Psihološka istraživanja, Vol. XXV (2) 2022.
183–203

UDK 616.22-008.5:616.89-008.454
TIP ČLANKA: Originalni naučni rad
DOI: 10.5937/PSISTRA25-39224

Acoustic features of voice in adults suffering from depression[1]

Gordana Calić[2]
Faculty of Special Education and Rehabilitation, University of Belgrade, Serbia

Mirjana Petrović-Lazić
Faculty of Special Education and Rehabilitation, University of Belgrade, Serbia

Tatjana Mentus
Faculty of Special Education and Rehabilitation, University of Belgrade, Serbia

Snežana Babac
Faculty of Special Education and Rehabilitation, University of Belgrade, Serbia

In order to examine the differences in people suffering from depression (EG, N=18) compared to the healthy controls (CG1, N=24) and people with the diagnosed psychogenic voice disorder (CG2, N=9), nine acoustic features of voice were assessed among the total of 51 participants using the MDVP software programme ("Kay Elemetrics" Corp., model 4300). Nine acoustic parameters were analysed on the basis of the sustained phonation of the vowel /a/. The results revealed that the mean values of all acoustic parameters differed in the EG compared to both the CG1 and CG2 as follows: the parameters which indicate frequency variability (Jitt, PPQ), amplitude variability (Shim, vAm, APQ) and noise and tremor parameters (NHR, VTI) were higher; only the parameters of fundamental frequency (F0) and soft index phonation (SPI) were lower (F0 compared to CG1, and SPI compared to CG1 and CG2). Only the PPQ parameter was not significant. vAm and APQ had the highest discriminant value for depression. The acoustic features of voice, analysed in this study with regard to the sustained phonation of a vowel, were different and discriminant in the EG compared to CG1 and CG2. In voice analysis, the

[1] This paper was a result of research within the projects of the Ministry of Education, Science and Technological Development of the Republic of Serbia (no. 451-03-68/2022-14).

[2] calicgordana@yahoo.com

parameters vAm and APQ could potentially be the markers indicative of depression. The results of this research point to the importance of the voice, that is, its acoustic indicators, in recognizing depression. Important parameters that could help create a programme for the automatic recognition of depression are those from the domain of voice intensity variation.

Keywords: acoustic features, vowel, depression, voice disorder, voice analysis

Introduction

As a complex psychophysical disorder affecting every aspect of life (affective, cognitive, motivational, physical, and social), depression is usually defined as a long-lasting low mood accompanied by a loss of interest in the activities a person used to enjoy, combined with an inability to perform daily activities for at least two weeks (WHO, 2017). Due to its high prevalence in the general population, ranging from 2% to 6% in the literature, with a proportion of 3.44% according to some reports (Ritchie & Roser, 2018), 3.8% (IHME, 2019) and 4.4% according to WHO (2017), which can further lead to suicide (accounting for 1.5% of the total mortality rate in the world according to WHO, 2017), scientists are trying to find the causing factors of depression in order to prevent it. A meta-analytical study (Bueno-Notivol et al., 2021), conducted during the COVID-19 pandemic, determined a depression rate between 7.45% and 48.30%, indicating almost seven times higher prevalence of depression than before the pandemic. All this points out to the significance of early detection of depression. Furthermore, in recent years, growing scientific interest has been focused on determining the ways to detect mental disorders through speech signals so that voice could be used as an objective biomarker in detecting these disorders instead of relying solely on the patients' subjective self-assessment and clinicians' experience. Thus, examining the acoustic features of voice may contribute to creating programmes for the automatic recognition of depression and other mental disorders.

What is the relation between voice and emotions?

Speech communication consists of direct and indirect channels. The direct (verbal) channel includes linguistic content (what is said), while the indirect (nonverbal) one refers to the paralinguistic content related to the speaker (how something is said) (Yang & Lugger, 2010). Emotions, quality of voice, and accentuation are just some of the paralinguistic features (Yang & Lugger, 2010). It is well known that emotions can be vocally expressed. According to Scherer's theory, physiological factors largely determine the nature of phonation and resonance in vocal expression (Scherer, 1986). Specific acoustic voice features may be expected with regard to the physiological condition. The three most common acoustic indicators are fundamental frequency (F0), vocal intensity, and speaking rate (Yang & Lugger, 2010).

Emotions can largely affect voice and phonation. Numerous studies have shown that acoustic features of voice are indicators of different emotions (Juslin & Laukka, 2003; Scherer, 2003; Scherer, Clark-Polner & Mortillaro, 2011; Patel & Scherer, 2013). This influence is related to the subcortical parts of the central nervous system (CNS) responsible for expressing emotions, while some also regulate the activity of the neurovegetative system. Due to this close relation, emotions directly influence the activity and tone of voice organs (Milutinović, 1997). It is possible to distinguish between the positive and negative effects of emotions on voice. The positive impact is manifested in the increased tone and activation of the CNS, which improves the coordination of phonation organs. The negative impact of low mood is reflected in the inhibiting effect on phonation organs (Milutinović, 1997).

Depression and acoustic features of voice

The assessment of depression is based on the patients' subjective self-reporting and clinicians' experience (Alghowinem et al., 2013; Mundt et al., 2007). An experienced clinician can subjectively perceive the vocal changes typical of depression, but the practice is not based on the assessment of objective voice parameters. However, with regard to the close relation between voice and emotions, scientists are becoming increasingly interested in determining the ways to detect mental disorders through speech signals, i.e., creating an algorithm for detecting depression with great precision (Alghowinem et al., 2013; Afshan et al., 2018; Cummins et al., 2015; Cummins et al., 2011; He & Cao, 2018; Jiang et al., 2017; Kiss & Jenei, 2020; Lopez-Otero & Docio-Fernandez, 2020; Nunes et al., 2010; Rejaibi et al., 2022; Sturim et al., 2011; Xing et al., 2022). Such studies on depression are suitable since they are not intrusive, do not require direct contact with participants, and are not expensive (Cummins et al., 2011).

