PAN
ARCHIVES OF ACOUSTICS
Vol. **43**, No. 2, pp. 245–251 (2018)

DOI: 10.24425/122372

The Acoustic Cues of Fear: Investigation of Acoustic Parameters of Speech Containing Fear

Turgut ÖZSEVEN

Department of Computer Engineering
Gaziosmanpaşa University
Tokat, Turkey; e-mail: turgutozseven@gmail.com

(*received March 4, 2017; accepted January 17, 2018*)

Speech emotion recognition is an important part of human-machine interaction studies. The acoustic analysis method is used for emotion recognition through speech. An emotion does not cause changes on all acoustic parameters. Rather, the acoustic parameters affected by emotion vary depending on the emotion type. In this context, the emotion-based variability of acoustic parameters is still a current field of study. The purpose of this study is to investigate the acoustic parameters that fear affects and the extent of their influence. For this purpose, various acoustic parameters were obtained from speech records containing fear and neutral emotions. The change according to the emotional states of these parameters was analyzed using statistical methods, and the parameters and the degree of influence that the fear emotion affected were determined. According to the results obtained, the majority of acoustic parameters that fear affects vary according to the used data. However, it has been demonstrated that formant frequencies, mel-frequency cepstral coefficients, and jitter parameters can define the fear emotion independent of the data used.

Keywords: emotion recognition; acoustic analysis; fear; speech processing.

1. Introduction

Many modern psychological theories of fear emphasize the two-fold reaction to dangerous stimuli: an immediate reaction caused by an automatically responding, conditioned system and a slightly slower reaction caused by a voluntary, more cognitive system (Hagenaars, van Minnen, 2005). The automatic system contains both physiological and behavioral reactions.

A vocalization is the result of actions of a great number of muscles in the chest, throat, and head, so any alterations in muscle tonus will affect vocal characteristics (Hagenaars, van Minnen, 2005).

The emotional state is the physiological response triggered when a person feels pain, receives a pleasing piece of news or fears in the face of an event, which reflects the voice and face of the person. The response which will be given to events is softened by means of an intermediate layer located between stimulant and emotional response (Scherer, 1984; Tompkins, 1962). At the same time, constant evaluation of stimulants and preparation of emotional reactions to these stimulants are allegedly one of the basic functions of emotion (Arnold, 1960; Scherer, 1982). Learning as a result of emotions and emotional reactions is specific to each person; however, basic features are consistent in every person (Sethu, 2009). Consequently, speech may be changed during exposure to emotional stimuli. This change may be a useful tool for providing information about emotion recognition.

The acoustic parameters obtained via acoustic analysis are used extensively in speech emotion recognition (SER) studies. An important part of these studies is classification based, and the emotion recognition success is examined through various classifiers and acoustic parameters. Although SER studies have been conducted for years, examining the effects of emotions on voice is still an actual topic. Some studies, have examined the effect of fear on acoustic parameters, but the number of used acoustic parameters is limited.

A summary of the results obtained by studies examining acoustic parameters affected by fear emotion is provided in Table 1.

According to Table 1, the F0 mean is generally increased (Diamond *et al.*, 2010; Drioli *et al.*, 2003;

Table 1. Summary of studies examining the relationship between fear and acoustic parameters in the literature.

Acoustic parameter	Change	Reference
F0 mean	Increase	(Diamond *et al.*, 2010; Drioli *et al.*, 2003; Goberman *et al.*, 2011; Murray, Arnott, 1993; Ruiz *et al.*, 1996; Ververidis, Kotropoulos, 2006; Weeks *et al.*, 2012)
F0 range	Irregular	(Drioli *et al.*, 2003; Protopapas, Lieberman, 1997; Ververidis, Kotropoulos, 2006)
F0 standard deviation	Irregular	(Goberman *et al.*, 2011; Hagenaars, van Minnen, 2005)
Duration	Decrease	(Ververidis, Kotropoulos, 2006)
Speech rate	Increase	(Murray, Arnott, 1993)
Pause rate	Decrease	(Goberman *et al.*, 2011; Laukka *et al.*, 2008)
Intensity	Irregular	(Drioli *et al.*, 2003; Laukka *et al.*, 2008; Murray, Arnott, 1993)
HNR (Harmonic to Noise Ratio)	Irregular	(Murray, Arnott, 1993)
Jitter/Shimmer	Increase	(Fuller *et al.*, 1992)

