Procena emocionalnog stanja govornika statističkom analizom fundamentalne frekvencije

Zoran Milivojević
Department for Information-communication technology
Academy of Applied Technical and Preschool Studies
Niš, Serbia
zoran.milivojevic@akademijanis.edu.rs

Bojan Prlinčević
Department of Information-communication technology
Kosovo and Metohija Academy of Applied Studies
Leposavić, Serbia
bojan.prlincevic@akademijakm.edu.rs

Dijana Kostić
Department for Information-communication technology
Academy of Applied Technical and Preschool Studies
Niš, Serbia
koricanac@yahoo.com

Sažetak—**U prvom delu rada opisan je algoritam za procenu emocionalnog stanja iz govora na bazi fundamentalne frekvencije F_0. Najpre je opisan algoritam za određivanje kriterijuma, odnosno linije odlučivanja, u ravnima (F_0, σ^2) i (F_0, T). Nakon toga je dat opis algoritma za klasifikovanje emocionalnog stanja govornika. U drugom delu rada je opisan Eksperiment u kojem je izvršena statistička procena preciznosti klasifikacija emocionalnog stanja govornika. Eksperimentalni rezultati su prikazani tabelarno i grafički. Statistička analiza eksperimentalnih rezultata obavljena je primenom matrice konfuzije. Na kraju je, na osnovu statističkih podataka, izvršena komparativna analiza preciznosti procene emocije u ravnima (F_0, σ^2) i (F_0, T).**

Ključne riječi-Fundamentalna frekvencija, emocija, matrica konfuzije

I. UVOD

U eri razvoja veštačke inteligencije pravljenje uređaja za prepoznavanje emocija iskazanih govorom nije novina. Prilikom verbalne i vizuelne komunikacije dve osobe prepoznavanje iskazanih emocija ne predstavlja problem. Međutim, računari i uređaji bazirani na veštačkoj inteligeniciji moraju na neki način oponašati mehanizme ljudske percepcije. Prva istraživanja u ovoj oblasti sprovedena su 1980-tih godina i bazirala su se na statističkim analizama akustičkih karakteristika [1], [2]. U periodu 1990-tih godina primenjeni su napredni algoritmi kojima su vršene procene akustičkih karakteristika govora [3], [4]. Tokom 2000-tih godina fokus istraživanja je na pronalaženju klasifikatora emocija kojima se unapređuje efikasnost primene algoritama u aplikacijama u svakodnevnoj primeni. U naučnoj literaturi opisan je veći broj metoda i kalsifikatora govornih emocija, kao što su: HMM (engl. *Hidden Markov model*) [5], GMM (engl. *The Gaussian mixture model*) [5], [6], SVM (engl. *Support vector machine*) [5], [7], [8], NB (engl. *Naive Bayes classifiers*) [9], [10], KNN (engl. *K-nearest Neighbours approach*) [11], [12] i ANN (engl. *Artificial Neural Network*) [13].

Najveći broj algoritama je baziran na opšte poznatoj činjenici da emocionalno stanje ima važan uticaj na govor. Najvažnije saznanje svih sprovedenih studija jeste da se prosečna fundamentalna frekvenicja F_0 povećava za emocije koje predstavljaju uzbuđenje (bes, strah, sreća..) a opada za emocije koje predstavljaju stanje manje uzbuđenosti (tuga, melanholija, dosada...) [14]. Međutim, u vremenu intenzivne interakcije čovek računar, istraživanje u oblasti prepoznavanja emocija iskazanih govorom predstavlja ogroman izazov. Fokus je baziran na istraživanjima koja se mogu primenjivati u realnim scenarijima, kao što su call-centri [15], [16] , terapeuti u medicini/psihologiji [17] i dr.

