
ESTIMATION OF EMOTIONS NORMAL/ANXIETY BY FUNDAMENTAL FREQUENCY TRAJECTORY ANALYSIS

Bojan Prlinčević

Kosovo and Metohija Academy of Applied Studies, Leposavić, Serbia,
bojan.prlincevic@akademijakm.edu.rs

Zoran Milivojević

Academy of Technical and Educational Studies, Niš, Serbia,
zoran.milivojevic@akademijanis.edu.rs

Vesna Simović

Kosovo and Metohija Academy of Applied Studies, Leposavić, Serbia,
vesna.simovic@akademijakm.edu.rs

Abstract: In this paper, the emotional state (normal state and anxiety state) of the speaker was detected using speech analysis. In the first part of the paper, an algorithm for evaluating the emotion from the speech is described, which is based on the analysis of the trajectory fundamental frequency F_0 . At first, the algorithm for determining the criteria, i.e. the decision line, in planes is described (F_0, σ^2) and (F_0, T). In the first phase of the training, based on the test signals, the decision criterion for the detection of the emotion (Normal/Anxiety) was defined. The second part of the paper describes an experiment in which a statistical assessment of the accuracy of classifications of the speaker's emotion was performed. In the paper, the confusion matrix was used to evaluate the effectiveness of the emotion assessment (TP, TN, FP, FN and ACC). In order to achieve greater efficiency in the assessment of emotion, the confusion matrix was applied to the results obtained by applying the logical operation on the results obtained in both fields of decision-making. Based on the detection precision parameters, a comparative analysis was performed with the results obtained for the detection of the emotion Normal/Anger. The experimental results are presented tabularly and graphically.

Keywords: fundamental frequency, emotion, emotional state, confusion matrix

PROCENA EMOCIJE NORMALNA/ANKSIOZNA ANALIZOM TRAJEKTORIJE FUNDAMENTALNE FREKVENCIJE

Bojan Prlinčević

Kosovo and Metohija Academy of Applied Studies, Leposavić, Serbia,
bojan.prlincevic@akademijakm.edu.rs

Zoran Milivojević

Academy of Technical and Educational Studies, Niš, Serbia,
zoran.milivojevic@akademijanis.edu.rs

Vesna Simović

Kosovo and Metohija Academy of Applied Studies, Leposavić, Serbia,
vesna.simovic@akademijakm.edu.rs

Sažetak: U ovom radu izvršena je detekcija emocija (normalna i anksiozna) govornika primenom analize govora. U prvom delu rada opisan je algoritam za procenu emocije iz govora koji je baziran na anlizi trajektorije fundamentalne frekvencije F_0 . Najpre je opisan algoritam za određivanje kriterijuma, odnosno linije odlučivanja, u ravnima (F_0, σ^2) i (F_0, T). U prvoj fazi obuke na osnovu testnih signala definisan je kriterijum odluke za detekciju emocije (Normalna/Anksiozna). U drugom delu rada je opisan eksperiment u kojem je izvršena statistička procena preciznosti klasifikacija emocija govornika. U radu je za procenu efikasnosti procene emocije korišćena matrica konfuzije (TP, TN, FP, FN i ACC). U cilju veće efikasnosti procene emocije, matrica konfuzije je primenjena na rezultate dobijene primenom logičke operacije "i" nad rezultatima dobijenim procenom u obe ravni odlučivanja. Na osnovu parametara preciznosti detekcije izvršena je komparativna analiza sa rezultatima dobijenim za detekciju emocija Normalna/Ljutnja. Eksperimentalni rezultati su prikazani tabelarno i grafički.