Goberman *et al.*, 2011; Murray, Arnott, 1993; Ruiz *et al.*, 1996; Ververidis, Kotropoulos, 2006; Weeks *et al.*, 2012). In the study in which non-verbal parts of speech were examined, the F0 mean decreased (Laukka *et al.*, 2008). According to studies which examined the changes in the F0 range, it was found out that no change occurred in this parameter (Protopapas, Lieberman, 1997), whereas some studies found out that the value of this parameter decreased (Drioli *et al.*, 2003; Ververidis, Kotropoulos, 2006) and yet some other studies detected an increase (Diamond *et al.*, 2010). This shows that the F0 range value is irregular in the case of fear. The same also applies for the F0 standard deviation, and there are studies which found decrease (Hagenaars, van Minnen, 2005) as well as the increase (Goberman *et al.*, 2011). According to the studies examining the parameters related to the physical structure of speech, it has been found that the duration has decreased (Ververidis, Kotropoulos, 2006), the speech rate value has increased very rapidly (Murray, Arnott, 1993) and the pause rate has decreased (Goberman *et al.*, 2011; Laukka *et al.*, 2008). In the case of fear, the intensity was found to be normal (Murray, Arnott, 1993), high (Drioli *et al.*, 2003) and low in non-verbal parts (Laukka *et al.*, 2008). It was found out that there were irregularities in the voice quality (Murray, Arnott, 1993) and increase in the Jitter and Shimmer values (Fuller *et al.*, 1992).

Studies examining the relationship between fear and acoustic parameters in the literature were performed through a limited number of acoustic parameters. In addition, whether the difference that was revealed depended on the used data has not been examined. In this context, the diversity of the acoustic parameter set, and the investigation of the relationship between fear and acoustic parameters on different data sets, will make a significant contribution to the literature. In addition, the use of different spoken languages in the data sets used will guide the SER researcher working in the language context.

In this study, the EMO-DB and EMOVO databases and formant frequency, bandwidth, the mel-frequency cepstral coefficient (MFCC), and linear prediction cluster coefficients (LPCC) were used. The purpose of selection of two different datasets is to investigate whether the relationship between fear and acoustic parameters changes according to language.

The purpose of the present study was to examine the change of the acoustical parameters that fear affects, by analyzing the acoustic features of speech. The acoustic analysis was performed using Praat (Boersma, Weenink, 2002). The statistical analysis was conducted with IBM SPSS Statistics 20 software.

In the 2nd section the methodology of the study is given. The 3rd section contains the results of statistical analysis and the relationship between fear and acoustic parameters. In the 4th section, the results are interpreted.

2. Method

2.1. The used data

One of the most important problems in SER involves obtaining a dataset with the tested performance that consists of natural emotional states. Current studies use readily available databases (such as EMO-DB, EMOVO, SUSAS, eNTERFACE) (Murray, Arnott, 1993; Ruiz *et al.*, 1996) or sound recordings that researchers have collected (Diamond *et al.*, 2010; Drioli *et al.*, 2003; Fuller *et al.*, 1992; Goberman *et al.*, 2011; Hagenaars, van Minnen, 2005; Protopapas, Lieberman, 1997; Weeks *et al.*, 2012).

