde{F}0_m$ in female subjects, (b) $\widetilde{F}0_m$ in male subjects, (c) $\widetilde{F}0_{iqr}$ in female subjects, (d) $\widetilde{F}0_{iqr}$ in male subjects. CT and MDD groups are marked in blue and red, respectively, while statistically significant differences of paired analysis are marked with one or two colored asterisks for $p < 0.05$ and $p < 0.001$, respectively.
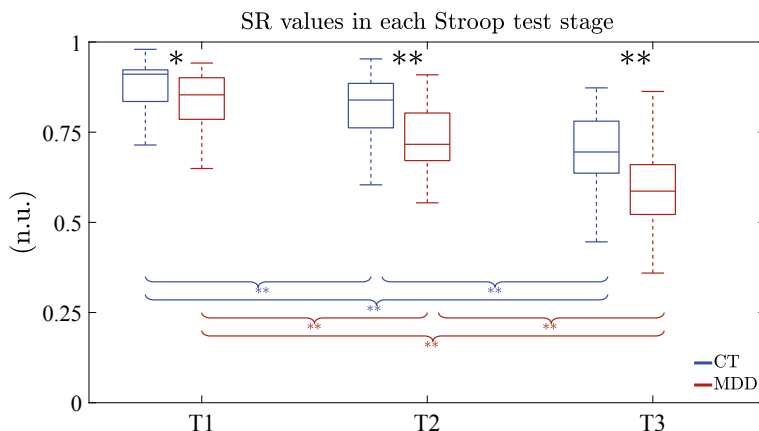
Fig. 2 Boxplot of Speech Ratio (SR). CT and MDD populations are marked in blue and red, respectively. Statistically significant differences for $p < 0.05$ and $p < 0.001$ are marked with one and two asterisks, respectively. Black asterisk correspond to unpaired analysis and colored to paired analysis.

(Fig. 1 (b, d)) only controls show higher values. The absence of significant differences in male subjects with MDD could be either because they do not respond to stressful stimuli, or due to the small number of male subjects (16 out of 40). In [7, 8], a significant negative trend for F0-related features was reported. However, most studies analysed these parameters using non-spontaneous speech, e.g. oral reading [6, 22], instead of performing a cognitive test. Thus, the absence of significant differences in F0-related parameters between MDD and CT group in this study, might be attributed to the stress induced by Stroop Test.

Moreover, SR is the most promising parameter as it shows significant differences between MDD and CT group in all tasks. Results show a decreasing trend with the increment of stress level in both groups. Note that SR speech parameter values are always lower in MDD group, which might be related to cognitive dysfunction in MDD patients [23]. The ability to inhibit cognitive interference in the Stroop Test has been used for measuring cognitive functions such as attention and processing speed among others [24]. Videbech et al. [25], reported that patients with depression had a greater difficulty when inhibiting interference compared to CT subjects, thus leading to lower values of SR in a fixed amount of time. A lower performance of MDD compared to CT subjects, measured as the time required for accomplishing a cognitive task, was reported in [26] for a subset of the present database.

5   Future Work

This study consist in a preliminary work that highlights the importance of the stress response for the monitoring and diagnosis of ambulatory MDD patients. In fact,

using speech analysis, depressive mood could be assessed in a non-invasive and comfortable way. Moreover, a joint analysis of additional parameters, such as jitter or shimmer, with other physiological signals, such as HRV features, can be conducted using different classifiers.

6 Conclusion

A preliminar study about speech analysis during a cognitive stress in a database recorded on MDD patients and matched controls has been presented. The analysis of speech parameters during Stroop Test tasks has revealed significantly decreased speech ratio values in MDD patients with respect to matched controls and differences in fundamental frequency related parameters within the same subject among different tasks. Thus, it can be concluded that the analysis of speech can be used for the objective diagnosis of MDD patients in a non-invasive, straightforward, and comfortable way. In conclusion, SR, not only because of its simplicity but also due to its robustness to inter-session variations due to recording conditions, can be thought as a suitable feature for distinguishing between populations in non-controlled environments, as it has shown the best performance among the analysed features.