Ključne reči: Fundamentalna frekvencija, emocija, emocionalno stanje, matrica konfuzije

1. 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 inteligenciji 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 (Bezooijen, 1984), (Tolkmitt & Scherer, 1986). U periodu 1990-tih godina primenjeni su napredni algoritmi kojima su vršene procene akustičkih karakteristika govora (Cairns & Hansen, 1994), (Womack & Hansen, 1996). Tokom 2000-tih godina fokus istraživanja je na pronalaženju klasifikatora emocija kojima se unapređuje efikasnost primene algoritama u aplikacijama u svakodnevnoj prameni. U naučnoj literaturi opisan je veći broj metoda i klasifikatora govornih emocija, kao što su: HMM (engl. Hidden Markov model) (El Ayadi et al., 2011), GMM (engl. The Gaussian mixture model) (El Ayadi et al., 2011), (Wanare & Dandare, 2014), SVM (engl. Support vector machine) (El Ayadi et al., 2011), (Zuber & Vidhya, 2022), (Xu et al., 2013), NB (engl. Naive Bayes classifiers) (Kyung Hak Hyun et al.), (Wang, 2015), KNN (engl. K-nearest Neighbours approach) (T, 2022), (Srinivas et al., 2014) i ANN (engl. Artificial Neural Network) (Abdul-Hadi & Waleed, 2020).

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 frekvencija F_0 povećava za emocije koje predstavljaju uzbuđenje (bes, strah, sreća...) a opada za emocije koje predstavljaju stanje manje uzbuđenosti (tuga, melanholija, dosada...) (Akçay & Oğuz, 2020). 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 (Sharma, 2021), (Chul Min Lee & Narayanan, 2005), terapeuti u medicini/psihologiji (France et al., 2000) 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.

U ovom radu je izvršena procena emocije na bazi trajektorije fundamentalne frekvencije F_0 . Kreiran je algoritam za procenu emocije 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 (Milivojević et al., 2023). Procena fundamentalne frekvencije, F_0 , govornog signala, odnosno trajektorije fundamentalne frekvencije $F_0(n)$, realizovana je primenom softverskog paketa Praat (Praat). 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 emocije. U sekciji III je opisan Eksperiment i prikazani su rezultati. Sekcija IV je Zaključak.

2. ALGORITMI PROCENE EMOCIONALNOG STANJA

U cilju procene emocije na osnovu govora korišćena je baza "Berlin database of emotional speech" (emo DB) nad kojom su primenjivana dva algoritma. Prvi algoritam je Algoritam za određivanje linije odlučivanja, a drugi algoritam je Algoritam za procenu emocije.

Algoritam za određivanje linije odlučivanja (Algoritam 1)

Algoritam za određivanje linije odlučivanja realizovan je u sledećim koracima:

Ulaz: \mathbf{x}_k -audio signal, $k = 1, 2, \dots, K$, 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 (Praat).

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 $T(k) = T_k$.

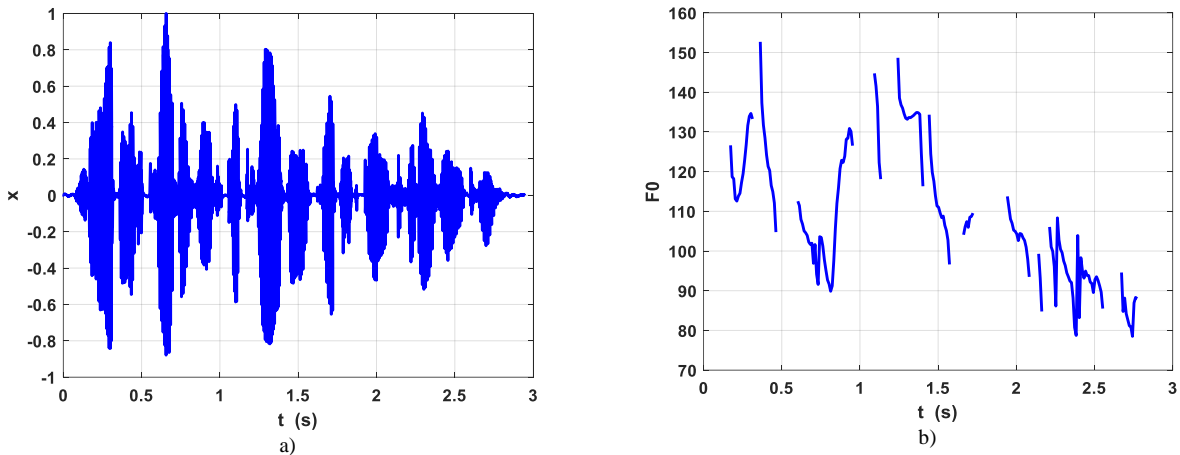
END k

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(F0,σ2), 'linear')`.

