ICEMIT
1st International Scientific Conference on Economy, Management and Information Technologies – ICEMIT 2023

Performance analysis of an algorithm for estimating emotions based on the trajectory of the fundamental frequency

Bojan Prlinčević[a*], Zoran Milivojević[b], Dijana Kostić[c]

[a] Kosovo and Metohija Academy of Applied Studies, Leposavić, Serbia
[b] Academy of Applied Technical and Preschool Studies Niš, Serbia
[c] Šargan inženjering d.o.o. Niš, Serbia

Article info

Original scientific paper

DOI:
https://doi.org/10.46793/ICEMIT23.287P

UDC/ UDK:
612.78:159.942

Abstract

In the presented study, an algorithm is analized to esttimate speakers' emotions using speech analysis techniques. The primary focus of the algorithm involves analyzing the trajectory of the fundamental frequency $F_0(t)$ to accurately determine a range of emotional states. This investigation includes a comprehensive analysis conducted within planes (F_0, σ^2) and (F_0, T). During the initial training phase, a clear decision criterion is established to differentiate between emotional states. This criterion is defined based on the analysis of test signals and is positioned within the designated planes. The performance evaluation of the algorithm is executed during the testing phase, utilizing a confusion matrix. This evaluation allows for a precise assessment of the algorithm's capability to detect emotional states. Furthermore, a comparative study is undertaken, comparing outcomes related to the identification of various emotional states such as Normal/Anger, Normal/Boredom, and Normal/Anxiety. To provide a comprehensive presentation of the algorithm's effectiveness in identifying emotional states, the results are presented in tabulles and graphs.

Keywords: *Fundamental frequency, emotional state, confusion matrix*

1. Introduction

During the era of artificial intelligence advancement, a well-established pursuit involves the development of devices capable of discerning emotions conveyed through speech. While humans can effortlessly recognize emotions in verbal and visual communication, enabling computers and AI-based devices to achieve this requires emulating human perceptual mechanisms. The quest for recognizing emotions in speech began in the 1980s with initial research grounded in statistical analysis of acoustic attributes (Bezooijen, 1984), (Tolkmitt & Scherer, 1986). Progressing into the 1990s, more sophisticated algorithms emerged, enabling better estimation of acoustic features in speech (Cairns & Hansen, 1994), (Womack & Hansen, 1999).

The 2000s marked a shift in focus among researchers toward identifying emotion classifiers that could enhance the practical application of algorithms in everyday contexts. The scientific literature documented numerous methods and classifiers for the analysis of emotional content in speech. Some of these methods include HMM (Hidden Markov Model) (Ayadi et al., 2011), GMM (Gaussian Mixture Model) (Ayadi et al., 2011), (Wanare & Dandare, 2014), SVM (Support Vector Machine) (Ayadi et al., 2011), (Pan et al., 2012), (Seehapoch & Wongthanavasu, 2013), NB (Naive Bayes classifiers) (Pan et al., 2012), (Wang, 2015), KNN (K-nearest Neighbors approach) (Nwe et al., 2023), (Srinivas et al., 2014), and ANN (Artificial Neural Network) (Panda et al., 2012).

A significant number of algorithms for recognizing emotions conveyed in speech heavily rely on the notable impact of emotional states on speech patterns. Notably, a consistent finding from various studies is that the average fundamental frequency (F_0) tends to rise for emotions associated with excitement, like anger, fear, and happiness, while it decreases

*Corresponding author
E-mail address: bojan.prlincevic@akademijakm.edu.rs

for emotions linked to lower levels of excitement, such as sadness, melancholy, and boredom (Ververidis & Kotropoulos, 2006).

In today's era of extensive human-computer interaction, emotional speech recognition presents a formidable challenge. Researchers are predominantly focused on creating practical applications for real-world scenarios, including call centers, medical or psychological therapy settings, and others (Lee & Narayanan, 2005), (France et al., 2000). For effective implementation of emotion recognition algorithms in speech, the development of emotional speech databases becomes paramount for algorithm training. These databases encompass diverse utterances from different speakers, facilitating accurate algorithm performance evaluation. Currently, emotional speech databases exist for various languages. These databases involve actors simulating the pronunciation of specific sentences while conveying diverse emotional states. The conveyed emotion's authenticity is assessed through methods such as the Mean Opinion Score (MOS) test, and sentences with unsatisfactory ratings are excluded.

