

Comparative analysis of LDA, QDA, SVM, algorithms and neural networks for boredom and fear detection in speech

Bojan Prlinčević^{a*}, Zoran Milivojević^b, Milan Mišić^a, Dejan Gurešić^c, Dijana Kostić^d

^a Academy of Applied Studies Kosovo and Metohija, Leposavić, Serbia

^b Academy of Applied Technical and Preschool studies, Niš, Serbia

^c University of Priština in Kosovska Mitrovica, Faculty of Sciences and Mathematics, Serbia

^d Šargan inženjering d.o.o, Niš, Serbia

Article info

Original scientific paper

DOI:

<https://doi.org/10.71159/icemit2565P>

This is an open access paper under the
license 

Abstract

This paper presents an analysis of speech emotion recognition focusing on distinguishing between boredom and fear emotional states. The methodology employs the fundamental frequency trajectory (F_0) and its derivatives to extract characteristics. Emotional states are classified using four machine learning algorithms: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). The experiment utilizes the Berlin Emotional Speech Database for training and evaluation. Results demonstrate that all classifiers effectively separate boredom and fear emotions, with ANN achieving the highest classification accuracy. Performance analysis are derived using confusion matrix analysis confirm. This study provides an efficient framework for detecting subtle emotional differences in speech signals, particularly between boredom and fear manifestations.

Keywords: speech emotion recognition, fundamental frequency, neural networks

1. Introduction

As artificial intelligence continues to evolve, the development of systems capable of understanding speech and interpreting human emotions has progressed rapidly. While humans naturally perceive emotions through verbal and visual cues, enabling machines to do so requires simulating the complex cognitive processes underlying human perception. The field of speech emotion recognition emerged in the 1980s, with early studies focusing primarily on statistical analyses of acoustic features (Tolkmitt et al., 1986). During the 1990s, researchers developed more sophisticated algorithms that significantly advanced the bottom-up understanding of speech's acoustic properties (Cairns & Hansen, 1994). By the 2000s, the focus shifted toward developing emotion classifiers capable of bridging laboratory research and real-world applications. Various machine learning approaches have been employed to detect emotional content in speech, including Hidden Markov Models (HMM) (Ayadi et al., 2011), Gaussian Mixture Models (GMM) (Ayadi et al., 2011; Wanare & Dandare, 2014), and more recently, a range of classifiers such as Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machines (SVM), and Artificial Neural Networks (ANN).

These methods have demonstrated particular effectiveness in handling the non-linear relationships often present in emotional speech data. The importance of emotion recognition lies in how emotional states significantly influence speech characteristics. Research has shown that the average fundamental frequency (F_0) correlates positively with high-arousal emotions such as fear and excitement, and negatively with low-arousal states like boredom or sadness (Ververidis & Kotropoulos, 2006). This relationship provides a crucial foundation for exploring more nuanced human-computer interactions in our increasingly technology-driven world.

Recent research has expanded into real-world applications across diverse fields including call centers, healthcare environments, and behavioral therapy. The development of specialized emotional speech databases has been instrumental in this progress, providing standardized datasets for algorithm training and evaluation. These databases contain utterances spoken by various actors who realistically express different emotions, available in multiple languages including German,

*Corresponding author

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

Chinese, and Taiwanese (Lee & Narayanan, 2005). This research utilizes the fundamental frequency (F_0) trajectory as a primary feature for emotion measurement (Prlinčević et al., 2023; Burkhardt et al., 2005). The study employs the Praat software package for F_0 extraction (Boersma & Weenink, 2023) and implements four distinct classification approaches—LDA, QDA, SVM, and ANN—to establish decision boundaries in the feature space. Algorithm performance is evaluated using confusion matrix analysis, providing comprehensive metrics for assessing classification effectiveness. By incorporating both the F_0 parameter and its derivative, this work advances beyond traditional linear classification methods, potentially contributing to enhanced emotional intelligence in human-AI interaction systems.

2. Algorithms for estimations of emotions

The "Berlin database of emotional speech" (Boersma & Weenink, 2023) 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.

Algorithms to Find the Decision Line: The LDA, QDA, SVM, and ANN decision lines, which are 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, σ) to determine classification either in terms of positive/accurate or negative/missed emotion.

The Algorithms are used for estimating the emotions estimate emotions expressed in the speech samples from the second part of sentences from berlin data base. For realisation of precision of estimation Algorithms includes processing the speech signals of the "Berlin database of emotional speech" (Boersma & Weenink, 2023), from which is extracted necessary features like the fundamental frequency (F_0) trajectory from the speech signals and its first derivative..

2.1. Algorithm for determining the decision line

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 trajectory, F_{0k} , of signals x_k , by applying the Praat (Boersma & Weenink, 2023).

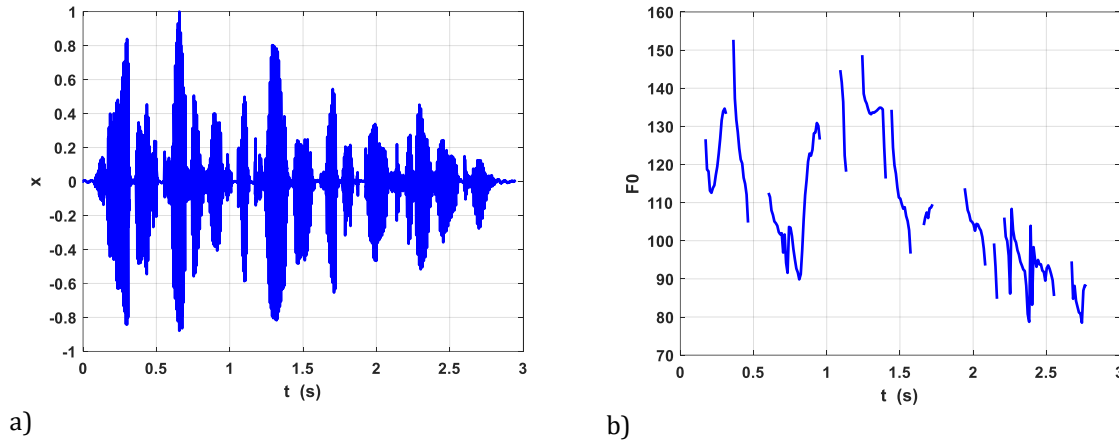
Step 2: Determining the mean value of the fundamental frequency trajectory $\overline{F_{0k}}$, and array generation $F_0(k) = \overline{F_{0k}}$.