U cilju efikasnosti primene algoritama za prepoznavanje emocija u govoru potrebno je kreirati baze emocionalnog govora, koje služe za obuku/treniranje algoritama. Kreiranje velikih baza emocionalnog govora, koje uključuju različite izgovore govornika, su neophodne, kako bi se verno procenio učinak algoritama za emocionalno prepoznavanje govora. U svetu su trenutno dostupne emocionalne baze govora za nemački, kineski, tajvanski ... Ove baze govora su kreirane tako što su glumci simulirali izgovor odabranih rečenica u definisanom emocionalnom stanju. Verifikacija verodostojnosti definisane emocije izgovorenih rečenica procenjivali su i ocenjivali slušaoci (MOS test). Na osnovu ocene slušalaca određena je korelacija između definisane i stvarne emocije. Rečenice koje nisu dobile zadovoljevajuću ocenu su uklonjene iz baze.

U ovom radu je izvršena procena emocionalnog stanja na bazi trajektorije fundamentalne frekvencije F_0. Kreiran je algoritam za procenu emocionalnog stanja koji se sastoji iz dva

dela: a) kreiranje baze za treniranje algoritma i kreiranje linije odlučivanja (engl. *decision line*) za procenu emocionalnog stanja u ravnima (F_0, σ^2) i (F_0, T), i b) testiranje efikasnosti algoritma procene emocija. Procena fundamentalne frekvencije, F_0, govornog signala, odnosno trajektorije fundamentalne frekvencije $F_0(n)$, realizovana je primenom softverskog paketa Praat [18]. Algoritam za testiranje je baziran na ugrađenoj Matlab funkciji *classify(trainingData)*. Dobijeni rezultati verifikovani su primenom matrice konfuzije.

Rad je organizovan na sledeći način: Sekcija II opisuje Algoritam za određivanje linije odlučivanja i Algoritam za procenu emocionalnog stanja. U sekciji III je opisan Eksperiment i prikazani su rezultati. Sekcija IV je Zaključak.

II. ALGORITMI PROCENE EMOCIONALNOG STANJA

U cilju procene emocionalnog stanja na osnovu govora korišćena je baza "Berlin database of emotional speech" [19] nad kojom su primenjivana dva algoritma. Prvi algoritam je Algoritam za određivanje linije odlučivanja, a drugi algoritam je Algoritam za procenu emocionalnog stanja.

A. *Algoritam za određivanje linije odlučivanja (Algoritam 1)*

Algoritam za određivanje linije odlučivanja realizovan je u sledećim koracima:

Ulaz: $\mathbf{x}_k$-audio signal, k = 1, 2, …, K, ukupan broj signala za testiranje.

Izlaz: ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$) - koeficijenti linije odlučivanja u ravni $P(F_0, \sigma^2)$, ($a_{F_0,T}$, $b_{F_0,T}$) - koeficijenti linije odlučivanja u ravni $P(F_0, T)$.

FOR $k = 1 : K$

Korak 1: Kreiranje trajektorije fundamentalne frekvencije, F_{0k}, signala $\mathbf{x}_k$, primenom Praat-a [18].

Korak 2: Određivanje srednje vrednosti trajektorije fundamentalne frekvencije $\overline{F_{0k}}$, i generisanje niza $F_0(k) = \overline{F_{0k}}$.

Korak 3: Određivanje varijanse trajektorije fundamentalne frekvencije σ_k^2 , i generisanje niza $\sigma^2(k) = \sigma_k^2$.

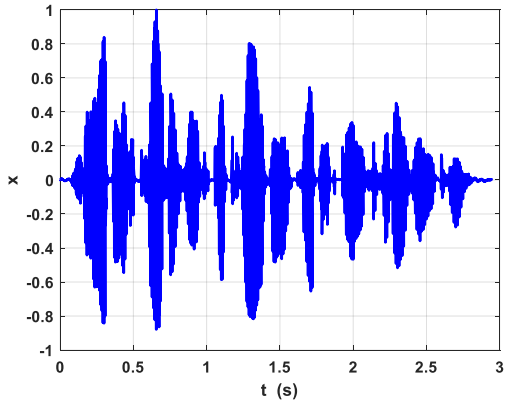
Korak 4: Određivanje trajanja T_k audio signala $\mathbf{x}_k$, i generisanje niza $T(k) = T_k$.