References

1. Vos T (2017) Global, regional, and national incidence, prevalence, and years lived with disability for 328 disease and injuries, 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390:1211–1259
2. World Health Organisation (WHO) (2011) Depression: let's talk. In: Website of World Health Association. Disorders Management, Depression. https://www.who.int/news-room/detail/30-03-2017--depression-let-s-talk-says-who-as-depression-tops-list-of-causes-of-ill-health. Accessed Jan 2020
3. American Psychiatric Association (1994) Diagnosis and Statistical Manual of Mental Disorders (DSM). 4th edn. Washington DC
4. Sperry SH, Kwapil TR, Eddington KM et al (2018) Psychopathology, everyday behaviours, and autonomic activity in daily life: An ambulatory impedance cardiography study of depression, anxiety, and hypomaniac traits. Int J Psychophysiol 129:67–75
5. Kräpelin E (1921) Manic-depressive insanity and paranoia, 2nd edn. Livingstone, Edinburgh
6. Cummins N, Scherer S, Krajewski J et al (2015) A review of depression and suicide risk assessment using speech analysis. Speech Commun 71:10–49

7. Hönig F et al (2014) Automatic modelling of depressed speech: relevant features and relevance of gender. In: 15th Proceedings of Interspeech, Singapore, 14–18 September 2014
8. Cannizzaro M, Harel B, Reilly N et al (2004) Automatic modelling of depressed speech: voice acoustical measurement of the severity of major depression. Brain Cogn 56:30–35
9. France DJ, Shiavi RG, Silverman S et al (2000) Acoustical properties of speech as indicator of depression and suicidal risk. IEEE T Bio Med Eng 47:309–319
10. Mundt JC, Snyder PJ, Cannizzaro MS et al (2007) Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguist 20:50–64
11. Taguchi T, Tachikawa H, Nemoto K, Suzuki M et al (2017) Major depressive disorder discrimination using vocal acoustic features. J Affect Disorders 225:214–220
12. Quatieri TF et al (2012) Vocal-source biomarkers for depression: a link to psychomotor activity. In: 13th Proceedings of Interspeech, Portland, OR, USA, 9–13 September 2012
13. Mundt JC, Vogel AP, Feltner DE et al (2012) Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry 72:580–587
14. Stroop JR (1992) Studies of interference in serial verbal reactions. J Exp Psychol 121:15–23
15. Resch B, Nilsson M, Ekman A et al (2007) Estimation of the Instantaneous Pitch of Speech. IEEE T Audio Speech 15:813–822
16. Eyben F, Wöllmer M, Schuller B (2010) openSMILE - the munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM international conference on multimedia, Firenze, Italy, 25–29 October 2010
17. Ramírez J, Górriz JM, Segura JC (2007) Voice activity detection. Fundamentals and speech recognition system robustness. In: Grimm M, Kroschel K (eds) Robust speech recognition and understanding. InTech
18. Klatt DH, Klatt LC (1990) Analysis, synthesis and perception of voice quality variations among female and male talkers. J Acoust Soc Am 87:820–857
19. Schuller B et al (2014) The INTERSPEECH 2014 computational paralinguistics challenge: cognitive & physical load. In: 15th Proceedings of Interspeech, Singapore, 14–18 September 2014
20. Yin B et al (2008) Speech-based cognitive load monitoring system. In: 2008 IEEE international conference on acoustics, speech, and signal processing, Las Vegas, NV, USA, 31 March–4 April 2008
21. Yap TF, Epps J, Ambikairajah E et al (2001) Formant frequencies under cognitive load: effects and classification. EURASIP J Adv Sig Pr
22. Williamson JR et al (2014) Vocal and facial biomarkers of depression based on motor incoordination and timing. In: AVEC 2014 Proceedings of the 4th international workshop on audio/visual emotion challenge, Orlando, Florida, USA, November 2014
23. Lam RW, Kennedy SH, McIntyre RS et al (2014) Cognitive dysfunction in major depressive disorder: effects on psychosocial functioning and implications for treatment. Can J Psychiatry 59:614–654
24. Scarpina F, Tagini S (2017) The stroop color and word test. Front Psychol 8:557
25. Videbech P, Ravnkilde B, Gammelgaard L et al (2014) The danish PET/depression project: performance on Stroop's test linked to white matter lesions in the brain. Psychiatry Res 130:117–130
26. Kontaxis S, Orini M, Gil E, Posadas-de Miguel M, Bernal ML, Aguiló J, de la Cámara C, Laguna P, Bailón R (2018) Heart rate variability analysis guided by respiration in major depressive disorder. In: 45th International conference of computing in cardiology, Maastricht, The Netherlands, 23–26 September 2018