Two emotional speech databases (EMO-DB and EMOVO) were used to remove the data set dependency of the findings in the study. A total of 316 speech expressions with fear and neutral emotions were used. EMO-DB was obtained via the expression of various emotions (anger, anxiety/fear, boredom, disgust, happiness, sadness, neutral) with actors in German. Emotional reflection qualities of sound recordings in EMO-DB were evaluated by perceptual analysis. For perceptual analysis, 20 participants listened and scored the voice recordings. Voice records have 16 kHz sampling frequency and 16-bit mono (Burkhardt *et al.*, 2005). EMOVO was obtained via the expression of various emotions (disgust, joy, fear, anger, surprise, sadness, neutral) with six actors and 14 sentences in German. The recordings were performed with a sampling frequency of 48 kHz, 16-bit stereo, wave format (Costantini *et al.*, 2014). In order to evaluate the emotional success of the voice recordings in EMOVO, the voice recordings were played by independent assessors and they were asked to select one of two possible emotions (ex: anger/joy). Table 2 contains detailed information on the used data.

Table 2. Distribution of used data.

Emotion		EMO-DB	EMOVO
Neutral	Male	39	42
	Female	40	42
	Count of different speaker	10	6
	Count of different sentence	10	14
	Language	German	Italian
	TOTAL	79	84
Fear	Male	36	42
	Female	33	42
	Count of different speaker	10	6
	Count of different sentence	10	14
	Language	German	Italian
	TOTAL	69	84
TOTAL		316	

2.2. Acoustic analysis

Acoustic analysis of speech records was performed with the help of Praat software (Boersma, Weenink, 2002), and 77 acoustic parameters were obtained from each speech recording. When the acoustic parameters were calculated, 25 ms frame size, 50% overlap, and Hamming windowing were used. Acoustic parameters obtained with acoustic analysis are given in Table 3.

F0 is defined as the number of opening and closing per second of glottis and the number of vibrations per second of the vocal cords. It gives the pitch of voice. The period is a primary factor in F0 perception. It is measured in loops/seconds and expressed in hertz (Hz). This value is 220–240 Hz in girls and boys before adolescence, whereas it is between 100–150 Hz and 150–250 Hz on average for adult men and women respectively (Sarıca, 2012). Formant frequency is the resonance in sound path, which provides spectral information on the quantitative features of vocal tract. The formant bandwidth interacts with the amplitude and the rate of absorption of the sound energy. As the absorption rate of the sound energy increases, the format bandwidth increases and the amplitudes of the sound wave decreases. Jitter is a parameter, which shows the difference between periods of F0. It consists of involuntary irregularities which occur in F0 (Kılıç, Okur, 2001). This parameter is defined as the changes of basic frequency between subsequent vibrant loops (Farrus, Hernando, 2009). The periodic variation between amplitude peaks is called shimmer. This parameter is defined as the amplitude changes of throat flow between loops with subsequent vibrations (Deshmukh *et al.*, 2005). HNR is the ratio of the total energy of F0 and harmonics, which are its multiples to the noise energy. The unvoiced frame rate is the ratio of silent sections during the speech. Voiced break is the number of pauses during the speech. Intensity indicates the resultant energy of the amplitude change in the speech signal. MFCC and LPCC are speech analysis methods based on the human hearing system (Sethu, 2009). Because it is a spectral feature, it relates to the structure of the vocal tract.

Table 3. The acoustic parameters used in this paper.

Acoustic parameters	Descriptive Statistics		
	Mean	Std. dev.	Median
Fundamental frequency (F0)	✓	✓	
Formant frequency (F1, F2, F3)	✓	✓	
Formant bandwidth			✓
Jitter (local, rap)/ Shimmer (local, apq3)	✓		
HNR (Harmonic to Noise Ratio)	✓	✓	
Unvoiced frame	✓		
Voiced break	✓		
Intensity	✓	✓	
MFCC0...14	✓	✓	
LPCC1...13	✓	✓	

2.3. Statistical analysis

In this study, the descriptive statistics is used at the stage of processing of each speech file. While processing speech files, speech utterance is divided into frames and feature extraction is performed on each frame. Thus, an acoustic parameter belonging to a speech file consists of values the number of which is equal to the number of used frames. The mean, standard deviation and median of these values are taken and the number of the parameter for each file is reduced to one.