END k

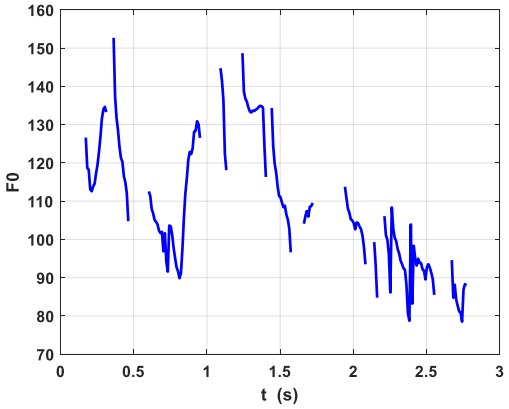
Korak 5: Izračunavanje koeficijenata ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$) linije odlučivanja u ravni $P(F_0, \sigma^2)$ primenom Matlab funkcije [$a_{F_0,\sigma}$, $b_{F_0,\sigma}$] = *classify*(*trainingData*(F_0,σ^2), '*linear*').

Korak 6: Izračunavanje koeficijenata ($a_{F_0,\sigma}$,T) linije odlučivanja u ravni $P(F_0, T)$ primenom Matlab funkcije [$a_{F_0,T}$, $b_{F_0,T}$] = *classify*(*trainingData*(F_0,T), '*linear*').

Primer audio signala **x,** nad kojim se primenjuje algoritam prikazan je na sl. 1.a. Trajektorija fundamentalne frekvencije F_0 prikazana je na sl.1.b (*Korak 1*). Na sl. 2.a prikazana je ravan $P(F_0, \sigma^2)$, linija odlučivanja i pozicije emocionalnih stanja nekih govornika iz testne baze (*Korak 5*). Na je slici 1.b prikazana ravan $P(F_0, T)$, linija odlučivanja i pozicije emocionalnih stanja nekih govornika iz testne baze (*Korak 6*).

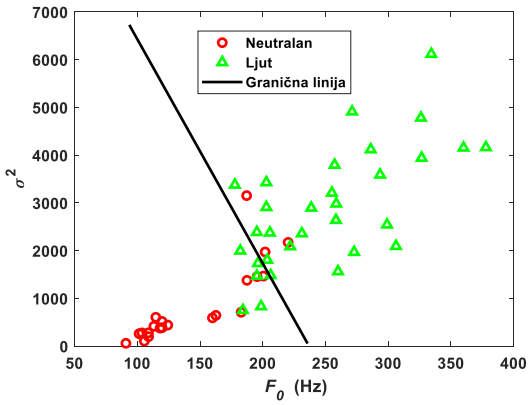


a)

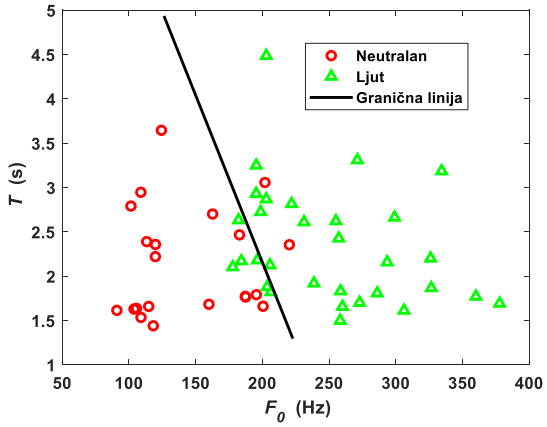


b)

Slika 1. Audio signal: a) vremenski oblik, b) trajektorija fundamentalne frekvencije.



a)



b)

Slika 2. Linija odlučivanja i pozicije emocionalnih stanja u: a) $P(F_0, \sigma^2)$ ravni, b) $P(F_0, T)$ ravni.