Acoustic analysis can provide objective data in addition to clinicians' subjective assessment. Such analysis is much more precise than the perceptive one since it provides quantitative measures (Petrović-Lazić et al., 2014) and may thus be used in both diagnostic and therapeutic clinical processes. Objective acoustic analysis is increasingly used in the literature on voice analysis in depression. Different speech tasks (sound phonation, spontaneous speech, reading texts, describing pictures) are used in the analysis. However, there is a discrepancy in determining the most suitable speech task for detecting depression. For example, according to some findings, acoustic features analysed on the basis of text reading have a high predictive value for depression measured by the Hamilton scale (Hashim et al., 2017) and, in others, those are acoustic features analysed on the basis of sound phonation for the Beck depression inventory (Silva et al., 2021).

Selecting a limited number of relevant variables for voice analysis and assessment has long been the main problem of speech analysis in psychiatry (Nilsonne, 1988). The existing literature has determined some acoustic variables indicating the difference between the participants with depression and the control group.

The most frequently examined acoustic features in people suffering from depression are the parameters which indicate frequency and its variation (F0, Jitter), amplitude variation (Shimmer) and noise and tremor parameter (NHR). The fundamental frequency of voice (F0) refers to the number of vibration cycles in one second (the frequency at which vocal cords open and close) (Baken & Orlikoff, 2000). The parameter of frequency variation from cycle to cycle (Jitter) measures short-term, cyclic irregularities of a voice period, while the amplitude variation of the sound wave (Shimmer) indicates amplitude variations during vocal cord vibrations (Zwetsch et al., 2006). The Noise-to-harmonics Ratio (NHR) is the ratio between harmonic and noise (non-harmonic) voice components (Ferrand, 2002). Most studies have examined the mean values of F0 and its deviations. Numerous studies indicate a smaller range of F0 variability in participants with depression, which, according to some authors, points to monotonous speech (Ellgring & Scherer, 1996; Moore et al., 2004; Mundt et al., 2007; Nilsonne, 1988; Silva et al., 2021). Several studies show that the mean value of F0 is lower in people with depression than in the typical control group (Mundt et al., 2007; Mundt et al., 2012; Wang et al., 2019), as well as that F0 decreases with the degree of depression severity (Yang et al., 2013). However, there are findings, although scarce, according to which there is no significance (Taguchi et al., 2018). Nilsonne suggests that smaller changes in F0 in the participants with depression compared to the control group could be the basic difference between these two groups of participants (Nilsonne, 1988). Most research studies show that the value of the Jitter parameter is higher in people suffering from depression (Nunes et al., 2010; Sahu & Espy-Wilson, 2016; Silva et al., 2021). The same is true for the Shimmer parameter (Sahu & Espy-Wilson, 2016; Silva et al., 2021), while in some studies, Shimmer was lower when analysing the voice in sadness (Nunes et al., 2010). Ozdas et al. (2004) point out that the increase in Jitter and glottal spectral tilt could be the acoustic features distinguishing between people with depression, suicidal people, and the control group. The NHR parameter is higher in the participants with depression than in the control group in some papers (Low et al., 2011), and lower in others (Quatieri & Malyska, 2012).

By reviewing the research findings, we have noticed that the results related to acoustic (vocal) analysis are not consistent. According to the literature, there is a need to determine additional acoustic features and different classifications in this field (Sahu & Espy-Wilson, 2016). The topic-related literature is quite inconsistent with regard to sample selection, methodology, type, and number of the analysed parameters. To our knowledge, scarce studies in the Serbian-speaking area (Ćuk-Jovanović, 2002; 2003; Popović, 2003) have found differences in specific acoustic voice features in people with depression compared to the control group, primarily in the intensity and duration of speech. Some authors recommend examining whether the acoustic features in depression other than F0 (which has the highest consistency of results) are consistent in different speech and cultural areas (Wang et al., 2019). In addition, the studies have only dealt with the comparison of voice between people with depression and the typical control group, not considering the group of participants with a voice disorder. Stress and depression are known to be some of the causes of hyperfunctional (psychogenic) dysphonia (Kosztyła-Hojna et al., 2018). This type of dysphonia involves excessive muscle tension due to an inadequate phonation process (Teixeira & Fernandes, 2015).

Thus, we wish to determine whether there are any differences in the acoustic features of voice between the people with psychogenic voice disorders and those with depression. Furthermore, we examine the acoustic features of voice in adults suffering from depression compared to the healthy controls in the control group in the Serbian-speaking area. We include the measures of frequency and its perturbations, amplitude, and the parameters of noise and tremor. We also want to determine whether the acoustic features of voice are discriminant for depression.

We hypothesize that the participants with depression differ in their acoustic parameters compared to the healthy controls and participants with psychogenic voice disorders.

Method

The sample

The research included 51 participants (F_{male} = 29.4%), divided into three subgroups: 18 participants with diagnosed depression (experimental group – EG, aged between 27 and 63, AS=51.83; SD=9.357), 24 healthy controls (control group 1 – CG1, aged between 18 and 34, AS=24.25; SD=4.286), and 9 with a diagnosed psychogenic voice disorder (control group 2 – CG2, aged between 34 and 61, AS=46.44; SD=9.671). The age range of all included participants was 18 to 65 years. There were no significant differences with regard to gender (χ^2 = 4.974, p> .05).

Table 1
Distribution of the participants (N=51) with regard to gender

Variable	Category	Frequency	%
Gender	Male	15	29.4
	Female	36	70.6
Group	Experimental group	18	35.3
	Control group 1	24	47.1
	Control group 2	9	17.6

To obtain the consent for conducting this study, we first sent a request to the head of the Psychiatry Clinic at the Zvezdara Medical Centre in Belgrade, thoroughly explaining the research aim, method, and procedure. After obtaining the consent, we sent a written request to the Medical Centre's Ethics Committee. The research was conducted after obtaining the consent of the local Ethics Committee of the Zvezdara Medical Centre (number IRB00009457). A speech and language pathologist who recorded voice, a psychiatrist who applied the MADRS Scale, and an otorhinolaryngologist who assessed the larynx participated in data collection.

