

AN IMPROVED APPROACH TO ASSESSING EMOTIONS FROM SPEECH

Bojan Prlinčević¹, Zoran Milivojević², Violeta Stojanović³, Dijana Kostić⁴

Abstract: In this paper, we conducted an analysis of speaker emotions through speech evaluation. The Introduction section introduces an algorithm designed to assess emotional speech, which is grounded in the examination of the fundamental frequency path (F_0). This assessment employs a decision line derived from the first derivative of the frequency, specifically utilizing a positive frequency increment. Initially, an algorithm that establishes criteria for detecting emotions in plane ($F_0, F_0^{(1)}$) is described. The decision criterion distinguishing between Normal and Anxiety emotions was defined during the preliminary training phase based on the testing emotions from Base. The next section of the paper presents an experiment focusing on the statistical evaluation of accuracy concerning speaker emotion classification. To evaluate the efficacy of emotional estimations, a confusion matrix was utilized. The results from these experiments are presented both in tabular and graphical formats.

Key words: emotion assessment, fundamental frequency, first derivative, emotion, confusion matrix

1. INTRODUCTION

As artificial intelligence continues to evolve, there has been a rapid development of machines that can understand speech and interpret human emotion. Humans naturally infer emotions in verbal or visual communications, but enabling computers to do so requires them to simulate the complex processes that accompany human perceptions. In fact, the very inception of emotion recognition from speech happened in the 1980s, with the first studies primarily being statistical analyses of acoustic features [1]. With the continued maturation of the field in the 1990s, researchers developed more complex algorithms that greatly improved the bottom-up understanding of speech's acoustic attributes [2]. By the 2000s, the emphasis was on developing emotion classifiers, systems that had the potential to integrate lab-based and real-world environments. Various approaches have been proposed to investigate emotional content in speech, such as, but not limited to, Hidden Markov Models (HMM) [3], Gaussian Mixture Models (GMM) [3], [4], Support Vector Machines (SVM) [3], Naive Bayes classifiers (NB) [5], K-nearest Neighbors (KNN) [6], and Artificial Neural Networks (ANN) [7]. One of the most important functionalities of emotion recognition lies in how emotional states significantly have an effect on the features of speech. Studies have shown that average fundamental frequency (F_0) in speech relates positively to higher emotional arousal (e.g., anger, fear, and happiness) and negatively to lower levels of emotional arousal, such as sadness and anxiety. This insight is important, as we explore the nuances of human-computer interaction in our tech-minded world.

Most researchers have started to focus on real-world applications of AI in a variety of fields, including call centers, the medical workplace, and behavioral therapy. Speech databases for emotions are established for emotion recognition algorithms as a crucial step. These databases have a variety of utterances spoken by various speakers and can be used to objectively assess how well the algorithms perform. Nowadays, there are emotional speech databases available in several languages (e.g., German, Chinese, Taiwanese, etc.), where recordings are usually done by actors who realistically express different emotions. This research aims to use the trajectory of the fundamental frequency (F_0) for emotion measurement [8], [9]. In particular, the authors introduced an emotion estimation algorithm based on two key components: (a) the creation of a training database and decision boundary in the ($F_0, F_0^{(1)}$) plane for emotional state assessment, (b) the determination of the effectiveness of the algorithm. The software package used for the analysis of the fundamental frequency (F_0) of speech signals was Praat, and the F_0 (n) was extracted [10]. The built-in Matlab function

¹dr Bojan Prlinčević, Kosovo and metohija academy of applied studies, bojan.prlincevic@akademijakm.edu.rs

²dr Zoran Milivojević, MB University, zoran.milivojevic@akademijanjs.edu.rs

³mr Violeta Stojanović, The Academy of applied technical and preschool studies, violeta.stojanovic@akademijanjs.edu.rs

⁴Dijana Kostić, Šargan d.o.o., dijanaaricija79@gmail.com

"classify(trainingData)" was used to test the algorithm, and the results were validated by means of the confusion matrix — an essential metric used to evaluate the performance of a classification algorithm. Thus, aiming at a improving algorithm for speech emotion recognition, this paper has studied the extracted F_0 as one parameter and, making the derivative of the F_0 trajectory as the second parameter. This keeps the linear decision line established as the classification boundary and is shown in contrast to previous studies which used linear decision lines on the plane (F_0, σ^2) [11]. This development can potentially improve emotional intelligence, which will enhance the interaction between humans and AI.