B. *Algoritam za procenu emocionalnog stanja (Algoritam 2)*

Algoritam za procenu emocionalnog stanja iz govora realizovan je u sledećim koracima:

Ulaz: **x**-audio signal, ($a_{F_0,\sigma}, b_{F_0,\sigma}$) - koeficijenti granične linije u ravni $P(F_0, \sigma^2)$, ($a_{F_0,T}, b_{F_0,T}$) - koeficijenti granične linije u ravni $P(F_0, T)$.

Izlaz: E_1, E_2 - emocija.

Korak 1: Kreiranje trajektorije fundamentalne frekvencije F_0 signala **x**.

Korak 2: Određivanje srednje vrednosti trajektorije fundamentalne frekvencije $\overline{F_0}$.

Korak 3: Određivanje varijanse trajektorije fundamentalne frekvencije σ^2.

Korak 4: Određivanje trajanja T audio signala **x**.

Korak 5: Klasifikacija emocije u $P(F_0, \sigma^2)$ ravni:

IF $\overline{F_0} <= (\sigma^2 - b_{F_0,\sigma}) / a_{F_0,\sigma}$

E_1 = 'Ljut'

ELSE

E_2 = 'Neutralan'

END

Korak 6: Klasifikacija emocije u $P(F_0, T)$ ravni:

IF $\overline{F_0} <= (T - b_{F_0,T}) / a_{F_0,T}$

E_1 = 'Ljut'

ELSE

E_2 = 'Neutralan'

END

III. EKSPERIMENTALNI REZULTATI I ANALIZA

A. *Eksperiment*

U cilju testiranja efikasnosti algoritama za prepoznavanje emocije iz govornog signala sproveden je Eksperiment. Eksperiment je realizovan na sledeći način: a) kreirana je baza govornih signala za treniranje algoritma procene emocije od baze [19], b) kreirane su linije odlučivanja u $P(F_0, \sigma^2)$ i $P(F_0, T)$ ravni primenom Algoritma 1; c) izvršeno je testiranje preciznosti procene emocionalnog stanja govornika primenom Algoritma 2; i d) izvršena je statistička analiza preciznosti procene emocionalnog stanja govornika korišćenjem matrice konfuzije.

Za potrebe definisanja kriterijuma za odlučivanju o tipu emocije (emocionalno stanje „Neutralno“ i „Ljutnja“) rečenice iz Test baze podeljene su u dve grupe. Prva grupa rečenica kreirana je za potrebe definisanja kriterijuma odlučivanja tipa emocije (Algoritam 1). Kao kriterijum odlučivanja korišćena je linija razdvajanja u ravnima $P(F_0, \sigma^2)$ i $P(F_0, T)$. Po slučajnom izboru odabrano je 60% rečenica i od njih formirana baza za treniranje Algoritma 1. Druga grupa rečenica iz Test baze (preostalih 40%) iskorišćeno je za testiranje efikasnosti, odnosno preciznosti klasifikacije emocionalnog stanja (Algoritam 2). Nad dobijenim rezultatima (detektovana emocija) primenjena je komparativna statistička analiza, primenom matrice konfuzije.

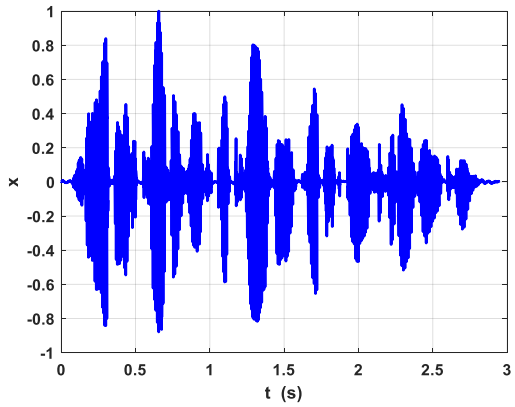
TABELA I. MATRICA KONFUZIJE

		DETEKTOVANA EMOCIJA	
	UKUPNO P + N	LJUT (PP)	NEUTRALAN (PN)
STVARNA EMOCIJA	LJUT (P)	STVARNO LJUT (TP)	LAŽNO NEUTRALAN (FN)
	NEUTRALAN (N)	LAŽNO LJUT (FP)	STVARNO NEUTRALAN (TN)