After each patient had consented to participate in the study, psychiatric history data were taken from the medical records (which included a psychiatric interview). Only participants who had a major depressive disorder without comorbidity were selected based on archive documents. The data indicating the major depressive disorder were collected by a psychiatrist based on the Montgomery-Asberg Depression Rating Scale (MADRS scale, Montgomery & Asberg, 1979) and used only as a confirmation of the condition of each participant (which was previously taken as archive data, from medical records), e.g. to certify the absence of remission. Participants in the control group had no previous psychiatric history.

Participants with a psychogenic voice disorder were included in the study upon their examinations at the Ear, Throat, and Nose Clinic of Zvezdara, where the authors had access. The study included only those participants who had given their consent and were examined by an otorhinolaryngologist and a vocal pathologist. Healthy control subjects were randomly selected, also upon giving consent and being examined.

The inclusion criteria were the absence of chronic diseases affecting the quality of voice, such as neurological, endocrine, and infectious, the absence of other psychiatric disorders and the absence of the aging voice. Only the participants with normal laryngoscopy findings, without organic voice disorders, were included in the study. Also, the participants who were vocal professionals were excluded from the research.

Materials and apparatus

The Multi-Dimensional Voice Programme (MDVP) of the Computerized speech lab ("Kay Elemetrics" Corp., model 4300) was used in the analysis of acoustic voice features. The programme offers a detailed graphical and numerical display for 33 acoustic parameters. Before voice recording, the examiner instructed the participants to calmly and spontaneously phonate the vowel /a/ (as the most commonly used one) for 3–4 seconds in a sitting position. According to the authors' recommendations, this procedure was repeated three times to get the best quality voice. A microphone was placed at a 5 cm distance from the participant's mouth. The signal was recorded directly on a computer.

We selected the sustained phonation of a vowel as a speech task in this research because previous studies have shown that continuous speech tasks, such as spontaneous speech, tend to have greater prosodic and segmental variability, and that phoning vowels provides more consistent results (Gerratt et al., 2015).

We analysed nine acoustic parameters: frequency variability (F0, Jitt, PPQ), amplitude variability (Shim, vAm, APQ), and noise and tremor parameters (NHR, VTI, SPI).

F0, Jitter, Shimmer, and NHR are the most frequently analysed acoustic features of voice in people with depression. The other analysed parameters (PPQ, APQ, VTI, and SPI) were selected due to their significance in studying voice disorders and insufficient research in the field of voice in depression.

The MDVP programme has frequently been used for voice assessment in various participants in the Serbian-speaking area, and has significant diagnostic and therapeutic implications, e.g. in monitoring the effects of treatment in vocal polyps (Petrović-Lazić et al., 2014; Petrović-Lazić et al., 2009).

Procedure

We explained the research aim and procedure to every participant. Only the participants who signed the informed consent for patients were included in the research with the possibility to withdraw at any time. The recording took place in a room isolated from noise. A Sony ECM-T150 microphone (Sony, Tokyo, Japan), attached to headphones at a 5.0 cm distance from the participant's mouth, was used for recording. All participants were instructed to sustain the phonation of the vowel /a/ for 3–4 seconds. Every phonation was recorded three times. In this way, the best quality recording was included in the analysis. The registered signal was recorded directly on a computer. The research was conducted during the optimal period, from 11 to 12 o'clock, when the participants were not tired.

Statistical analysis

SPSS 23.00 was used for statistical analysis. The obtained values of acoustic parameters were presented by descriptive statistics methods: mean and standard deviation. Analysis of variance was used to test the differences between the experimental and control groups. We used the Tukey post hoc test to further examine partial comparison of the groups. Cohen's d was used to measure the effect size of the differences. Discriminant analysis determined which of the parameters had the highest discriminant value for the groups. Power was calculated by the G-power programme and it was indicated that for the effect size analysis of 0.40, the measurement error of α = 0.05 (and therefore the power of the test of β = 0.95), 3 groups of participants and 0 covariates, the predicted smallest sample size was 162.

Results

Differences in acoustic features of voice between groups

The one-way analysis of variance (ANOVA) was used to examine whether the EG was different from CG1 and CG2 in the acoustic features of voice. The independent variable in the analysis was the group (EG, CG1, and CG2), while the dependent variables were acoustic voice features (9 of them). There were 18 participants in the EG, 24 in the CG1, and 9 in the CG2. Cohen's d and eta-squared tests measured the effect size of the differences between the groups. The results of group comparisons (EG vs. CG1, CG1 vs. CG2, and EG vs. CG2) were tested by the Tukey post hoc test. Table 2 shows the results of the analyses.

Table 2
Results of the variance analysis examining the differences in the acoustic features of voice (F0, Jitter, Shimmer, NHR, vAm, APQ, PPQ, VTI, SPI, dependent variable) in three groups of participants (EG, CG1, CG2) (the factor or independent variable)

Parameter / group	Experimental group M (SD)	Control group 2 M (SD)	Control group 1 M (SD)
F0	163.58 (43.487)	153.07 (53.821)	232.82 (47.189)
Jitter	3.35 (2.638)	1.63(0.618)	0.55(0.267)
Shimmer	11.96(4.704)	6.26(3.350)	2.10(1.024)
NHR	0.31(0.191)	0.17(0.055)	0.11(0.013)
vAm	27.07(8.479)	11.41(4.462)	9.41(4.275)
APQ	9.51(3.235)	4.50(2.178)	1.49(0.694)
PPQ	2.09(1.691)	0.95(0.371)	0.32(0.158)
VTI	0.18(0.105)	0.08(0.048)	0.04(0.015)
SPI	4.59(2.416)	11.64(3.486)	6.58(3.273)

Notes: Acoustic features of voice: F0, Jitter, Shimmer, NHR, vAm, APQ, PPQ, VTI, SPI; three groups of participants: experimental group, control group 1, and control group 2; M – mean, SD – standard deviation

The results showed that the EG had significantly higher values than CG2 and CG1 on many scales of acoustic parameters: NHR (0.31 vs. 0.17 vs. 0.11; F(2/48)=15.627, p<.01), Shimmer (11.96 vs. 6.26 vs. 2.10; F(2/48)=49.060, p<.001), Jitter (3.35 vs. 1.63 vs. 0.55; F(2/48)=15.744; p<.001), VTI (0.18 vs. 0.08 vs. 0.04; F(2/48)=20.923; p<.001), APQ (9.51 vs. 4.50 vs. 1.49; F(2/48)=70.082; p<.001), vAm (27.07 vs. 11.41 vs. 9.41; F(2/48)=15.296, p<.001). In F0 and SPI parameters, the EG had lower mean values than the other two groups (control group 1 and control group 2) and statistically significant differences in the acoustic parameters (F0 (163.58 vs. 232.82 vs. 153.07; F(2/48)=15.296, p<.001), SPI (4.59 vs. 11.64 vs. 6.58; F(2/48)=16.251, p<.001)). These results indicate that the greatest differences between EG, CG1, and CG2 (expressed in η^2) were found in Shimmer, APQ, and vAm parameters.