For normality test of the data used in the study, Shapiro-Wilk test is used. Following normality test, with the purpose of evaluating the relation of parameters with emotion groups, Mann-Whitney U test or variance analysis was preferred depending on the distribution of data. In statistical analysis 95% reliability level ($p < 0.05$) was used.

Shapiro and Wilk (1965) test was originally restricted for sample size of less than 50. This test was the first test that was able to detect departures from normality due to either skewness or kurtosis, or both (Althouse *et al.*, 1998). It has become the preferred test because of its good power properties (Mendes, Pala, 2003). Shapiro-Wilk test was made usable by being modified by Royston with data which consist of a sample size of 2000, data consisting of samples larger than 50 and data consisting of sample between 3 and 5000 (Razali, Wah, 2011). Mann-Whitney U test is used with the purpose of comparing measures as regards a dependent variable consisting of two groups and determining whether there is a significant difference between two contributions. Variance analysis is used in comparing mean values of more than two samples. With this method interaction between variables and experimental mistakes of various change sources that contribute to the total change are examined.

3. Results

The analysis was performed on a laptop with the 2.7 GHz processor and 16 GB of RAM. Analysis of each audio file took 6.5 seconds on average. According to the results of the Shapiro-Wilk test, two emotion groups in both EMOVO and EMO-DB databases did not obey normal distribution. For this reason, Mann-Whitney U test is used in the analysis. Parameters that show difference according to groups and the degree of relations are given in Tables 4 and 5.

The F0 mean, F0 std, F1 mean, F2 mean, F3 mean, F1 std, F1 bandwidth, F3 bandwidth, jitter(local), unvoiced frame, intensity std, MFCC std and MFCC

Table 4. The results of Mann-Whitney U test for EMOVO.

Acoustic parameters	Mean Rank		MWU	p
	Neutral	Fear		
F0 mean	74.51	93.60	2689.00	0.011*
F0 std	76.19	91.90	2830.00	0.036*
F1 mean	72.80	95.34	2545.00	0.003*
F2 mean	70.82	97.34	2379.00	0.001*
F3 mean	63.33	104.92	1750.00	0.000*
F1 std	70.61	97.55	2361.00	0.000*
F2 std	90.88	77.04	2908.00	0.064
F3 std	79.94	88.11	3145.00	0.275
F1 bandwidth	70.80	97.36	2377.00	0.001*
F2 bandwidth	76.79	91.30	2880.00	0.052
F3 bandwidth	72.99	95.14	2561.00	0.003*
Jitter (local)	76.54	91.55	2859.00	0.045*
Jitter (rap)	83.32	84.69	3428.50	0.854
Shimmer (local)	83.15	84.86	3415.00	0.820
Shimmer (apq3)	78.11	89.96	2991.50	0.114
HNR mean	87.46	80.49	3195.00	0.352
HNR std	77.86	90.22	2970.00	0.099
Unvoiced frame	67.63	100.57	2110.50	0.000*
Voiced break	85.60	82.39	3352.00	0.668
Intensity mean	82.92	85.10	3395.00	0.771
Intensity std	73.58	94.54	2611.00	0.005*
MFCC std	90.74	77.18	2523.43	0.009*
MFCC mean	83.75	84.25	2562.27	0.011*
LPCC mean	78.79	89.28	3048.00	0.161
LPCC std	79.61	88.45	3117.00	0.238

* $p < 0.05$, MWU: Mann-Whitney U.

Table 5. The results of Mann-Whitney U test for EMO-DB.