2. ALGORITHMS FOR ESTIMATIONS OF EMOTIONS

The "Berlin database of emotional speech" [12] was used in this paper for the identification of the speech emotions. Using this database, we applied two different algorithms to detect the emotions in the speech samples.

Algorithm to Find the Decision Line:

The linear decision line, which is a key differentiation in evaluating the emotional state in this paper with respect to the F_0 trajectory, was found with the first algorithm. This algorithm uses the boundary/threshold of the plane $(F_0, F_0^{(1)})$ to determine classification either in terms of positive/accurate or negative/missed emotion.

The second Algorithm, for estimating the emotions, is developed to estimate emotions expressed in the speech samples. For realisation of this Algorithm includes processing the speech signals of the "Berlin database of emotional speech" [12], from which is extracted nessesery features like the fundamental frequency (F_0) trajectory from the speech signals and its first derivate. It intends to improve the performance of emotion detection in speech with a high, accurate, and effective method by implementing these two algorithms on the "Berlin database of emotional speech" [12].

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

The algorithm for determining the decision line, was implemented through the following steps:

Input: x_k -audio signal, $k = 1, 2, \dots, K$, total number of signal for testing.

Output: $(a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}})$ - coefficients of the decision line in the plane $P(F_0, F_0^{(1)})$.

FOR $k = 1 : K$

Step 1: Creating a fundamental frequency trajector, F_{0k} , of signals x_k , by applying the Praat [12].

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 firs derivative of the fundamental frequency trajectory $F_0^{(1)}$, and array generation $F_0^{(1)}(k) = F_{0-k}^{(1)}$.

END k

Step 5: Calculation the coefficients $(a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}})$ of decision lines in the plane $P(F_0, F_0^{(1)})$ using the Matlab function $[a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}}] = \text{classify}(\text{trainingData}(F_0, F_0^{(1)}), 'linear')$.

In the context of emotion recognition in speech, the researchers conducted an experiment using an audio signal labeled as " x ". This audio signal represents speech from one of the speakers in the test database. The trajectory of the fundamental frequency (F_0) extracted from this audio signal is shown in Figure 1b. (*Step 1*). This trajectory depicts how the fundamental frequency varies over time during the speech.

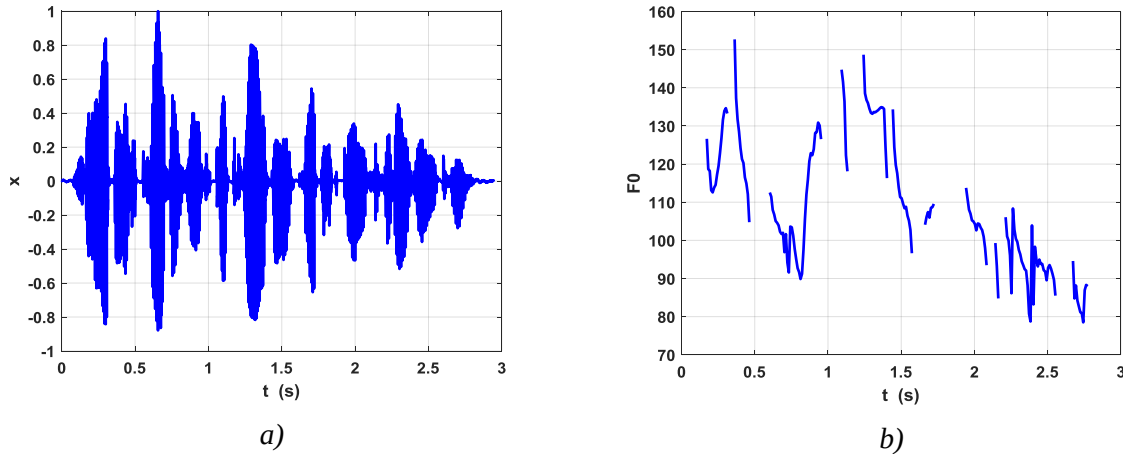


Figure 1. - Audio signal: a) time shape, b) fundamental frequency trajectory.