We used discriminative analysis using the stepwise method, in which predictors are included one by one, making the layers of the analysis. Results of the discriminative analysis via the stepwise method indicated a model with two significant discriminative functions (Box'M = 53,153; F(12/3036)=3,917, p<.01).

Table 3.
Results of the canonical standardized function coefficients and function at group centroids

		Function	
		1	2
Standardized Canonical Discriminant Function Coefficients	vAm	.418	-.526
	APQ	.716	.659
	SPI	-.083	.969
	group	**1**	**2**
Function at the group centroid	EG	2.349	-.234
	CG1	-1.551	-.515
	CG2	-.561	1.841

The final model included three layers. At the first layer, only the APQ voice feature was significant (Wilks L=.255; F(2/48)= 70.082, p<.01), at the second, SPI was significant (Wilks L=.160, F(4/94)=35.233, p<.01), while at the third it was vAm (Wilks L=.129, F (6/92)= 27.344, p<.01).

The first discriminative function has its own eigenvalue of 3.331 and explained 80.9% of variance of the criteria (Wilks L=.129 ($\chi^2(6)$=96.221, p<.01), while the second has its own eigenvalue of .789 and explained 19.1% of variance of the criteria (Wilks L=.559; ($\chi^2(2)$=27.334, p<.01). The value of the canonical correlation of the first discriminative function is .877, while that of the second is .664.

At the first discriminative function, the parameters vAm and APQ have the highest value of the standardized coefficient of the canonical discriminative function, while at the second the parameter SPI has the highest value (Table 3).

Table 4.
Results of the predicted group membership according to the discriminant functions

	group	EG	CG1	CG2
Original	count	17	0	1
		0	21	3
		0	0	9
	%	94.4	0	5.6
		0	87.5	12.5
		0	0	100
Cross validated	count	16	1	1
		0	21	3
		0	0	9
	%	88.9	5.6	5.6
		0	87.5	12.5
		0	0	100

The first discriminative function is for the depressive, while the other is for voice disorders (Table Coefficients of the matrix structure of the discriminative functions also indicated such a pattern of the results). Results of the classification of the partcipants according to the results of the discriminative function indicate that one participant from the depressive group should be classified in voice disorder, and the three participants classified in the depressive should be in voice disorders (Table 4).

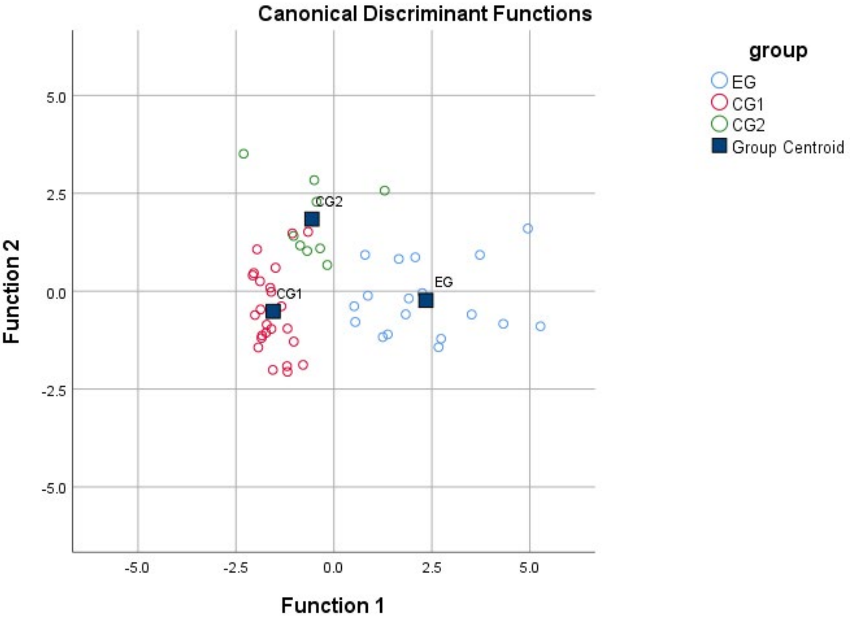

Figure 1. The overall presentation of the results is organized in a coordinate space defined by two discriminative functions. The first (Function 1) represents the X-axis, while the second (Function 2) represents the Y-axis. The EG is marked by blue circles, CG2 by red, and CG1 by green circles.

The figure indicates three separate groups of participants. It can be clearly seen that the EG is the most prominent in the first discriminative function (hence the highest discriminative function coefficients for this group of disorders in the table that represents the group centroids). In addition, we can notice that voice disorders are in a completely different quadrant of the coordinate system from depression. However, they have the highest scores in the second discriminative function according to the position in group centroids (where they have the highest coefficients on the Y-axis).

Discussion

Although with some inconsistent results, the literature largely indicates the existence of differences in acoustic features of voice in people with depression compared to a control group of healthy controls (Low et al., 2011; Mundt et al., 2007; Nunes et al., 2010; Sahu & Espy-Wilson, 2016; Taguchi et al., 2018; Quatieri & Malyska, 2012). We wanted to examine whether introducing an additional subgroup of participants with psychogenic voice disorders could significantly strengthen the existing results and contribute to the literature on depression and its relation to acoustic features of voice. Thus, we conducted an acoustic analysis of voice in adults suffering from depression (EG), aiming to compare their differences in relation to healthy controls (CG1) and adults with a psychogenic voice disorder (CG2).