In this particular paper, the authors employ the trajectory of the fundamental frequency (F_0) to estimating emotional states. The emotion assessment algorithm consisting of two main components: a) creation of a training database and establishment of an emotional state evaluation decision boundary in the (F_0, σ^2) and (F_0, T) planes, and b) testing the efficacy of the emotion estimation algorithm. The software package Praat (Praat, 2023) is utilized to evaluate the fundamental frequency (F_0) of the speech signal, specifically its trajectory F_0 (n). The testing algorithm employs the built-in Matlab function "classify(trainingData)." The algorithm's results are validated using a confusion matrix, a tool assessing classification algorithm accuracy. In previous works, the authors analyzed the assessment of the emotions Normal/Angry, Normal/Anxiety and Normal/Boredom in particular, so in this paper a comparative analysis of the estimation of emotions was performed. This study strives to develop an effective algorithm for recognizing emotions in speech, with the trajectory of the fundamental frequency (F_0) as a pivotal parameter. By crafting emotional speech databases and meticulously assessing the expressed emotions in sentences, the researchers aim to enhance emotion recognition algorithm performance, thereby facilitating practical deployment in diverse real-world scenarios.

The paper is organization as follows: Introduction: This section provides an introduction to the topic of emotions recognition and the importance of the research. Algorithm for Determining the Decision Line: Section 2 elaborates on the algorithm developed to determine the decision line for evaluating the emotions based on the trajectory of the fundamental frequency (F_0) in the planes (F_0, σ^2) and (F_0, T). Algorithm for Evaluating the Emotions: This section explains the algorithm used for estimating the emotions in speech. It involves processing the speech signal and extracting features related to the fundamental frequency (F_0) trajectory to estimate the emotions. Experiment and Results: Section 3 describes the experiment, including the creation of the emotional speech database and the testing procedure. The obtained results are then presented and analyzed, showcasing the performance of the emotion estimation algorithm. Conclusion: In Section 4, the authors summarize their results, discuss the implications of the results, and potentially suggest areas for future research in emotion's recognition.

2. Algorithms for emotion estimation

In this study, the authors utilized the "Berlin database of emotional speech" (Milivojević et al., 2023) to assess emotions through speech analysis. They employed two distinct algorithms to analyze the emotions in the speech samples from this database.

Algorithm for Establishing the Decision Line:
The first algorithm's purpose was to determine a decision line crucial for estimation emotion's based on the trajectory of the fundamental frequency (F_0). This algorithm involved identifying a boundary or threshold within the (F_0, σ^2) and (F_0, T) planes. This boundary is used in classifying various emotional states.
Algorithm for Estimating Emotion's:

The second algorithm was used to assess the emotion's conveyed within the speech samples. This entailed processing the speech signals extracted from the "Berlin database of emotional speech" (Milivojević et al., 2023). Relevant features, such as the trajectory of the fundamental frequency (F_0), were extracted from the signals. These features then are used to estimate the emotions.

By applying these two algorithms to the "Berlin database of emotional speech" (Milivojević et al., 2023), the researchers aimed to enhance the precision and efficiency of emotion recognition in speech. The outcomes derived from these experiments were used to verify the efficiency of the proposed algorithms and their potential practicality in real-world settings, such as call centers, medical therapy environments, or psychological analyses.

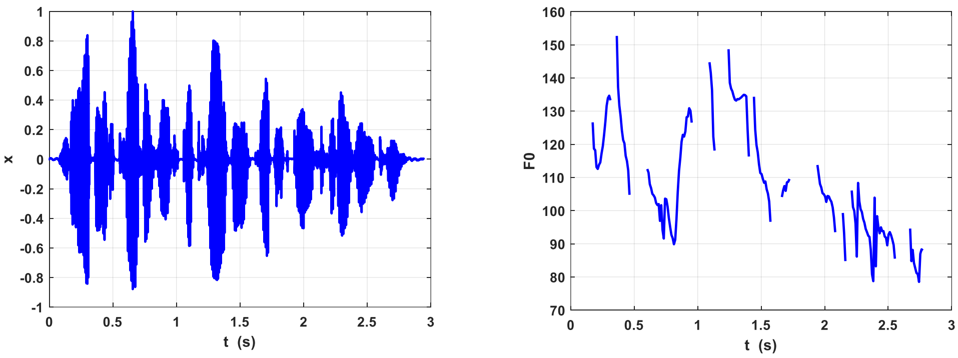
2.1. Algorithm for determining the decision line (Algorithm 1)