2.2. Algorithm for estimating emotions (Algorithm 1)

The algorithm for estimating emotions from speech was implemented through follows steps:

Input: $\mathbf{x}$ -audio signal, $(a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}})$ - boundary decision line coefficients in the plane $P(F_0, F_0^{(1)})$.

Output: E_1, E_2 - emotions.

Step 1: Creating the trajectory of the fundamental frequency F_0 of the signal $\mathbf{x}$.

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

Step 3: Determining the variance of the first derivative of frequency trajectory F_0 .

Step 4: Classification of emotion in the $P(F_0, F_0^{(1)})$ plane:

IF $\overline{F_0} \leq (F_0^{(1)} - b_{F_0, F_0^{(1)}}) / a_{F_0, F_0^{(1)}}$

$E_1 = \text{'Anxiety'}$

ELSE

$E_2 = \text{'Normal'}$

END

END.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Experiment

To evaluate the effectiveness of algorithms for emotion recognition from speech signals, an experiment was conducted, comprising the following steps:

a) Database Creation: A database of speech signals was created to train the emotion assessment algorithm. This database includes emotional speech samples from the "Berlin database of emotional speech" [12].

b) Decision Line Creation (linear): Decision lines were established in the plane: $P(F_0, F_0^{(1)})$ using Algorithm 1. These decision lines act as boundaries to separate emotions based on the combination of fundamental frequency F_0 in the planes $P(F_0, F_0^{(1)})$.

c) Emotions Assessment: Algorithm 2 was employed to assess the emotions of speakers. The algorithm processes the speech signals, extracts features F_0 , depending of $F_0^{(1)}$, and uses them to predict the emotions of each speaker in the test database.

d) Accuracy Testing: The accuracy of the emotions assessment was tested using Algorithm 2. This step involves applying the trained algorithm to the test database (the remaining 40% of sentences) and determining how accurately it can classify speakers' emotions.

e) Statistical Analysis: To analyze the results, a statistical approach was utilized, specifically the confusion matrix.

f) Defining Emotion-Type Decision Criteria: To establish criteria for deciding the type of emotion (e.g., "Neutral" and "Anxiety"), sentences from the test database were divided into two groups. The first group of sentences was used to define emotion-type decision criteria through Algorithm 1. The decision line in the planes $P(F_0, F_0^{(1)})$ was used as the criterion for categorizing emotions. 60% of the sentences were randomly selected and formed the training data for Algorithm 1.

g) Efficiency Testing: The second group of sentences (40% of the test database) was used to test the efficiency and accuracy of emotions classification using Algorithm 2.

h) Comparative Statistical Analysis: A comparative statistical analysis was applied to the results obtained from Algorithm 2, which includes the detected emotions. The confusion matrix was used in this analysis to assess the accuracy of the emotion classification and evaluate the algorithm's overall performance.

The experiment involved evaluating the accuracy of emotions recognition using a confusion matrix. The labels in confusion matrix represent different categories: P denotes the number of sentences with the emotional state "Anxiety," N represents the number of sentences with the emotional state "Normal," PP indicates the number of correctly detected states classified as "Anxiety", and PN represents the number of correctly detected states classified as "Normal."

Statistical parameters were calculated for comparing the accuracy of Algorithm 2: a) True Positive Rate (TPR); b) True Negative Rate (TNR).