The analysis of variance showed that practically all mean values of the parameters significantly differed between the EG and CG1, and between the EG and CG2. Almost all parameters in the EG (parameters of frequency variability (Jitt, PPQ), amplitude variability (Shim, vAm, APQ) and noise and tremor parameters (NHR, VTI)) had higher values except for fundamental frequency (F0) in relation to CG1 and soft index phonation (SPI) in relation to both control groups. These results indicated that the EG differed from both control groups in all measures of frequency and its variability, amplitude variability, and noise and tremor assessment in this research.

The fundamental frequency (F0) of voice (subjectively observed as voice pitch) is one of the most frequently examined acoustic variables. The EG had lower mean values in F0 compared to CG1 and higher compared to CG2. This is in accordance with numerous studies indicating that F0 is lower in depression (Mundt et al., 2007; Mundt et al., 2012; Silva et al., 2021; Wang et al., 2019). The fundamental frequency of voice refers to the number of vibration cycles in one second (the frequency at which vocal cords open and close). Lower F0 can result from slower vocal cord vibrations due to bad mood and stress. Slow and long vibrations produce a low tone of voice. Being directly associated with vocal cord vibrations, F0 can be observed through the relation to the speaker's general muscle tension. The strong relation between F0 and muscle tension is particularly significant (Ellgring & Scherer, 1996). Taguchi et al. (2018) did not find significant differences in F0 between the participants with depression and the control group. However, they used the reading numbers task in their study and explained that the short pauses between the numbers possibly affected the F0 value (Taguchi et al., 2018). Based on vowel phonation, our research found that F0 was significant for depression. Interestingly, the value of this parameter in our study was higher in the EG than in CG2 (when considering mean values, see Table 2). Also, our study showed a lower variability of the fundamental frequency. Moore et al. (2004) point out that a smaller range of F0 indicates monotonous speech in people with depression. This decrease in the F0 range and its variability can, according to some authors, be explained by the increased rigidity in the phonation mechanism based on high muscle tension (Ellgring & Scherer, 1996). From the paralinguistic aspect, speech intelligibility depends on prosodic segments. F0 variations are considered the main carrier of prosodic information (Ellgring & Scherer, 1996). Ozdas et al. (2000) assume that the difference between perceptive descriptions of a suicidal voice heard by clinicians is based on long-term F0 variations rather than its fluctuations between periods.

The results of our research indicate a statistically significant increase in the mean values of Jitter and Shimmer parameters in the EG compared to both control groups, which is in accordance with research findings in this field (Low et al., 2011; Sahu & Espy-Wilson, 2016). One of the possible effects of high laryngeal tension caused by emotional stress is the irregularity of vocal cord vibrations that can lead to increased Jitter parameter (Scherer, 1986). Jitter measures short-term, cyclic irregularities of a voice period (Zwetsch et al., 2006). A study aiming to determine suicidal tendency on the basis of voice features showed that the Jitter parameter was higher in participants with this tendency than in the control group (Ozdas et al., 2000). Shimmer indicates amplitude variations during vocal cord vibrations (Zwetsch et al., 2006). A recent study also found an increase in Shimmer and Jitter, which is in accordance with our results (Silva et al., 2021). The literature is mainly consistent with regard to the values of the mentioned parameters. We are familiar with one study in which Shimmer was lower, but the parameter was analysed on a speech sample for the emotion of sadness and not in participants with depression (Nunes et al., 2010).

According to some authors, different airflow during speech increases the noise-to-harmonics ratio in people with depression compared to healthy controls (Low et al., 2011). This was also confirmed in our study in relation to both control groups. However, some studies indicate that this parameter is lower in participants suffering from depression (Quatieri & Malyska, 2012), explaining that laryngeal muscle tension decreases in depression due to a motor impairment (weakening) which causes greater glottal turbulence. NHR (noise-to-harmonics ratio) is the ratio between harmonic and noise (non-harmonic) voice components. Noise may come from the glottis, resulting from supraglottic constriction, and may be mixed. An increase in vocal intensity occurs in greater vocal cord occlusion due to excessive muscle tension, increased vibration amplitude, and increased subglottal pressure. NHR is an objective indicator of breathiness as a perceptual feature.

Although F0, Jitter, Shimmer and NHR are the most commonly examined acoustic parameters in people with depression, it is also important to examine other parameters of frequency and amplitude variability, and noise and tremor parameters. The selected additional parameters are significant in researching voice disorders. Apart from these most frequently examined parameters, our results show a significant increase in the APQ and vAm (variation of amplitude, Cappellari and Cielo, 2008) parameters, indicating the possible effect of increased laryngeal tension due to emotional stress and amplitude irregularities. APQ is the amplitude perturbation quotient in percentages and is thus an appropriate Shimmer measure (Petrović-Lazić et al., 2014). This parameter usually increases in breathy and hoarse voice. The vAm indicates long-term variations in the peak amplitude expressed in percentages (Cappellari & Cielo, 2008). PPQ is the pitch period perturbation quotient (Cappellari & Cielo, 2008). This parameter measures short-term irregularities in this period.

PPQ is the only parameter that was not statistically significant in our research. In addition to NHR, VTI is also a good indicator of breathiness and was higher in the EG in our study. VTI measures the ratio between the energy density of high-frequency noise (in the range of 1800–5800 Hz) and the spectral energy density of harmonics (in the range of 70–4200 Hz) (Cappellari & Cielo, 2008). It represents the voice turbulence index.

Apart from the lower F0, our research also showed a decrease in the soft phonation index (SPI) in the EG compared to both control groups. This parameter refers to the approximation and tightening of vocal cords during phonation (Heđever, 2012). It implies the average ratio between low-frequency harmonics (in the range of 70–1550 Hz) and high-frequency harmonics (in the range of 1600–4200 Hz). The increased SPI value usually indicates insufficient closure and tightening of vocal cords during phonation. On the other hand, low values of this parameter imply an increase in vocal cord adduction (Roussel & Lobdell, 2006). However, the literature shows that this parameter decreases with the increase of F0. Thus, it is possible that the sample size is the reason for the discrepancy between these parameters.