The algorithm for determining the decision line, which is crucial for evaluating the emotions based on the trajectory of the fundamental frequency (F_0), was implemented through the following steps:
Input: $\mathbf{x}_k$-audio signal, k = 1, 2, …, K, total number of signals for testing.
Output: ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$) - coefficients of the decision line in the plane $P(F_0, \sigma^2)$, ($a_{F_0,T}$, $b_{F_0,T}$) - coefficients of the decision line in the plane $P(F_0, T)$.
FOR k = 1 : K
Step 1: Creating a fundamental frequency trajector, F_{0k}, of signals $\mathbf{x}_k$, by applying the Praat (Praat, 2023).
Step 2: Determining the mean value of the fundamental frequency trajectory $\overline{F_{0k}}$, and array generation $F_0(k) = \overline{F_{0k}}$.
Step 3: Determining the variance of the fundamental frequency trajectory σ_k^2, and array generation $\sigma^2(k) = \sigma_k^2$.
Step 4: Determining the duration T_k of the audio signal $\mathbf{x}_k$, and array generation $T(k) = T_k$.
END *k*
Step 5: Calculation the coefficients ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$) of decision lines in the plane $P(F_0, \sigma^2)$, using the Matlab function [$a_{F_0,\sigma}$, $b_{F_0,\sigma}$] = classify(trainingData(F_0, σ^2), 'lin ear').
Step 6: Calculation the coefficients ($a_{F_0,T}$, $b_{F_0,T}$) of decision lines in the plane $P(F_0, T)$ using the Matlab function [$a_{F_0,T}$, $b_{F_0,T}$] = classify(trainingData(F_0,T), 'linear').

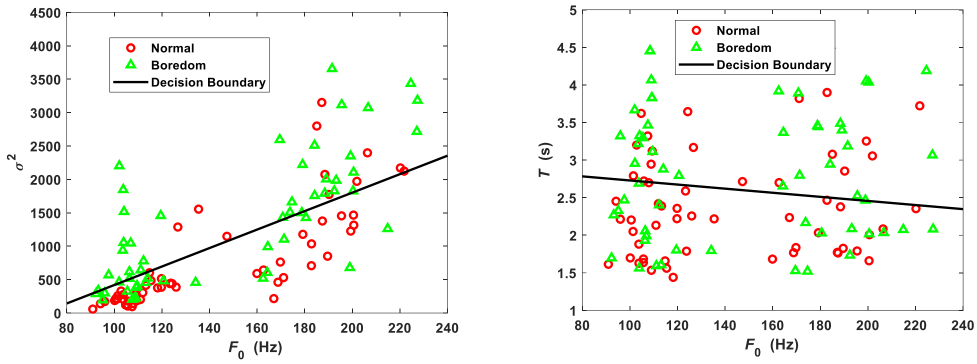
In the realm of speech-based emotion recognition, the researchers conducted an experiment centered an audio signal termed as "x." This audio signal represents speech sourced from a speaker within the test database. The trajectory of the fundamental frequency (F_0) extracted from this audio signal is depicted in Figure 1b (Step 1). This trajectory illustrates the dynamic fluctuations of the fundamental frequency throughout the speech duration.

In Figure 2a (Step 5), an visual representation is showed, featuring a two-dimensional plane denoted as $P(F_0, \sigma^2)$. Here, F_0 signifies the fundamental frequency, while σ^2 signifies the variance or standard deviation of the fundamental frequency. At this plane, a perceptible decision line is evident, serving as a delineation between distinct emotional states. Notably, emotional states of select speakers from the test database are indicated as data points within this plane. Each data point corresponds to a specific emotion, and its placement is determined by the respective F_0 and σ^2 values corresponding to that particular speech segment. Furthermore, Figure 2b (Step 6) introduces another two-dimensional plane labeled as $P(F_0, T)$, with T symbolizing an additional feature related to duration time. Similar to Figure 2a, this plane also incorporates a decision line designed to segregate diverse emotional states. The emotional state positions for specific speakers from the test database are denoted as data points on this plane. These data points shed light on the emotional characteristics embedded within the respective speech excerpts.

Importantly, it's emphasized that Figures 1, and 2 serve the purpose of illustrating the algorithm's efficacy in emotion recognition within speech. This visual presentation a clearer represent of how the fundamental frequency and other pertinent features collaboratively contribute to precise emotion classification, and are presented to enhance the understanding of the algorithm's functioning and its role in recognizing emotions in speech.



a) b)
Figire 1. Audio signal: a) time shape, b) fundamental frequency trajectory.



a) b)

Figire 2. Decision line and position of emotions in: a) $P(F_0, \sigma^2)$ plane, and b) $P(F_0, T)$ plane.

2.2. Algorithm for estimating emotions (Algorithm 2)

The algorithm for estimating emotions from speech was implemented through a series of defined steps, which are as follows:

Input: **x**-audio signal, ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$) - boundary decision line coefficients in the plane $P(F_0, \sigma^2)$, ($a_{F_0,T}$, $b_{F_0,T}$) - boundary decision line coefficients in the plane $P(F_0, T)$.

Output: E_1, E_2 - emotions.

Step 1: Creating the trajectory of the fundamental frequency F_0 of the signal **x**.

Step 2: Determining the mean value of the trajectory of the fundamental frequency $\overline{F_0}$.

Step 3: Determining the variance of the fundamental frequency trajectory σ^2.

Step 4: Determining the duration T of the audio signal **x**.