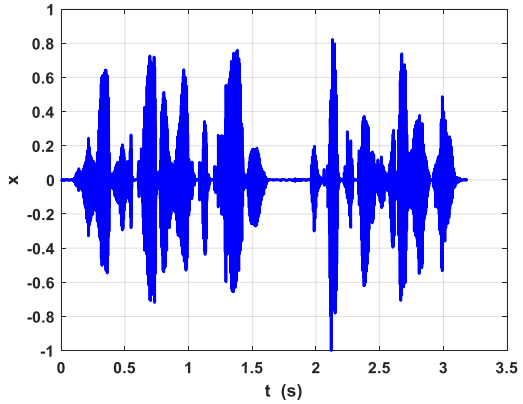
Matrica konfuzije prikazana je u Tabeli I. Oznake u tabeli su: P – broj rečenica sa emocionalnim stanjem „Ljuto“, N – broj rečenica sa emocionalnim stanjem „Neutralno“, PP – broj detektovanih stanja „Ljut“ i PN – broj detektovanih stanja „Neutralan“. Tačnost verifikacije je izračunata na osnovu parametara na sledeći način: TP (True Positive) – stvarno detektovana emocija „Ljut“, TN (True Negative) – stvarno detektovana emocija „Neutralan“, FP (False Positive) – lažno detektovana emocija „Ljut“, FN (False Negative) – lažno detektovana emocija „Neutralan“. Statistički parametri za komparaciju preciznosti algoritma 2: a) TPR = TP / (TP + FN) – (engl. true positive rate - TPR) udeo tačno detektovanih emocija „Ljut“, b) TNR = TN / (TN + FP) – (engl. true negative rate - TNR) udeo tačno detektovanih emocija „Neutralan“, PPV = TP / (TP + FP) – (engl. positive predictive value - PPV) prediktivna vrednost emocije „Ljut“, NPV = TN / (TN + FN) – (engl. negative predictive value - NPV) prediktivna vrednost emocije „Neutralan“, i ACC = (TP + TN) / (TP + TN + FP + FN) – (engl. Accuracy - ACC) tačnost procene (emocija).

B. *Test baza*

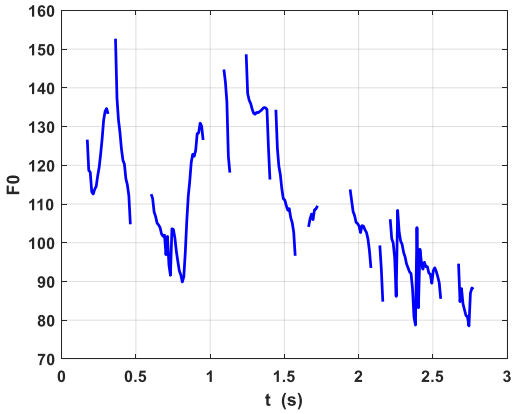
Test baza kreirana je od dela rečenica iz "Berlin database of emotional speech" [19]. Uzete su rečenice koje se odnose na dva emocionalna stanja, i to emocionalno stanje „Neutralno" i „Ljuto". Kao primer na sl.3 prikazan je vremenski oblik signala i trajektorije fundamentalne frekvencije F_0(t), za slučaj kada je govornik muška osoba, izgovorila rečenicu "Sie haben es gerade hochgetragen unf jetz gehen sie wieder runter" (srp. "Samo su ga poneli i sada opet idu dole"), sa emocionalnim stanjem: a) ljutnja (sl. 3.a), b) neutralno (sl. 3.b).