3.2. Experiment

Sentences from the "Berlin database of emotional speech" [12] were used. In particular, two emotional states, "Normal" and "Anxiety" were selected, and sentence represent states were taken. For example, the time-domain waveform and the F0(t) trajectory for a male speaker saying the sentence "Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter" (English translation: "They just carried it up, and now they're going down again") through the following emotional states. The researchers also evaluated the obtained waveforms and trajectories of the F0 during various emotional states of the same sentence to investigate the relation between emotions and the way of speech, and establish a relationship that could be used for emotion recognition. Such examples illustrate the differences between pairs of emotions, "Anxiety" and "Normal," in speech characteristics for a female speaker voicing a particular speech utterance.

4. RESULTS

Applying the Algorithm 1, the coefficients of the linear decision line were determined for plane: a) $P(F_0, F_0^{(1)})$ (represented as $(a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}})$). In Figure 2.a, the plane $P(F_0, F_0^{(1)})$ is visualized along with the training Algorithm and obtained decision line based on the coefficients $(a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}})$, in Figure 2.b, the plane $P(F_0, F_0^{(1)})$ are visualized results obtained from algorithm for estimation along

with the decision line based on the coefficients $(a_{F_0, F_0^{(1)}}, b_{F_0, F_0^{(1)}})$. Statistical results and elements of the confusion matrix are presented in Table 1, showcasing the performance of the emotion recognition algorithm for planes $P(F_0, F_0^{(1)})$ and $P(F_0, \sigma^2)$ in order to compare results.

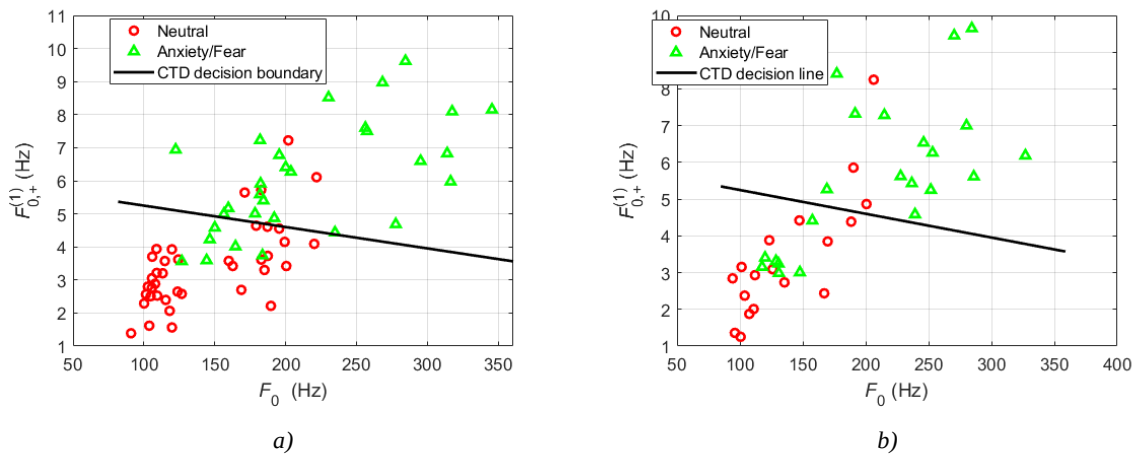


Figure 2. - Positioning of emotions and decision lines of the base of emotions in $P(F_0, F_0^{(1)})$ plane: a) training Algorithm, and b) estimatig Algorithm.

Table 1 - Results of confusion matrix in planes $P(F_0, F_0^{(1)})$ and $P(F_0, \sigma^2)$ for linear decision line.

		P	N	TP	FP	TN	FN	PP	PN	TPR	TNR
Plane	$P(F_0, F_0^{(1)})$	22	18	15	3	15	7	18	22	0.6818	0.8333
	$P(F_0, \sigma^2)$	22	18	14	3	15	8	17	23	0.6364	0.8333

The presented statistical analysis, together with the decision lines, offers valuable insights into the effectiveness of the algorithm using the first derivative of frequency in recognizing emotions from speech signals, which they intend to continue their research.