Step 5: Classification of emotion in the $P(F_0, \sigma^2)$ plane:

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

E_1 = 'Emotion 1'

ELSE

E_2 = 'Emotion 2'

END

Step 6: Classification of emotion in the $P(F_0, T)$ plane:

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

E_1 = 'Emotion 1'

ELSE

E_2 = 'Emotion 2'

END.

3. Experimental results and analysis

3.1. Experiment

To assess the efficiancy of emotion recognition algorithms from speech signals, an experiment was conducted, through the following phases:

a) Database Formation: The initial step involved creating a database of speech signals aimed at training the emotion assessment algorithm. This database was compiled using emotional speech excerpts derived from the "Berlin database of emotional speech" (Milivojević et al., 2023).

b) Decision Boundary Establishment: Decision boundaries were established in two planes, namely $P(F_0, \sigma^2)$ and $P(F_0, T)$, utilizing Algorithm 1. These decision lines were crucial for demarcating emotional states, formed by the fundamental frequency F_0 within the planes $P(F_0, \sigma^2)$ and $P(F_0, T)$.

c) Emotions Estimation: Algorithm 2 was then implemented to estimate the emotions expressed by speakers. This algorithm processed the speech signals, extricated features F_0, σ^2, and T, and utilized them to predict the emotional states of each speaker present in the test database.

d) Accuracy Verification: The accuracy of emotion estimation was tested employing Algorithm 2. This phase encompassed the application of the trained algorithm to the test database, constituting the remaining 40% of the speech samples. The goal was to ascertain the algorithm's capability to accurately classify the emotions conveyed by speakers.

e) Statistical Insight: To comprehensively evaluate the results, a statistical approach was employed, specifically involving the utilization of a confusion matrix. This matrix allowed for an comparison between the predicted emotions estimated by the algorithm and the actual emotions present within the test database.

f) Formulating Emotion-Type Criteria: Criteria for determining emotion types (e.g., "Neutral", "Anxiety", "Angry" and "Boredom") were established. This was achieved by segregating sentences from the test database into two groups. The first group was utilized to define emotion-type decision criteria via Algorithm 1. The decision lines within the planes $P(F_0, \sigma^2)$ and $P(F_0, T)$ were employed as the basis for categorizing emotions. The 60% of the sentences were randomly chosen to form the training data for Algorithm 1.
g) Efficiency of the Estimation: The second group of sentences, constituting 40% of the test database, was employed to assess the efficiency and accuracy of emotion classification using Algorithm 2.
h) Comparative Statistical Analisys: A comparative statistical analysis was conducted on the outcomes obtained from Algorithm 2. This involved using the confusion matrix to evaluate the precision of emotion classification and to gauge the overall performance of the algorithm.

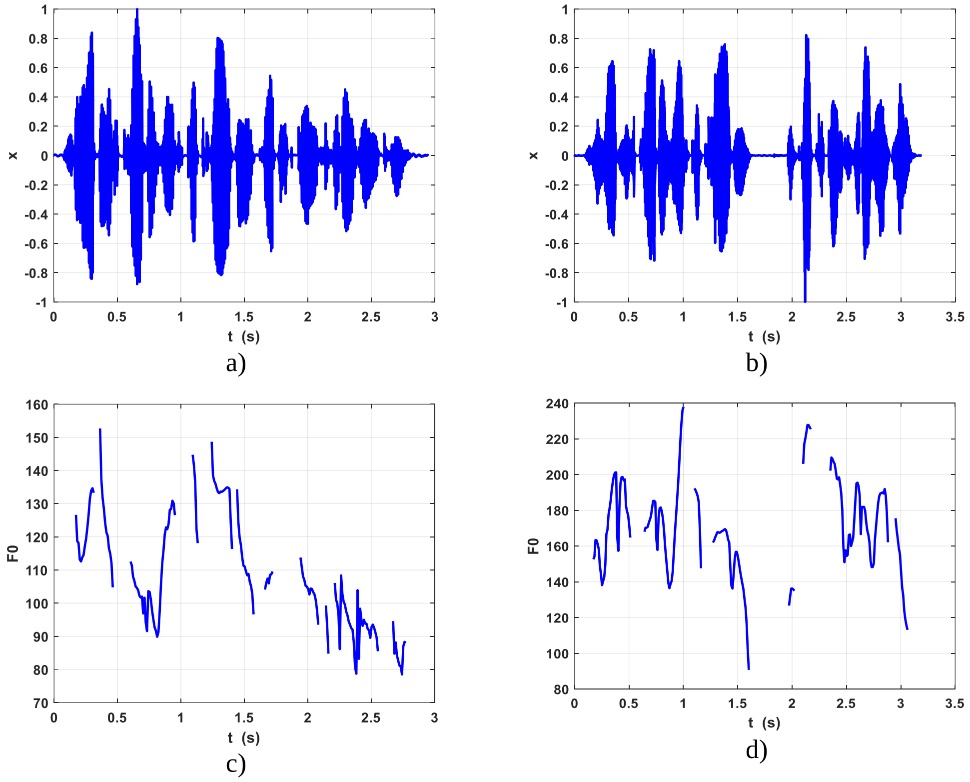
Through conduced experiments in previous works (Milivojević et al., 2023), (Prlinčević et al., 2023), (Prlinčević et al., 2023), the author aimed to substantiate the precision and efficiency of the emotion recognition algorithms. In this paper authors where comparative analisys of the results obtained in papers (Milivojević et al., 2023), (Prlinčević et al., 2023), (Prlinčević et al., 2023).