In our research, amplitude variability parameters – vAm and APQ – had the highest discriminant value for voice in depression (Figure 1). These findings are in accordance with a study indicating that Shimmer (the parameter from the domain of amplitude variability) was the parameter most associated with depression, unlike Jitter (Quatieri & Malyska, 2012). However, some authors point out that lowering the fundamental tone and reducing the variability range of F0 may be the main objective parameter indicating depressed speech, i.e., the basic measure of the difference between participants with depression and healthy controls (Nilsonne, 1988). Ozdas et al. (2004) point out that the increase in Jitter and glottal spectral tilt can be observed as the acoustic features distinguishing between people with depression, suicidal people, and healthy controls (Ozdas et al., 2004). Also, a recent paper shows that Jitter is the parameter that predicts most whether participants have depression (Silva et al., 2021). The authors believe that neurophysiological changes in depression cause glottis irregularities and thus affect motor and dynamic coordination of the larynx. Our results could potentially provide significant insights with regard to predicting depression based on the acoustic features of voice, which would be our focus in future research.

Our findings confirm previous findings that there are differences in the acoustic features of voice between the EG and CG1, and also between the EG and the added CG2. These new findings, along with the confirmed previous ones, preliminarily indicate that voice in depression is a separate entity independent of psychogenic voice disorders. This could facilitate the diagnostic process by differentiating this mental disorder from other comorbid diagnoses. By pointing out important acoustic parameters associated with depression, this pilot research could contribute to creating a programme for automatic recognition of depression. However, the practical implementation requires significantly more systematic research in this field, where we see the future of multidisciplinary cooperation of experts in mental health, voice, and computer engineering.

Limitations

According to the power calculation and the small sample, it is necessary to conduct further research in this field on a larger sample in order to generalize the results. Also, it would be potentially significant to divide the sample into subgroups according to the severity of depression. As a variable that can make a difference in acoustic values, gender was not taken into account, as well as whether the history of smoking and taking medications can be confounding variables, which would be significant to examine in future studies. There are indications of medication effects on voice, although they have not been shown to affect the ability to discriminate between the depressed and control groups (Silva et al., 2021). Therefore, future research should certainly consider a research design that would include both the participants who are taking and who are not taking medicine. The relation between acoustic analysis and other voice and speech analyses (e.g. spectral analysis) and depression should also be examined in this sample in future studies. Further, the study should be replicated on other sounds, e.g. words and different types of speech tasks (such as reading, continuous speech, etc.), to confirm the results. It could be useful to compare depression voice with other types of voice disorders. In future studies, it is necessary to thoroughly examine depressive states in all groups, not only based on self-reports.

Conclusion

The acoustic features of voice analysed on the basis of the sustained phonation of the vowel /a/ in this research were different and discriminant in the EG compared to CG1 and CG2. Practically all analysed acoustic parameters were different between these groups. Discriminant analysis indicated that vAm and APQ had the highest discriminant value for depression. The results of this research point to the importance of the voice, that is, its acoustic indicators in the recognition of depression. Important parameters that could serve to create a programme for automatic recognition of depression are those from the domain of voice intensity variation.

Acknowledgments

This manuscript was a result of research within the projects of the Ministry of Education, Science and Technological Development of the Republic of Serbia (no. 451-03–68/2022–14).

References

Afshan, A., Guo, J., Park, S.J., Ravi, V., Flint, J., & Alwan, A. (2018, september). *Effectiveness of voice quality features in detecting depression*. In Interspeech 2018. ISCA, Hyderabad, India (pp. 1676–1680.) https://doi.org/10.21437/Interspeech.2018–1399.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2013). Detecting depression: A comparison between spontaneous and read speech. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp.7547–7551). https://doi.org/10.1109/ICASSP.2013.6639130

American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Washington DC: American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596.

Baek, Y.-S., Kim, S.-J., Kim, E., & Choi, Y. (2012). Vocal acoustic characteristics of speakers with depression. *Korean Society of Speech Sciences*, *4*(1), 91–98. https://doi.org/10.13064/KSSS.2012.4.1.091

Baken, R. J., & Orlikoff, R. F. (2000). Clinical measurement of speech and voice (2nd ed.). San Diego, CA: Singular Thomson Learning.

Bueno-Notivol, J., Gracia-García, P., Olaya, B., Lasheras, I., López-Antón, R., & Santabárbara, J. (2021). Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. *International Journal of Clinical and Health Psychology*, *21*(1), 100196. https://doi.org/10.1016/j.ijchp.2020.07.007

Cappellari,V. M. & Cielo,C. A. (2008). Vocal acoustic characteristics in pre-school aged children. *Brazilian Journal of Otorhinolaryngology*,*74* (2), 265–272. https://doi.org/10.1016/S1808–8694(15)31099–5.

Ćuk-Jovanović, L. (2002). Akustička analiza govornog signala pacijenata sa depresivnim poremećajem – karakteristike trajanja (The acoustic analysis of the speech signal of the patients with a depressive dissorder: Characteristics of duration). *Engrami*, *24*(2), 15–23.

Ćuk-Jovanović, L. (2003). Intenzitet govornog signala pacijenata sa depresivnim poremećajem (The intensity of the speech signal of the patients with a depressive disorder). *Govor i jezik* (pp.217–223). Beograd: Institut za eksperimentalnu fonetiku i patologiju govora.

Cummins, N., Epps, J., Breakspear, M., & Goecke, R. (2011). An Investigation of Depressed Speech Detection: Features and Normalization. Proceedings of the INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association. Florence, Italy: International Speech Communication Association (pp.2997–3000). https://doi.org/10.21437/Interspeech.2011–750

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. https://doi.org/10.1016/j.specom.2015.03.004.