4.1 Analysis of the results

Examining the results presented in Table 1, conclusions are drawn regarding the correct and incorrect classification rate of emotions in the plane $P(F_0, F_0^{(1)})$ compared to the plane $P(F_0, \sigma^2)$:

a) correctly detected emotions: The number of times "Anxiety" is correctly classified is with a ratio, $TP_{F_0, F_0^{(1)}} / TP_{F_0, \sigma^2} = 15 / 14 = 1,07$. The number of "Normal" correctly classified emotions is with a same ratio, $TN_{F_0, F_0^{(1)}} / TN_{F_0, \sigma^2} = 15 / 15 = 1$.

b) incorrectly detected emotions: The number of times "Anxiety" is incorrectly classified is same, $FP_{F_0, F_0^{(1)}} / FP_{F_0, \sigma^2} = 3 / 3 = 1$, times. The number of "Normal" incorrectly classified emotions is $FN_{F_0, F_0^{(1)}} / FN_{F_0, \sigma^2} = 7 / 8 = 0.875$.

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 $P(F_0, F_0^{(1)})$ compared to $P(F_0, \sigma^2)$ are for correctly detected emotions: a) "Anxiety", are higher 0.6818 compared to 0.6364; b) For "Normal," the with the same ratio 0.8333.

5. CONCLUSION

The Paper conducted analysis of 2 algorithms, Algorithm 1 and Algorithm 2, which provide efficiency analysis of estimation of emotions. Classification Accuracy of the emotions for Planes $P(F_0, F_0^{(1)})$ and $P(F_0, \sigma^2)$: In contrast to the plane $P(F_0, \sigma^2)$, the overall accuracy in classifying the Neutral emotions for plane $P(F_0, F_0^{(1)})$ by Algorithm 2 was relatively higher 1,07 times higher. Also using planes $P(F_0, F_0^{(1)})$ for emotion estimating gives better results in accuracy of false negative (neutral emotional state) in comparison with plane $P(F_0, \sigma^2)$, 0.875 false estimates cases. Through the analysis of these outcomes, the authors were able to extract relations between the plane for mappings of emotional states and the efficacy of the emotion recognition algorithms.

6. REFERENCES

- [1] Tolkmitt F.J. and Scherer K.R., Effect of experimentally induced stress on vocal parameters. J. Exp. Psychol. [Hum. Percept.] 12 (3), 302–313, **1986**.
- [2] Cairns D. and Hansen J.H.L, Nonlinear analysis and detection of speech under stressed conditions. J. Acoust. Soc. Am. 96 (6), 3392–3400, **1994**.
- [3] Ayadi M.E., Kamel M.S. and Karray F., "Survey on speech emotion recognition: Features, classification schemes, and databases", Pattern Recognition, pp. 572–587, **2011**.
- [4] Wanare A.P., Dandare S.N., "Human Emotion Recognition From Speech", Int. Journal of Engineering Research and Applications, vol. 4, Issue 7, pp.74–78, July **2014**.
- [5] Wang K.C., "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**.
- [6] Srinivas V., Rani C.S. and Madhu T., "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**.
- [7] Panda B., Padhi D., Dash K., Mohanty S., "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**.
- [8] Ververidis D., Kotropoulos C. "Emotional speech recognition: Resources, features, and methods", Speech Communication, Vol. 48, pp. 1162-1181, **2006**.
- [9] Lee C.M., Narayanan S.S., "Toward detecting emotions in spoken dialogs" IEEE Trans. Speech Audio Process. 13 (2), 293–303, **2005**.
- [10] <https://praat.en.softonic.com/>.
- [11] Prlinčević B., Milivojević Z., and Simović V. "Estimation of emotions normal/anxiety by fundamental frequency trajectory analysis2023", KNOWLEDGE - International Journal, Vol. 58 (3), pp. 495-500, **2023**.
- [12] <https://www.tu.berlin/en/kw/research/projects/emotional-speech>.