Table 1. Confusion matrix

		DETECTED EMOTION	
	TOTAL P + N	POSITIVE (EMOTION 1) (PP)	NEUTRAL (EMOTION 2) (PN)
REAL	POSITIVE (EMOTION 1) (P)	TRUE POSITIVE (EMOTION 1) (TP)	FALSE NEUTRAL (EMOTION 2) (FN)
REAL	NEUTRAL (EMOTION 2) (N)	FALSE POSITIVE (EMOTION 1) (FP)	TRUE NEUTRAL (EMOTION 2) (TN)

The experiments involved evaluating the accuracy of emotions recognition using a confusion matrix, as shown in Table 1. The labels in the table represent different categories: P denotes the number of sentences with the emotional state "Emotion 1", N represents the number of sentences with the emotional state "Emotion 2", PP indicates the number of detected states classified as "Emotion 1", and PN represents the number of detected states classified as "Emotion 2."
Statistical parameters were calculated for comparing the accuracy of Algorithm 2:
a) True Positive Rate (TPR): Also known as sensitivity or recall, it represents the proportion of correctly detected emotions labeled as "Emotion 1". It is calculated as TPR = TP / (TP + FN).
b) True Negative Rate (TNR): Also known as specificity, it represents the proportion of correctly detected emotions labeled as "Emotion 2". It is calculated as TNR = TN / (TN + FP).
c) Positive Predictive Value (PPV): Also known as precision, it measures the predictive value of emotions classified as "Emotion 1". It is calculated as PPV = TP / (TP + FP).
d) Negative Predictive Value (NPV): It measures the predictive value of emotions classified as "Emotion 2". It is calculated as NPV = TN / (TN + FN).
e) Accuracy (ACC): The overall accuracy of emotion assessment, representing the proportion of correctly classified emotions (both "Emotion 1" and "Emotion 2") over the total number of emotions tested. It is calculated as ACC = (TP + TN) / (TP + TN + FP + FN).
Using these calculated statistical parameters and analyzing the confusion matrix, the researchers gained valuable insights into the accuracy and performance of Algorithm 2 in recognizing emotions in speech. These measures allow for a comprehensive evaluation of the algorithm's ability to correctly identify emotions and serve as a basis for comparison with other emotion recognition methods.
3.2. Test base
The test database was created from a subset of sentences taken from the "Berlin database of emotional speech" (Milivojević et al., 2023). Specifically, sentences related to emotional states were selected: "Normal", Anxiety", Angry", and "Boredom." An example of one of this is illustrated in Figure 3, which shows the time-domain waveform and the trajectory of the fundamental frequency $F_0(t)$ for a male speaker uttering the sentence "Sie haben es gerade hochgetragen unf jetz gehen sie wieder runter" (English translation: "They just carried it up and now they're going down again") with the following emotional states:
a) Boredom (Figure 3.a): This figure shows the time-domain waveform and the trajectory of the fundamental frequency $F_0(t)$ for the same sentence when expressed with the emotional state of anger.
b) Normal (Figure 3.b): Similarly, this figure shows the time-domain waveform and the trajectory of the fundamental frequency $F_0(t)$ for the same sentence when expressed with a neutral emotional state.

By comparing the waveforms and the trajectories of the fundamental frequency (F_0) for the same sentence under different emotional states, the researchers aimed to examine how emotions influence speech patterns and how these patterns can be utilized for emotion recognition.



Figire 3. Display of the audio signal spoken by a 31-year-old man: a) "Normal" speech, and b) "Boredom" speech; and presentation of the trajectory of the fundamental frequency of the spoken sentence for: c) "Normal" speech, d) "Boredom" speech.