Korak 6: Izračunavanje koeficijenata ($a_{F_0,T}, b_{F_0,T}$) linije odlučivanja u ravni $P(F_0, T)$ primenom Matlab funkcije [$a_{F_0,T}, b_{F_0,T}$] = `classify(trainingData(F0,T), 'linear')`.

Primer audio signala \mathbf{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. su prikazane je ravni $P(F_0, \sigma^2)$ i $P(F_0, T)$, gde je prikazan primer prikaza linija odlučivanja i pozicije emocionalnih stanja nekih govornika iz testne baze (Korak 5 i Korak 6).

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



Algoritam za procenu emocionalnog stanja (Algoritam 2)

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

Ulaz: \mathbf{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 \mathbf{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 \mathbf{x} .

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

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

$E_1 = \text{'Anksiozan'}$

ELSE

$E_2 = \text{'Neutralan'}$

END

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

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

$E_1 = \text{'Anksiozan'}$

ELSE

$E_2 = \text{'Neutralan'}$

END

3. EKSPERIMENTALNI REZULTATI I ANALIZA

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 (emo DB), b) kreirane su linije odlučivanja u $P(F_0, \sigma^2)$ i $P(F_0, T)$ ravnima primenom Algoritma 1;

c) izvršeno je testiranje preciznosti procene emocionalnog stanja govornika primenom Algoritma 2; i d) izvršena je statistička analiza preciznosti procene emocije govornika korišćenjem matrice konfuzije.

MATRICA KONFUZIJE

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

Matrica konfuzije prikazana je u Tabeli I. Oznake u tabeli su: P – broj rečenica sa emocionalnim stanjem „Anksiozan“, N – broj rečenica sa emocionalnim stanjem „Neutralno“, PP – broj detektovanih stanja „Anksiozan“ i PN – broj detektovanih stanja „Neutralno“. Tačnost verifikacije je izračunata na osnovu parametara na sledeći način: TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative). Statistički parametri za komparaciju preciznosti algoritma 2: a) TPR (True positive rate), b) TNR (True negative rate), (Positive predictive value), NPV (Negative predictive value), i ACC (Accuracy).

Test baza

Test baza kreirana je od dela rečenica iz “Berlin database of emotional speech“ (emo DB). Uzete su rečenice koje se odnose na dva emocionalna stanja, i to emocionalno stanje „Neutralno“ i „Anksiozno/Strah“. Dve trećine izgovorenih rezultata je iskorišćeno za treniranje algoritma a 1/3 za proveru preciznosti procene emocije.