Acoustic parameters	Mean Rank		MWU	p
	Neutral	Fear		
F0 mean	53.29	98.78	1050.00	0.000*
F0 std	68.11	81.81	2221.00	0.053
F1 mean	58.94	92.32	1496.00	0.000*
F2 mean	60.77	90.22	1641.00	0.000*
F3 mean	65.89	84.36	2045.00	0.009*
F1 std	70.18	79.45	2384.00	0.189
F2 std	90.14	56.59	1490.00	0.000*
F3 std	79.87	68.35	2301.00	0.103
F1 bandwidth	54.13	97.83	1116.00	0.000*
F2 bandwidth	71.90	77.48	2520.00	0.430
F3 bandwidth	75.29	73.59	2663.00	0.810
Jitter (local)	77.62	70.93	2479.00	0.343
Jitter (rap)	88.01	59.04	1658.50	0.000*
Shimmer (local)	71.73	77.67	2507.00	0.401
Shimmer (apq3)	72.45	76.85	2563.50	0.534
HNR mean	79.68	68.57	2316.00	0.116
HNR std	81.27	66.75	2191.00	0.040*
Unvoiced frame	71.72	77.69	2505.50	0.398
Voiced break	74.71	74.26	2709.00	0.949
Intensity mean	79.65	68.61	2319.00	0.118
Intensity std	70.56	79.01	2414.00	0.231
MFCC std	75.64	73.19	1917.19	0.009*
MFCC mean	86.49	60.77	1625.04	0.008*
LPCC mean	66.16	84.04	1471.57	0.001*
LPCC std	87.53	59.58	1696.00	0.004*

* $p < 0.05$, MWU: Mann-Whitney U.

mean parameters outlined in Table 4 can be used for the purpose of discriminating two emotions, as they reveal differences depending on the emotional state. When the mean rank values of these parameters are examined, MFCC std is more effective for the neutral emotion, whereas other parameters are more effective for fear. When the Mann-Whitney U value and its significance is examined (p), the strongest and weakest parameters in discriminating neutral and fear emotions are the F3 mean and Jitter (local) parameters respectively.

The F0 mean, F1 mean, F2 mean, F3 mean, F2 std, F1 bandwidth, jitter (rap), HNR std, MFCC and LPCC parameters outlined in Table 5 can be used for the purpose of discriminating two emotions, as they reveal differences depending on the emotional state. When the mean rank values of these parameters are examined, it can be seen that neutral emotion is more effective on the F2 std, Jitter (rap), HNR std, MFCC mean, MFCC std, and LPCC std parameters whereas in other effective parameters fear emotion is more influential. When Mann-Whitney U value and its significance is examined (p), the strongest and weakest parameters in discriminating neutral and fear feelings are F0 mean and HNR std parameter respectively.

The results obtained from both databases are examined according to their significance and rank values and the changes in the parameters are given in Table 6.

Table 6. The effect on acoustic parameters of fear.

Acoustic parameters	Change	
	EMOVO	EMO-DB
F0 mean	⇗	⇑
F0 std	⇗	x
F1 mean	⇗	⇑
F2 mean	⇗	⇗
F3 mean	⇑	⇗
F1 std	⇗	x
F2 std	x	⇘
F3 std	x	x
F1 bandwidth	⇗	⇑
F2 bandwidth	x	x
F3 bandwidth	⇗	x
Jitter (local)	⇗	x
Jitter (rap)	x	⇘
Shimmer (local)	x	x
Shimmer (apq3)	x	x
HNR mean	x	x
HNR std	x	⇘
Unvoiced frame	⇑	x
Voiced break	x	x
Intensity mean	x	x
Intensity std	⇗	x
MFCC std	⇘	⇔
MFCC mean	⇔	⇘
LPCC mean	x	⇗
LPCC std	x	⇘

⇓: strong decrease, ⇘: decrease, ⇑: strong increase, ⇗: increase, ⇔: no change, x: ineffective ($p \geq 0.05$).