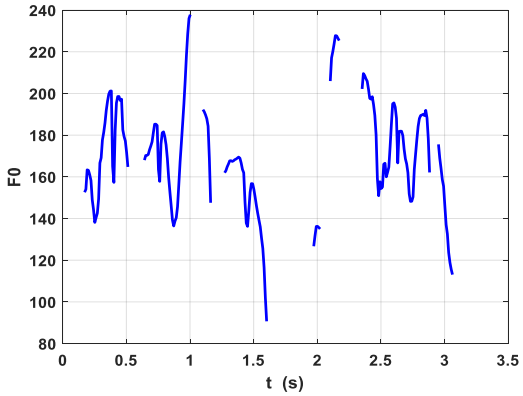
a)



b)



c)



d)

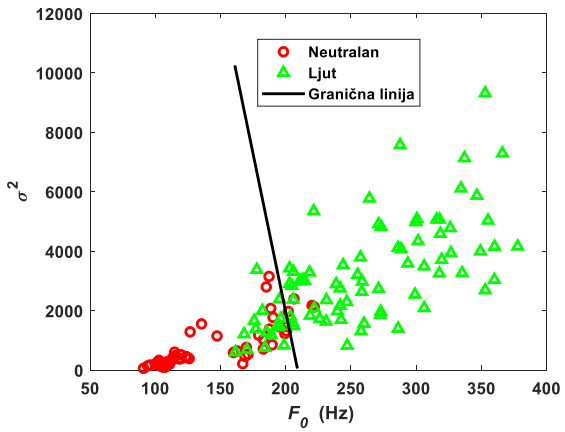
Slika 3. Prikaz audio signala izgovorenog od strane muškarca 31g: a) "Neutralan" govor , i b) "Ljut" govor; i prikaz trajektorije fundamentalne frekvencije izgovorene rečenice za: c) "Neutralan" govor, d) "Ljut" govor.

C. *Razultati*

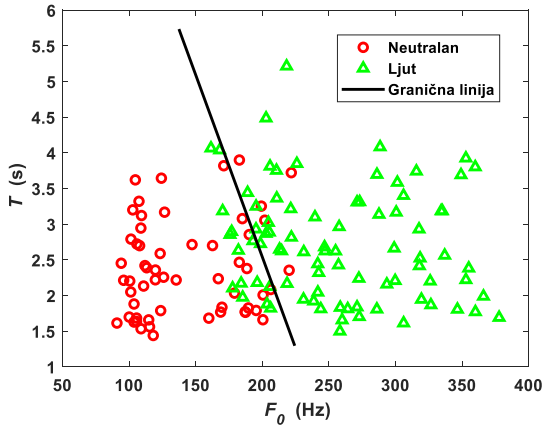
Primenom Algoritma 1 određeni su koeficijenti linije odlučivnja za ravni: a) $P(F_0, \sigma^2)$ ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$) i b) $P(F_0, T)$ ($a_{F_0,\sigma}$,T). Na sl. 4.a prikazana je ravan $P(F_0, \sigma^2)$ i linija odlučivanja. Na sl. 4.b prikazana je ravan $P(F_0, T)$ i linija odlučivanja. Statistički rezultati, elementi matrice konfuzije prikazani su u Tabeli II. Srednje vreme trajanja govornog signala za test rečenice iz baze: a) $\overline{T_N}$ = 2.379 s (emocija „Neutral") i b) $\overline{T_A}$ = 2.711 s (emocija „Ljutnja").

TABELA II. REZULTATI MATRICE KONFUZIJE U RAVNIMA $P(F_0, \sigma^2)$ I $P(F_0, T)$.

Emocija	Ravan P	
	$P(F_0, \sigma^2)$	$P(F_0, T)$
P	31	31
N	21	21
TP	27	28
FP	1	2
TN	20	19
FN	4	3
PP	28	30
PN	24	22
TPR	0.870968	0.903226
TNR	0.952381	0.904762
PPV	0.964286	0.933333
NPV	0.833333	0.863636



a)



b)

Slika 4. Pozicioniranje emocija i linija odlučivanja baze emocionalnih stanja u: a) ravni $P(F_0, \sigma^2)$, b) ravni $P(F_0, T)$.