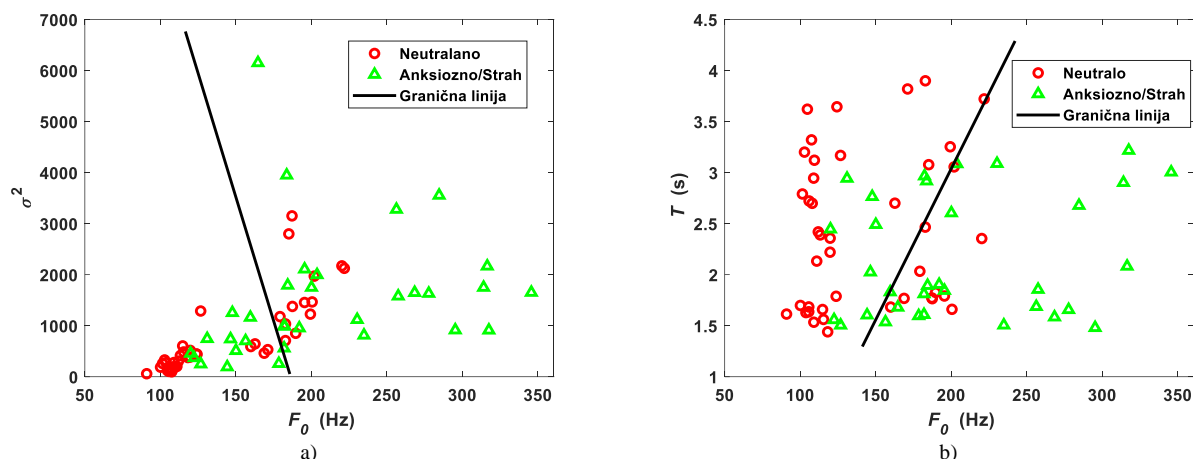
Razultati

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. 2.a prikazana je ravan $P(F_0, \sigma^2)$ i linija odlučivanja, a na sl. 2.b prikazana je ravan $P(F_0, T)$ i linija odlučivanja za analizirane emocije. Statistički rezultati, elementi matrice konfuzije prikazani su u Tabeli II. Srednje vreme trajanja govornog signala za test rečenice iz baze: a) $\bar{T}_N = 2.394$ s (emocija „Neutral“) i b) $\bar{T}_A = 2.114$ s (emocija „Anksiozan“).

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

Emocija	Ravan P		
	$P(F_0, \sigma^2)$	$P(F_0, T)$	$P(F_0, \sigma^2) \& P(F_0, T)$
P	14	14	14
N	18	18	18
TP	7	7	7
FP	4	4	5
TN	14	14	13
FN	7	7	7
PP	11	11	12
PN	21	21	20
TPR	0.5	0.5	0.5
TNR	0.778	0.778	0.722
PPV	0.636	0.636	0.583
NPV	0.667	0.667	0.650
ACC	0.6563	0.6563	0.6250

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



Analiza rezultata

Na osnovu srednjih vremena trajanja izgovorenih rečenica zaključuje se da je trajanje izgovorenih rečenica pod emocijom „Anksiozan“ kraće od trajanja izgovorenih rečenica pod emocijom „Neutralan“ $\overline{T}_A / \overline{T}_N = 2.114 / 2.394 = 0.88$ puta.

Na osnovu rezultata prikazanih u Tabeli II zaključuje se da su u ravni $P(F_0, \sigma^2)$ u odnosu na ravan $P(F_0, T)$, ispravna klasifikacija i neispravna klasifikacija emocije jednake : a) „Anksiozna“, TP b) „Neutralna“, TN c) „Anksiozna“, FP i d) „Neutralna“, FN. Primenom logičke operacije "i" za rezultate analizirane u dvema ravnima zaključuje se da u odnosu na rezultate procene dobijene samo analizom fundamentalne frekvencije u pojedinačnim ravnima ima veći broj lažno detektovanih emocija "Anksiozna" FP = 5, a manji broj tačno detektovanih emocija "Normalna" TN = 13. Broj tačno detektovanih emocija "Anksiozna" je veća PP = 12, a broj tačno detektovanih emocija PN = 20 je manja, primenom logičke operacije "i" nad dobijenim rezultatima.

Rezultati dobijeni primenom matrice konfuzije pokazuju da su vrednosti klasifikacije u ravni $P(F_0, \sigma^2)$ u odnosu na ravan $P(F_0, T)$, jednake: a) „Anksiozna“, TPR, b) „Neutralna“, TNR, kao i prediktivne vrednosti emocije: a) „Anksiozna“, PPV, i b) „Neutralna“, NPV. Tačnost procene emocija je jednaka za procenu emocija u ravni $P(F_0, \sigma^2)$ i $P(F_0, T)$, ACC = 0.6563. Primenom logičke operacije "i" za rezultate analizirane u dvema ravnima zaključuje se da u odnosu na rezultate procene dobijene samo analizom fundamentalne frekvencije u pojedinačnim ravnima ima veći broj lažno detektovanih emocija "Anksiozna" FP = 5, a manji broj tačno detektovanih emocija "Normalna" TN = 13, atačnost procene je smanjena na vrednost ACC = 0.625.

4. CONCLUSIONS

Analiza emocija govornika dobijena je primenom algoritma za određivanje linije odlučivanja (Algoritam 1) i algoritma za procenu emocija (Algoritam 2). Analizom dobijenih rezultata može se zaključiti da je srednje vreme trajanja govora izgovoreno emocijom „Anksiozna“ u odnosu na emociju „Neutralna“ 0,88 puta kraće. Analizom rezultata, dobijenih iz matrice konfuzije, prikazanih u Tabeli II, zaključuje se da primena Algoritam 2 u ravnima $P(F_0, \sigma^2)$ i $P(F_0, T)$ daje iste istu preciznost za rečenice izgovorene sa obe analizirane emocije. Primenom logičke operacije "i" nad rezultatima dobijenim u ravnima $P(F_0, \sigma^2)$ i $P(F_0, T)$ zaključuje se da se povećava broj lažno pogođenih emocija "Anksiozna" a smanjuje se broj tačno pogodjenih emocija "Neutralna". Može se zaključiti da se primenom logičke opreacije "i" smanjuje preciznost pogođenih emocija.