According to the results provided in Table 6, fear increases the value of the parameters related to F0. Regarding the states of the formant frequencies, the mean and bandwidth values are increasing, whereas standard deviation reveals a variance. If Jitter (local) has validity in the used data, it demonstrates an increase, whereas Jitter (rap) has decreased. HNR std. reveals a decrease, as it was effective in EMO-DB database. The unvoiced frame and intensity std values are effective in EMOVO, which shows an increase. The LPCC mean and LPCC std values do not reveal an impact on EMOVO. Meanwhile, they showed variance in EMO-DB and helped with defining the fear emotion, which does not lead to any change in the shimmer, voiced break, intensity mean and loudness parameters in both databases.

4. Discussion and conclusion

This study investigated the effects of fear on acoustic parameters. According to the results we obtained, the strongest and weakest parameters in distinguishing neutral and fear emotions in the EMO-DB database are the F0 mean and HNR std respectively. As for the EMOVO database, the strongest and weakest parameters are the F3 mean and jitter (local) parameters respectively. According to the used data, the strongest and weakest parameters show variance; however, in the literature, acoustic parameters related to fear emotion are effective in both databases. This case shows that the F0, speech rate and jitter parameters widely used in studies can define the fear emotion independent from the used data.

In previous studies, it was determined that the shimmer value showed the increase in fear. Unlike the literature, out results did not find any relation between fear and shimmer in both databases. In addition, the voiced break and intensity mean parameters do not show variance in the case of fear. Interpreting the results obtained for the F0 and F0 mean showed a resemblance to the results obtained in the literature and showed an increase in both databases. Because F0 is associated with the rate of the glottis, it will change in the case of emotional arousal. This change suggests that fear increases subglottal pressure.

According to the results of formant frequencies, in addition to the literature, the mean and bandwidth values increase in the case of fear. On the other hand, standard deviation values show variance depending on the used data. This change in formant frequencies indicates that the vowel sounds affect the tongue when one feels fear. The HNR and intensity std parameters reveal a resemblance to the literature and vary according to the used data.

According to the literature, fear increases the number of pauses in speaking. Our results showed that, unlike the literature, pauses (unvoiced frame) increased in EMOVO, whereas it did not have an effect on EMO-DB.

In previous studies, the relation of fear with MFCC and LPCC parameters was not investigated. In our study, these parameters investigated, and we found that the LPCC parameter did not have any effect on EMOVO but increased in EMO-DB. On the other hand, MFCC parameters decreased in the case of fear independent from the acquisition method of the data. In the case of fear, the change in MFCC parameters indicates that fear is reflected in speech. This means people will be able to perceive fear while listening to a conversation.

In addition to showing the feasibility and validity of the results obtained in previous studies, this study has also found changes in different parameters in the case of fear. According to the results obtained on the two corpora, the parameters mostly used in the literature can define the emotion independently of the data. The differences in the results obtained on the two corpora vary with the spoken language and accent.

References

1. Althouse L.A., Ware W.B., Ferron J.M. (1998), *Detecting departures from normality: a monte carlo simulation of a new omnibus test based on moments*, presented at the Annual Meeting of the American Educational Research Association, San Diego, California, https://eric.ed.gov/?id=ED422385.

2. Arnold M.B. (1960), *Emotion and personality*, Columbia University Press, New York.

3. Boersma P., Weenink D. (2002), *Praat, a system for doing phonetics by computer*, Glot International, **5**, 9–10, 341–345.

4. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W.F., Weiss B. (2005), *A database of German emotional speech*, Interspeech, **5**, 1517–1520.

5. Costantini G., Iaderola I., Paoloni A., Todisco M. (2014), *EMOVO Corpus: an Italian Emotional Speech Database*, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 3501–3504.

6. Deshmukh O., Espy-Wilson C.Y., Salomon A., Singh J. (2005), *Use of temporal information: Detection of periodicity, aperiodicity, and pitch in speech*, IEEE Transactions on Speech Audio Process, **13**, 5, 776–786.