4. Results

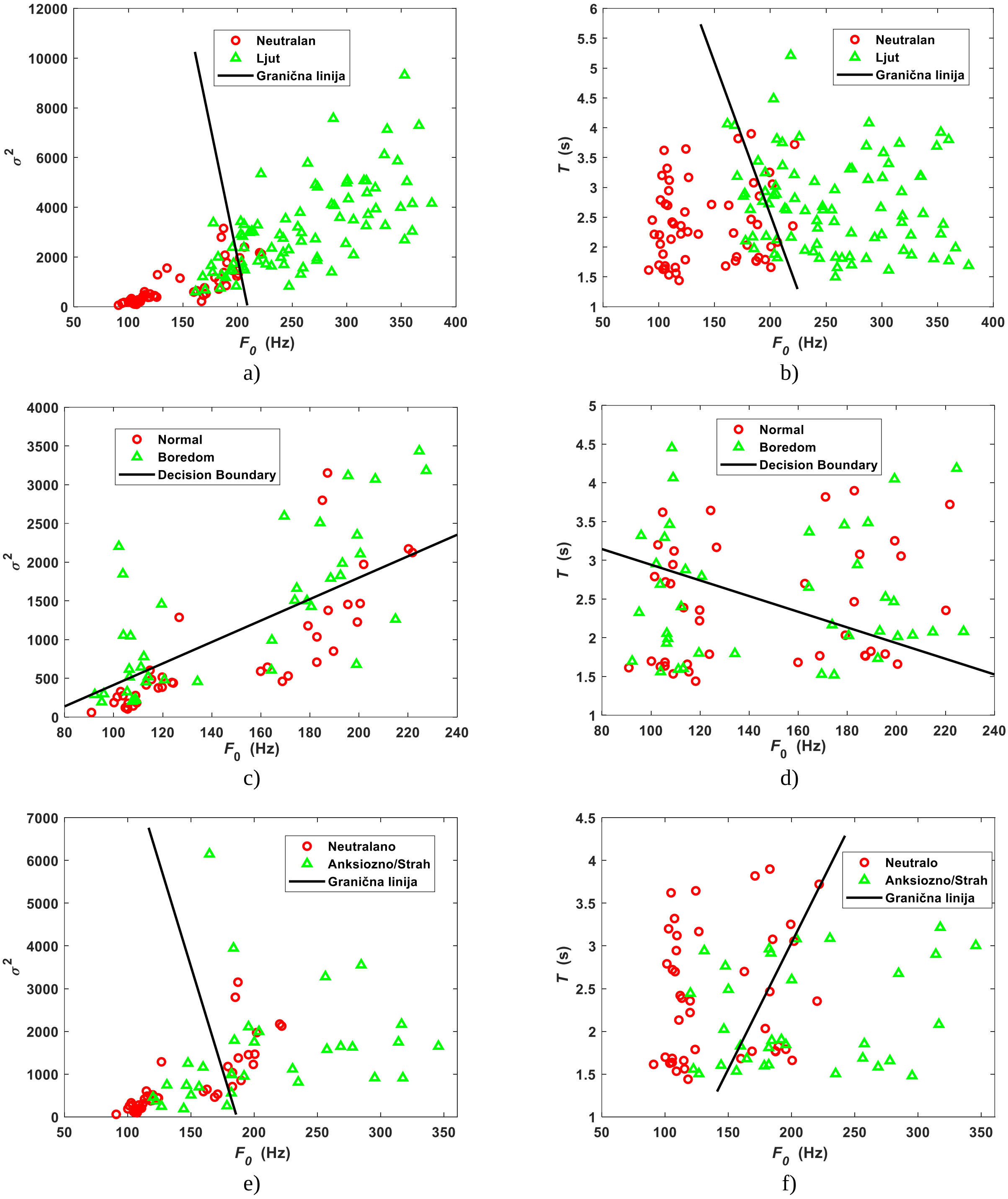
Using the Algorithm 1, the coefficients of the decision line were determined for two planes: a) $P(F_0, \sigma^2)$ (represented as ($a_{F_0,\sigma}$, $b_{F_0,\sigma}$)), Figure 4.(a,c,e), and b) $P(F_0, T)$ (represented as ($a_{F_0,T}$, $b_{F_0,T}$)), Figure 4.(b,d,f).

Statistical results and elements of the confusion matrix are presented in Table 2, showcasing the performance of the emotion recognition algorithm.

The mean duration of the speech signal for the test sentence from the base is provided as follows: a) for the "Normal" emotion, the mean duration is $\bar{t}$ = 2.37 seconds; b) for the "Boredom" emotion, the mean duration is $\bar{t}$ = 2.592 seconds, c) for the "Anxiety" emotionmean duration is $\bar{t}$ = 2.114 s, and d) for. The the "Anger" emotion duration is $\bar{t}$ = 2.711s.