D. Analiza razultata

Na osnovu srednjih vremena trajanja izgovrenih rečenica zaključuje se da je trajanje izgovorenih rečenica pod emocijom „Ljutnja“ veće od trajanja izgovorenih rečenica pod emocijom „Neutralan“ $\overline{T_A} / \overline{T_N} = 2.711 / 2.379 = 1.14$ puta.

Na osnovu rezultata prikazanih u Tabeli II zaključuje se da je u ravni $P(F_0, \sigma^2)$ u odnosu na ravan $P(F_0, T)$, ispravna klasifikacija emocije:

a) „Ljutnja“ veća $TP_{F_0,s^2} / TP_{F_0,T} = 27 / 28 = 0{,}96$ puta,

b) „Negativna“ veća $TN_{F_0,s^2} / TN_{F_0,T} = 20 / 19 = 1{,}052$ puta,

dok je neispravna klasifikacija emocije:

a) „Ljutnja“ veća $FP_{F_0,s^2} / FP_{F_0,T} = 1 / 2 = 0{,}5$ puta.

b) „Neutralan“ veća $FN_{F_0,s^2} / FN_{F_0,T} = 4 / 3 = 1{,}33$ puta.

Rezultati dobijeni primenom matrice konfuzije pokazuju da su vrednosti klasifikacije u ravni $P(F_0, \sigma^2)$ u odnosu na ravan $P(F_0, T)$, veće kod tačno detektovanih emocija:

a) „Ljutnja“, $TPR_{F_0,s^2} / TPR_{F_0,T} = 0{,}8709867 / 0{,}903226 = 0{,}986$ puta, i

b) „Neutralan“, $TNR_{F_0,s^2} / TNR_{F_0,T} = 0{,}8709867 / 0{,}903226 = 0{,}986$ puta,

dok je prediktivna vrednost emocije veća kod emocionalnog stanja:

a) „Ljut“, $PPV_{F_0,s^2} / PPV_{F_0,T} = 0{,}964286 / 0{,}933333 = 1{,}033$ puta, i

b) „Neutralan“, $NPV_{F_0,s^2} / NPV_{F_0,T} = 0{,}833333 / 0{,}863636 = 0{,}965$ puta.

Naredna istraživanja odnosiće se na analizu emocionalnog stanja govornika na srpskom jeziku što podrazumeva kreiranje baze emocionalnog govora.

IV. ZAKLJUČAK

Analiza emocionalnog stanja govornika dobijena je primenom algoritma za određivanje linije odlučivanja (Algoritam 1) i algoritma za procenu emocionalnog stanja (Algoritam 2). Analizom dobijenih rezultata može se zaključiti da je srednje vreme trajanja govora izgovoreno emocijom „Ljutnja“ u odnosu na emociju „Neutralan“ 1.14 puta veće. Analizom rezultata, dobijenih iz matrice konfuzije, prikazanih u Tabeli II, zaključuje se da Algoritam 2 u ravni $P(F_0, \sigma^2)$ ima veću preciznost ispravne klasifikacije emocije „Neutral“ u odnosu na emocije „Ljutnja“ 1.33 puta. Predikciona vrednost Algoritma 2 u ravni $P(F_0, T)$ ima veću vrednost za emociju „Ljutnja“ 1.033 puta u odnosu na emociju „Neutralan“.

LITERATURA

[1] V. Bezooijen, 1984. The Characteristics and Recognizability of Vocal Expression of Emotions. Foris, Drodrecht, The Netherlands.

[2] F.J. Tolkmitt and K.R. Scherer, Effect of experimentally induced stress on vocal parameters. J. Exp. Psychol. [Hum. Percept.] 12 (3), 302–313. 1986.

[3] D. Cairns and J.H.L Hansen, Nonlinear analysis and detection of speech under stressed conditions. J. Acoust. Soc. Am. 96 (6), 3392–3400, 1994.