LITERATURA

Abdul-Hadi, M. H., & Waleed, J. (2020). Human speech and facial emotion recognition technique using SVM. 2020 International Conference on Computer Science and Software Engineering (CSASE).

<https://doi.org/10.1109/csase48920.2020.9142065>.

Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116, 56–76.

<https://doi.org/10.1016/j.specom.2019.12.001>.

Bezooijen, R. (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Walter de Gruyter.

- Cairns, D. A., & Hansen, J. H. (1994). Nonlinear analysis and classification of speech under stressed conditions. *The Journal of the Acoustical Society of America*, 96(6), 3392–3400. <https://doi.org/10.1121/1.410601>.
- Chul Min Lee, & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, 13(2), 293–303. <https://doi.org/10.1109/tsa.2004.838534>.
- El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3), 572–587. <https://doi.org/10.1016/j.patcog.2010.09.020>.
- France, D. J., Shiavi, R. G., Silverman, S., Silverman, M., & Wilkes, M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Biomedical Engineering*, 47(7), 829–837. <https://doi.org/10.1109/10.846676>.
- Kyung Hak Hyun, Eun Ho Kim, & Yoon Keun Kwak. (n.d.). Improvement of emotion recognition by bayesian classifier using non-zero-pitch concept. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. <https://doi.org/10.1109/roman.2005.1513797>.
- Milivojević, Z. N., Prlinčević, B. P., & Kostić, D. (2023). Procena emocionalnog stanja govornika statističkom analizom fundamentalne frekvencije. *2023 22st International Symposium INFOTEH-JAHORINA (INFOTEH). Phonetic reduction and elaboration in emotional speech (emo DB)*. EmoDB - Database of Emotional Speech | Chair of Speech Communication. (n.d.). <https://www.tu.berlin/en/kw/research/projects/emotional-speech>.
- Praat. Softonic. (n.d.). <https://praat.en.softonic.com/>.
- Sharma, S. (2021). Emotion recognition from speech using artificial neural networks and recurrent neural networks. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. <https://doi.org/10.1109/confluence51648.2021.9377192>.
- Srinivas, V., Santhirani, Ch., & Madhu, T. (2014). Neural network based classification for speaker identification. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 7(1), 109–120. <https://doi.org/10.14257/ijsp.2014.7.1.11>.
- T, T. (2022). Automatic speech recognition system using hybrid hidden Markov model and Human Emotion Recognition System. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4021329>.
- Tolkmitt, F. J., & Scherer, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 302–313. <https://doi.org/10.1037/0096-1523.12.3.302>.
- Wanare, A. P., & Dandare, S. N. (2014). Human Emotion Recognition From Speech. *Journal of Engineering Research and Applications*, 4(7), 74–78.
- Wang, K.-C. (2015). Time-frequency feature representation using multi-resolution texture analysis and acoustic activity detector for real-life speech emotion recognition. *Sensors*, 15(1), 1458–1478. <https://doi.org/10.3390/s150101458>.
- Womack, B. D., & Hansen, J. H. L. (1996). Classification of speech under stress using target driven features. *Speech Communication*, 20(1–2), 131–150. [https://doi.org/10.1016/s0167-6393\(96\)00049-0](https://doi.org/10.1016/s0167-6393(96)00049-0).
- Xu, S., Liu, Y., & Liu, X. (2013). Speaker recognition and speech emotion recognition based on GMM. *Proceedings of the 3rd International Conference on Electric and Electronics*. <https://doi.org/10.2991/eeic-13.2013.102>.
- Zuber, S., & Vidhya, K. (2022). Detection and analysis of emotion recognition from speech signals using decision tree and comparing with support vector machine. *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES)*. <https://doi.org/10.1109/icses55317.2022.9914046>.