Table 2. Results of confusion matrix in planes $P(F_0, \sigma^2)$ and $P(F_0, T)$

Emotion	Plane *P*: Neutral/Anger		Neutral/Boredom			Neutral/Anxiety		
	$P(F_0, \sigma^2)$	$P(F_0, T)$	$P(F_0, \sigma^2)$	$P(F_0, T)$	$P(F_0, \sigma^2)$ & $P(F_0, T)$	$P(F_0, \sigma^2)$	$P(F_0, T)$	$P(F_0, \sigma^2)$ & $P(F_0, T)$
P	31	31	14	14	14	14	14	14
N	21	21	18	18	18	18	18	18
TP	27	28	7	7	7	10	11	8
FP	1	2	4	4	5	13	6	17
TN	20	19	14	14	13	5	12	1
FN	4	3	7	7	7	4	3	6
PP	28	30	11	11	12	23	17	25
PN	24	22	21	21	20	9	15	7
TPR	0.87097	0.9032	0.5	0.5	0.5	0.7143	0.7857	0.5714
TNR	0.95238	0.9048	0.778	0.778	0.722	0.2778	0.6667	0.0556
PPV	0.96429	0.9333	0.636	0.636	0.583	0.4348	0.6471	0.3200
NPV	0.83333	0.8636	0.667	0.667	0.650	0.5556	0.8000	0.1429



Figire 4. Positioning of emotions and decision lines of the base of emotions in: a,c,e) $P(F_0, \sigma^2)$ plane, and b,d,f) $P(F_0, T)$ plane.

4.1. Analysis of the results

Based on the analysis of the average duration of spoken sentences, it is observed that sentences spoken under the emotion "Anger" and "Boredom" have a longer duration compared to sentences spoken under the emotion "Normal, and sentences spoken under emotion "Anxiety" have lower duration compared to sentence spoken under emotion "Normal". The duration of "Anger" sentences is approximately 1.14 times greater than that of "Normal" sentences (2.711 s / 2.37 s = 1.14 times), the duration of "Boredom" sentences is approximately 1.094 times greater than that of "Normal" sentences (2.592 s / 2.37 s = 1.094 times), the duration of "Anxiety" sentences is approximately 0.89 times greater than that of "Normal" sentences (2.114 s / 2.37 s = 0.89 times).

Analisyng the results presented in Table 2, conclusions are drawn regarding the correct and incorrect classification rate of emotions in the plane $P(F_0, \sigma^2)$ according to the classification on the plane $P(F_0, T)$:

a) correctly detected emotions: The number of times "Anger" is correctly classified is greater, with a ratio of TP_{F_0,s^2} / $TP_{F_0,T}$ = 27 / 28 = 0.964, times. The number of times "Normal" is correctly classified is also greater, with a ratio of TN_{F_0,s^2} / $TN_{F_0,T}$ = 20 / 19 = 1.05, times.
b) correctly detected emotions: The number of times "Boredom" is correctly classified is greater, with a ratio TP_{F_0,s^2} / $TP_{F_0,T}$ = 7 / 7 = 1, times. The number of times "Normal" is correctly classified is also greater, with a ratio of TN_{F_0,s^2} / $TN_{F_0,T}$ = 14 / 14 = 1, times.
c) correctly detected emotions: The number of times "Anxiety" is correctly classified is greater, with a ratio TP_{F_0,s^2} / $TP_{F_0,T}$ = 10 / 11 = 0.91, times. The number of times "Normal" is correctly classified is also greater, with a ratio of TN_{F_0,s^2} / $TN_{F_0,T}$ = 5 / 12 = 0.42, times.
d) incorrectly detected emotions: The number of times "Anger" is incorrectly classified is greater, FP_{F_0,s^2} / $FP_{F_0,T}$ = 1 / = 0.5, times. The number of times "Normal" is incorrectly classified is also greater, with a ratio FN_{F_0,s^2} / $FN_{F_0,T}$ = 4 / 3 = 1.33 times.
e) incorrectly detected emotions: The number of times "Boredom" is incorrectly classified is greater, FP_{F_0,s^2} / $FP_{F_0,T}$ = 4 / 4 = 1.0, times. The number of times "Normal" is incorrectly classified is also greater, with a ratio FN_{F_0,s^2} / $FN_{F_0,T}$ = 7 / 7 = 1.0 times.
f) incorrectly detected emotions: The number of times "Anxiety" is incorrectly classified is greater, FP_{F_0,s^2} / $FP_{F_0,T}$ = 13 / 6 = 2.17, times. The number of times "Normal" is incorrectly classified is also greater, with a ratio FN_{F_0,s^2} / $FN_{F_0,T}$ = 4 / 3 = 1.33 times.
By using the logical operator "and", it is concluded that the rate of correctly detected emotions are not identical in both planes.
The results obtained from the confusion matrix for true positive and true negative rate (TPR and TNR) demonstrate that the classification values in the plane $P(F_0, T)$ compared to $P(F_0, \sigma^2)$ are higher for correctly detected emotions: a) "Anger," the ratio of correct classification is $TPR_{F_0,T}$ / TPR_{F_0,s^2} = 0.9032 / 0.87097 = 1.037 times; b) "Boredom," the ratio of correct classification is $TPR_{F_0,T}$ / TPR_{F_0,s^2} = 0.5 / 0.5 = 1 times; c) "Anxiety," the ratio of correct classification is $TPR_{F_0,T}$ / TPR_{F_0,s^2} = 0,7857 / 0,7143 = 1.099 times.