Darby, J. K., Simmons, N., & Berger, P. A. (1984). Speech and voice parameters of depression: A pilot study. *Journal of Communication Disorders*, *17*(2), 75–85. https://doi.org/10.1016/0021–9924(84)90013–3, Ellgring, H., & Scherer, R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, *20*(2), 83–110. https://doi.org/10.1007/BF02253071.

Ferrand, C. T. (2002). Harmonics-to-Noise Ratio. *Journal of Voice*, *16*(4), 480–487. doi:10.1016/s0892–1997(02)00123–6

Fuller, B. F., Horii, Y., & Conner, D. A. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Research in Nursing & Health*, *15*(5), 379–389. https://doi.org/10.1002/nur.4770150507

Hashim, N. W., Wilkes, M., Salomon, R., Meggs, J., & France, D. J. (2017). Evaluation of voice acoustics as predictors of clinical depression scores. *Journal of Voice*, *31*(2), 256.e1–256.e6. https://doi.org/10.1016/j.jvoice.2016.06.006

He, L., & Cao, C. (2018). Automated depression analysis using convolutional neural networks from speech. *Journal of Biomedical Informatics*, *83*, 103–111. https://doi.org/10.1016/j.jbi.2018.05.007

Heđever, M. (2012). *Govorna akustika* (Speech acoustics*)*. Zagreb: Zagreb University, Faculty of Education and Rehabilitation Sciences

Institute of Health Metrics and Evaluation. Global Health Data Exchange (GHDx). http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b Accessed February 2022.

Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., Kang, H., & Li, X. (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, *90*, 39–46. https://doi.org/10.1016/j.specom.2017.04.001

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033–2909.129.5.770

Kiss, G., & Jenei, A. Z. (2020). Investigation of the accuracy of depression prediction based on speech processing. 2020 43rd International Conference on Telecommunications and Signal Processing (TSP) (pp.129–132.) https://doi.org/10.1109/TSP49548.2020.9163495

Kosztyła-Hojna, B., Moskal, D., Łobaczuk-Sitnik, A., Kraszewska, A., Zdrojkowski, M., Biszewska, J., & Skorupa, M. (2018). Psychogenic voice disorders. *Otolaryngologia polska*, *72*(4), 26–34. https://doi.org/10.5604/01.3001.0012.0636

Kuny, S., & Stassen, H. H. (1993). Speaking behavior and voice sound characteristics in depressive patients during recovery. *Journal of Psychiatric Research*, *27*(3), 289–307. https://doi.org/10.1016/0022–3956(93)90040–9

Lopez-Otero, P., & Docio-Fernandez, L. (2020). Analysis of gender and identity issues in depression detection on de-identified speech. *Computer Speech & Language*, 101118. https://doi.org/10.1016/j.csl.2020.101118

Low, L.-S. A., Maddage, M. C., Lech, M., Sheeber, L. B., & Allen, N. B. (2011). Detection of clinical depression in adolescents' speech during family interactions. IEEE Transactions on Biomedical Engineering, *58*(3), 574–586. https://doi.org/10.1109/TBME.2010.2091640

Milutinovic, Z. (1997). *Klinički atlas poremećaja glasa: Teorija i praksa* (Clinical atlas of voice disorders: Theory and practice). Belgrade: Institute for textbook publishing and teaching aids

Moore, E. I. I., Clements, M., Peifer, J., & Weisser, L. (2004). Comparing objective feature statistics of speech for classifying clinical depression. The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, *1*, 17–20. https://doi.org/10.1109/IEMBS.2004.1403079

Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, *20*(1), 50–64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Nilsonne, A. (1988). Speech characteristics as indicators of depressive illness. *Acta Psychiatrica Scandinavica*, *77*(3), 253–263. https://doi.org/10.1111/j.1600–0447.1988.tb05118.x

Nunes, A., Coimbra, R. L., & Teixeira, A. (2010). Voice quality of European Portuguese emotional speech. Computational Processing of the Portuguese Language, International Conference on Computational Processing of the Portuguese Language, 6001, (pp.142–151.) https://doi.org/10.1007/978–3–642–12320–7_19

Ozdas, A., Shiavi, R. G., Silverman, S. E., Silverman, M. K., & Wilkes, D. M. (2004). Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Transactions on Biomedical Engineering*, *51*(9), 1530–1540. https://doi.org/10.1109/TBME.2004.827544

Ozdas, A., Shiavi, R. G., Silverman, S. E., Silverman, M. K., & Wilkes, D. M. (2000). Analysis of fundamental frequency for near term suicidal risk assessment. SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. "Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions", *5*, 1853–1858. https://doi.org/10.1109/ICSMC.2000.886379

Patel, S., & Scherer, K. R. (2013). Vocal behaviour. In: Hall JA, Knapp ML, editors. *Handbook of nonverbal communication*. Berlin: Mouton-DeGruyter (pp.167–204.) https://doi.org/10.1515/9783110238150.167

Petrović-Lazić, M., Babac, S., Ivanković, Z., & Kosanović, R. (2009). Multidimenzionalna akustička analiza patološkog glasa (Multidimensional Acoustic Analysis of Pathological Voice). *Srpski arhiv za celokupno lekarstvo*, *137*(5–6), 234–238. https://doi.org/10.2298/SARH0906234P

Petrović-Lazić, M., Jovanović, N., Kulić, N., Babac, S., & Jurisić, V. (2014). Acoustic and perceptual characteristics of the voice in patients with vocal polyps after surgery and voice therapy. *Journal of Voice*, *29*(2), 241–246. https://doi.org/10.1016/j.jvoice.2014.07.009

Popović, M. (2003). Akustičke karakteristike govora i psihološko-emocionalni faktori (Acoustic characteristics of speech and psychological-emotional factors). *Govor i jezik* (pp.210–216.), Beograd: Institut za eksperimentalnu fonetiku i patologiju govora

Quatieri, T., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity, In Interspeech 2012, 13th Annual Conference of the

International Speech Communication Association Portland, OR, USA https://doi.org/10.21437/Interspeech.2012–311

Rejaibi, E., Komaty, A., Meriaudeau, F., Agrebi, S., & Othmani, A. (2022). MFCC-based recurrent neural network for automatic clinical depression recognition and assessment from speech. *Biomedical Signal Processing and Control*, *71*, 103107. https://doi.org/10.1016/j.bspc.2021.103107

Ritchie, H. & Roser, M. (2018). *Mental Health. Our World in Data.* https://ourworldindata.org/mental-health Accessed June 2022.