[4] B.D. Womack and J.H.L., Hansen, Classification of speech under stress using target driven features. Speech Comm. 20, 131–150, 1996.

[5] M.E. Ayadi, M.S. Kamel and F. Karray, ”Survey on speech emotion recognition: Features, classification schemes, and databases”, Pattern Recognition, pp. 572–587, 2011.

[6] A.P.Wanare, S.N. Dandare, ”Human Emotion Recognition From Speech”, Int. Journal of Engineering Research and Applications, vol. 4, Issue 7, pp.74–78, July 2014.

[7] Y. Pan, P. Shenand L. Shen, ”Speech Emotion Recognition Using Support Vector Machine”, International Journal of Smart Home vol.6, no.2, April, 2012.

[8] S. Xu,Y. Liu and X. Liu, ”Speaker Recognition and Speech Emotion Recognition Based on GMM”, 3rd International Conference on Electric and Electronics, 2013.

[9] K.H. Hyun, E.H. Kim, Y.K. Kwak, ”Improvement of Emotion Recognition by Baysian classifier using non-zero-pitch concept”, 2005 IEEE International Workshop on Robots and Human Interactive Communication, 2005.

[10] K.C. Wang, ”Time-Frequency Feature Representation Using Multi-Resolution Texture Analysis and Acoustic Activity Detector for Real-Life Speech Emotion Recognition”, Sensors, vol.15, no. 1, pp.1458–1478, 2015.

[11] B.I. Ashish, S. Chaudhari, ”Speech Emotion Recognition using Hidden Markov Model and Support Vector Machine”, International Journal of Advanced Engineering research Study, vol.1, pp. 316–318, 2012.

[12] V. Srinivas, C.S. Rani and T. Madhu, ”Neural Network based Classification for Speaker Identification”, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol.7, no.1, pp.109–120, 2014.

[13] B. Panda, D. Padhi, K. Dash, S. Mohanty, ”Use of SVM Classifier MFCC in Speech Emotion Recognition System”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, Issue 3, pp.225–230, March 2012.

[14] D. Ververidis, C. Kotropoulos “Emotional speech recognition: Resources, features, and methods”, Speech Communication, Vol. 48, pp. 1162-1181, Aprlil 2006.

[15] V.A. Petrushin,. “Emotion in speech recognition and application to call centers” In: Proc. Artificial Neural Networks in Engineering (ANNIE ’99), Vol. 1, pp. 7–10, 1999.

[16] C.M. Lee, S.S. Narayanan, “Toward detecting emotions in spoken dialogs” IEEE Trans. Speech Audio Process. 13 (2), 293–303, 2005.

[17] D.J. France, R.G. Shiavi, S. Silverman, M. Silverman, M. Wilkes, , “Acoustical properties of speech as indicators of depression and suicidal risk” IEEE Trans. Biomed. Eng. 7, 829–837, 2000.

[18] https://praat.en.softonic.com/

[19] https://www.tu.berlin/en/kw/research/projects/emotional-speech.

ABSTRACT

In the first part of the paper, an algorithm for estimating emotional state from speech based on fundamental frequency F0 is described. First, the algorithm for determining criteria, i.e. decision lines, in the planes (F_0, σ^2) and (F_0, T) is presented. Then, a description of the algorithm for classifying the speaker's emotional state is given. In the second part of the paper, an experiment is described in which a statistical estimation of the accuracy of the speaker's emotional state classification was performed. The experimental results are presented in tables and graphs. Statistical analysis of the experimental results was performed using a confusion matrix. Finally, a comparative analysis of the accuracy of emotion estimation in the (F_0, σ^2) and (F_0, T) planes was conducted based on statistical data.

ESTIMATING THE SPEAKER'S EMOTIONAL STATE BY STATISTICAL ANALYSIS OF THE FUNDAMENTAL FREQUENCY

Zoran Milivojević, Bojan Prlinčević, Dijana Kostić