7. Diamond G.M., Rochman D., Amir O. (2010), *Arousing primary vulnerable emotions in the context of unresolved anger: "Speaking about" versus "speaking to"*, Journal of Counseling Psychology, **57**, 4, 402–410.

8. Drioli C., Tisato G., Cosi P., Tesser F. (2003), *Emotions and voice quality: experiments with sinusoidal modeling*, [in:] *Voice Quality: Functions, Analysis and Synthesis (VOQUAL'03), ISCA Tutorial and Research Workshop*, Christophe d'Alessandro, Klaus R. Scherer [Eds.], Geneva, Switzerland, August 27–29, 2003, ISCA Archive, http://www.isca-speech.org/archive_open/voqual03, pp. 127–132.

9. Farrus M., Hernando J. (2009), *Using jitter and shimmer in speaker verification*, IET Signal Processing, **3**, 4, 247–257.

10. 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 and Health, **15**, 5, 379–389.

11. Goberman A.M., Hughes S., Haydock T. (2011), *Acoustic characteristics of public speaking: anxiety and practice effects*, Speech Communication, **53**, 6, 867–876.

12. Hagenaars M.A., van Minnen A. (2005), *The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia*, Journal of Anxiety Disorders, **19**, 5, 521–537.

13. Kiliç M.A., Okur E.K.M. (2001), *Comparison of fundamental frequency and perturbation data analyzed by CSL and Dr. Speech systems*, The Turkish Journal of Ear Nose and Throat, **8**, 2, 152–157.

14. Laukka P. *et al.* (2008), *In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech*, Journal of Nonverbal Behavior, **32**, 195–214.

15. Mendes M., Pala A. (2003), *Type I error rate and power of three normality tests*, Information Technology Journal, **2**, 2, 135–139.

16. Murray I.R., Arnott J.L. (1993), *Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion*, Journal of Acoustical Society of America, **93**, 2, 1097–1108.

17. Protopapas A., Lieberman P. (1997), *Fundamental frequency of phonation and perceived emotional stress*, Journal of Acoustical Society of America, **101**, 4, 2267–2277.

18. Razali N.M., Wah Y.B. (2011), *Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests*, Journal of Statistical Modeling and Analytics, **2**, 1, 21–33.

19. Ruiz R., Absil E., Harmegnies B., Legros C., Poch D. (1996), *Time-and spectrum-related variabilities in stressed speech under laboratory and real conditions*, Speech Communication, **20**, 1–2, 111–129.

20. Sarica S. (2012), *Acoustic parameters that used in voice analysis*, PhD Thesis, Kahramanmaras Sutcu Imam University.

21. Scherer K.R. (1982), *Emotion as a process: Function, origin and regulation*, Social Science Information, **21**, 4, 555–570.

22. Scherer K.R. (1984), *On the nature and function of emotion: a component process approach*, [in:] Klaus R. Scherer, Paul Ekman [Eds.], *Approaches to emotion*, pp. 293–317, Hillsdale, NJ: Erlbaum.

23. Sethu V. (2009), *Automatic emotion recognition: an investigation of acoustic and prosodic parameters*, PhD Thesis, The University of New South Wales.

24. Shapiro S.S., Wilk M.B. (1965), *An analysis of variance test for normality (complete samples)*, Biometrika, **52**, 3–4, 591–611.

25. Tompkins S. (1962), *Affect Imagery Consciousness. Vol. I: The Positive Affects*, Springer Publishing Company.

26. Ververidis D., Kotropoulos C. (2006), *Emotional speech recognition: Resources, features, and methods*, Speech Communication, **48**, 9, 1162–1181, doi: 10.1016/j.specom.2006.04.003.

27. Weeks J.W. *et al.* (2012), *"The sound of fear": Assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder*, Journal of Anxiety Disorders, **26**, 8, 811–822.