The results obtained from the confusion matrix for predictive value (PPV and NPV) demonstrate that the classification values in the plane $P(F_0, T)$ compared to $P(F_0, \sigma^2)$ are higher for correctly detected emotions: a) "Anger," the ratio of correct classification is PPV_{F_0,s^2} / $PPV_{F_0,T}$ = 0.96429 / 0.9333 = 1.033 times; b) "Boredom," the ratio of correct classification is PPV_{F_0,s^2} / $PPV_{F_0,T}$ = 0.636 / 0.636 = 1,488 times; c) "Anxiety," the ratio of correct classification is PPV_{F_0,s^2} / $PPV_{F_0,T}$ = 0.4348 / 0.6471 = 0.672 times.

4. Conclusion

The study employed two algorithms, Algorithm 1 and Algorithm 2, to analyze the emotions of the speaker. After analyzing the results, the authors drew several conclusions: Average Duration: The duration of speech associated with the emotion "Anger" was found to be approximately 1.14 times longer than that of the emotion "Normal" , emotion "Boredom" was found to be approximately 1.094 times longer than that of the emotion "Normal", "Boredom" was found to be approximately 0.89 times longer than that of the emotion "Normal". This suggests that the duration of spoken sentences can vary depending on the emotional state expressed. Prediction Value in the plane $P(F_0, T)$: the Algorithm 2 demonstrated a higher predictive value for the both analysed emotions. For emotion "Normal" after using logical operator "and" prediction value is signicifly decreasing for emotion "Boredom", and for emotion "Anxiety"have aproximatly same value, as well as a value equal to the value in the plane where this value is lower. By analyzing these results, the authors gained valuable insights into the relationships between emotions, speech characteristics, and the performance of the emotion recognition algorithms. The study emphasizes the significance of considering various factors, such as duration and acoustic features, in accurately assessing emotional states from speech signals.

References

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
Bezooijen, R. V. (1984). *Characteristics and recognizability of vocal expressions of emotion*. 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

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

Hyun, K. H., Kim, E. H., & Kwak, Y. K. (2005, August). Improvement of emotion recognition by Bayesian classifier using non-zero-pitch concept. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (pp. 312-316). IEEE. https://doi.org/10.1109/ROMAN.2005.1513797

Lee, C. M., & 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

Milivojević, Z. N., Prlinčević, B. P., & Kostić, D. (2023). Procena emocionalnog stanja govornika statističkom analizom fundamentalne frekvencije. In *2023 22st International Symposium INFOTEH-JAHORINA (INFOTEH)*.

Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. *Speech communication*, *41*(4), 603-623. https://doi.org/10.1016/S0167-6393(03)00099-2

Seehapoch, T., & Wongthanavasu, S. (2013, January). Speech emotion recognition using support vector machines. In *2013 5th international conference on Knowledge and smart technology (KST)* (pp. 86-91). IEEE. https://doi.org/10.1109/KST.2013.6512793

Panda, B., Padhi, D., Dash, K., & Mohanty, S. (2012). Use of SVM classifier & MFCC in speech emotion recognition system. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(3), 225-230.

Praat (2023). *Praat*. Softonic. https://praat.en.softonic.com/

Prlinčević, B. P., Milivojević, Z. N., Simović, V., & Kostić, D. (2023). *Estimation of emotional Normal/Boredom state by fundamental frequency trajectory analysis*. http://www.fmns.swu.bg/en/index.html

Prlinčević, B., Milivojević, Z., & Simović, V. (2023). Estimation of emotions normal/anxiety by fundamental frequency trajectory analysis. *KNOWLEDGE-International Journal*, *58*(3), 495-500.

Seehapoch, T., & Wongthanavasu, S. (2013, January). Speech emotion recognition using support vector machines. In *2013 5th international conference on Knowledge and smart technology (KST)* (pp. 86-91). IEEE. https://doi.org/10.1109/KST.2013.6512793

Srinivas, V., & Madhu, T. (2014). Neural network-based classification for speaker identification. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *7*(1), 109-120.

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. https://psycnet.apa.org/doi/10.1037/0096-1523.12.3.302

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech communication*, *48*(9), 1162-1181. https://doi.org/10.1016/j.specom.2006.04.003

Wanare, M. A. P., & Dandare, S. N. (2014). Human emotion recognition from speech. Int. 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. (1999). 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