Roussel, N. C., & Lobdell, M. (2006). The clinical utility of the soft phonation index. *Clinical Linguistics & Phonetics*, *20*(2–3), 181–186. https://doi.org/10.1080/02699200400026942

Sahu, S., & Espy-Wilson, C. (2016). Speech features for depression detection. The Interspeech 2016, 17th Annual Conference of the International Speech Communication Association (pp.1928–1932.) https://doi.org/10.21437/Interspeech.2016–1566

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256. https://doi.org/10.1016/S0167–6393(02)00084–5

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. https://doi.org/10.1037/0033–2909.99.2.143

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*(6), 401–435. https://doi.org/10.1080/00207594.2011.626049

Silva, W. J., Lopes, L., Galdino, M. K. C., & Almeida, A. A. (2021). Voice acoustic parameters as predictors of depression. *Journal of Voice*, Article in Press https://doi.org/10.1016/j.jvoice.2021.06.018

Sturim, D.E., Torres-Carrasquillo, P.A., Quatieri, T., & Malyska, N. (2011). Automatic detection of depression in speech using Gaussian Mixture Modeling with factor analysis. Interspeech 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy https://doi.org/10.21437/Interspeech.2011–746

Taguchi, T., Tachikawa, H., Nemoto, K., Suzuki, M., Nagano, T., Tachibana, R., Nishimura, M., & Arai, T. (2018). Major depressive disorder discrimination using vocal acoustic features. *Journal of Affective Disorders*, *225*, 214–220. https://doi.org/10.1016/j.jad.2017.08.038

Teixeira, J. P., & Fernandes, P. O. (2015). Acoustic analysis of vocal dysphonia. *Procedia Computer Science*, *64*, 466–473. https://doi.org/10.1016/j.procs.2015.08.544

Wang, J., Zhang, L., Liu, T., Pan, W., Hu, B., & Zhu, T. (2019). Acoustic differences between healthy and depressed people: a cross-situation study. *BMC Psychiatry*, *19*(1). https://doi.org/10.1186/s12888–019–2300–7

World Health Organization. Depression and other common mental disorders: Global health estimates. World Health Organization; 2017. http://www.who.int/iris/handle/10665/254610 Accessed August 2021.

Xing, Y., Liu, Z., Li, G. Ding, Z., & Hu, B. (2022). 2-level hierarchical depression recognition method based on task-stimulated and integrated speech features. *Biomedical Signal Processing and Control*, *72*, 103287. https://doi.org/10.1016/j.bspc.2021.103287

Yang, B., & Lugger, M. (2010). Emotion recognition from speech signals using new harmony features. *Signal Processing*, *90*(5), 1415–1423. https://doi.org/10.1016/j.sigpro.2009.09.009

Yang, Y., Fairbairn, C., & Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, *4*(2), 142–150. https://doi.org/10.1109/T-AFFC.2012.38

Zwetsch, I., Fagundes, R., Russomano, T., & Scolari, D. (2006). Digital signal processing in the differential diagnosis of benign larynx diseases. *Scientia Medica*, *16*(3), 109.

DATUM PRIJEMA RADA: 18.7.2022. DATUM PRIHVATANJA RADA: 28.11.2022.

Akustičke karakteristike glasa kod odraslih osoba sa depresivnim poremećajem

Gordana Calić[3]

Fakultet za specijalnu edukaciju i rehabilitaciju, Univerzitet u Beogradu, Srbija

Mirjana Petrović-Lazić

Fakultet za specijalnu edukaciju i rehabilitaciju, Univerzitet u Beogradu, Srbija

Tatjana Mentus

Fakultet za specijalnu edukaciju i rehabilitaciju, Univerzitet u Beogradu, Srbija

Snežana Babac

Fakultet za specijalnu edukaciju i rehabilitaciju, Univerzitet u Beogradu, Srbija

U cilju utvrđivanja razlika između grupe osoba sa depresivnim poremećajem (EG, N=18) u odnosu na grupu osoba iz tipične populacije (CG1, N=24) i grupu osoba sa dijagnostikovanim psihogenim poremećajem glasa (CG2, N=9) analizirano je 9 akustičkih karakteristika glasa primenom MDVP softverskog programa ("Kay Elemetrics" Corp., model 4300) na uzorku od 51 ispitanika. Devet akustičkih parametara analizirano je na osnovu produženog foniranja vokala /a/. Rezultati istraživanja pokazuju da se srednje vrednosti svih akustičkih parametara razlikuju između osoba sa depresivnim poremećajem u odnosu na obe kontrolne grupe i to: parametri varijabilnosti frekvencije (Jitter, PPQ), varijabilnosti amplitude (Shimmer, vAm i APQ), i parametri procene šuma i tremora (NHR i VTI) imaju više vrednosti; samo su parametar fundamentalne frekvencije (F0) i indeks prigušene fonacije (SPI) niži (F0 u odnosu na CG1, i SPI u odnosu na CG2). Samo se parametar PPQ nije pokazao značajnim. Parametri vAm i APQ imaju najveću

[3] calicgordana@yahoo.com

diskriminativnu vrednost za depresivni poremećaj. Akustičke karakteristike glasa analizirane na osnovu produženog foniranja vokala u ovom istraživanju razlikuju i diskriminišu EG i u odnosu na CG1 i u odnosu na CG2. U vokalnoj analizi parametri vAm i APQ bi potencijalno mogli biti markeri koji ukazuju na depresivni poremećaj. Rezultati ovog istraživanja ukazuju na značaj glasa, odnosno njegovih akustičkih pokazatelja, u prepoznavanju depresije. Važni parametri koji bi mogli da pomognu u kreiranju programa za automatsko prepoznavanje depresije su oni iz domena varijacije intenziteta glasa.

Ključne reči: akustičke karakteristike, vokal, depresija, poremećaj glasa, vokalna analiza