Step 3: Determining the first derivative of the fundamental frequency trajectory $F_{0,+}^{(1)}$, and array generation $F_{0,+}^{(1)}(k) = F_{0-k}^{(1)}$.

END k

Step 4: 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 functions.

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. Algorithms for estimating emotions

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

Input: 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 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 $F_{0,+}^{(1)}$.

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 = 'Fear'

ELSE

E_2 = 'Boredom'

END

END.

3. Experimental results and analysis

An experiment was conducted in which the ETC algorithm was used for emotion detection within the (F_0, ETC) plane. To compute the decision boundary—that is, to define the boundary in the plane between emotional states—the following algorithms were employed: (a) LDA, (b) QDA, (c) SVM and (d) ANN.

The experiment was carried out in two phases: (a) Training phase (Algorithm 1) – during which the classifiers were trained and the decision boundary was created, and (b) Testing phase (Algorithm 2) – in which the trained models were applied to test speech signals.

In the experiment, a specialized version of a feedforward neural network designed for pattern recognition was created and trained. The neural network was implemented in MATLAB using the patternnet function. The architecture employed the default tangent-sigmoid (tansig) transfer function in both the hidden and output layers. The selected network structure is with architecture [2 3 2 2]. Through supervised learning, the network mapped the input feature space to a two-class output space, represented as $[Emotion1, Emotion2]^T$. For an input vector belonging to a given category, the corresponding neuron in the output layer—representing that emotion class—was expected to output a 1, while the other neuron would output a 0.

The algorithms were applied to test signals from a purpose-built emotional speech database. As a result of training and testing, evaluation parameters were generated based on confusion matrices (Confusion Parameters – CP). The efficiency of algorithms in emotion estimation is measured via the confusion matrix, a) in the training phase, and b) in the testing phase.

3.1. Speech signal database

The recording setup included a Sennheiser MKH 40 P 48 microphone and a Tascam DA-P1 portable DAT recorder. In addition to the pure audio data, electroglottograms were captured using a portable laryngograph (Laryngograph Ltd.). Recordings were made at a sampling frequency of 48 kHz and subsequently downsampled to 16 kHz.

Recordings were taken with a sampling frequency of 48 kHz and later downsampled to 16 kHz. In this paper, two emotional states—neutral and anger—were selected for analysis. Representative utterances for each state were chosen, and the analysis includes both the speech signal and the fundamental frequency contour ($F_0(t)$) for a male and a female speaker pronouncing the sentence: "Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter" (Translation on english: "They just carried it upstairs, and now they are coming down again.")

3.2. Results

Using the LDA, QDA, SVM, and ANN algorithms for creating decision lines and detecting emotions, the locations of points with coordinates (F_0, σ_i) in the plane were determined for training speech signals (Figures 2.,3.,4., and 5.): (a) training phase, and (b) testing phase. Confusion matrix parameters for all training and testing using LDA, QDA, SVM, and ANN classifiers are presented in Table 1.

Figure 2. Decision boundaryfor LDA algorithm in the (F_0, σ) plane:(a) training phase, (b) testing phase

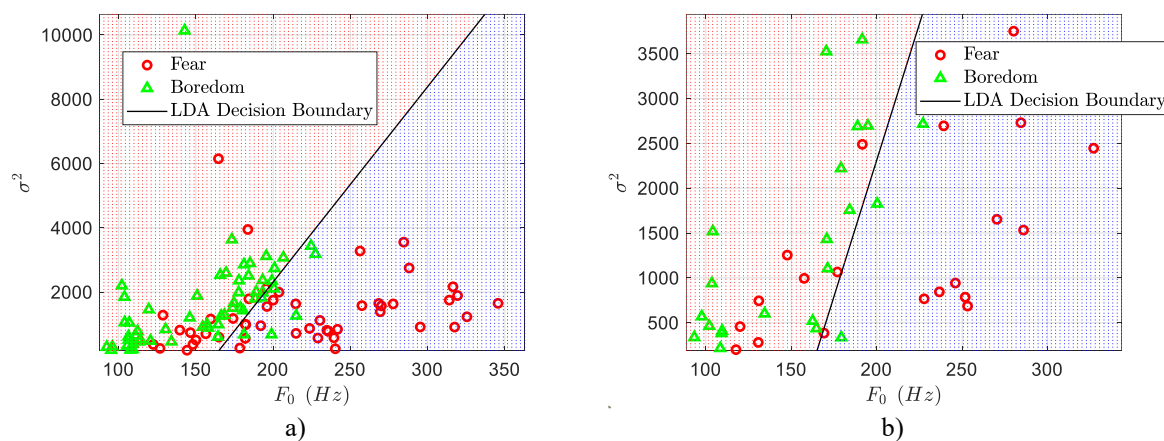


Figure 3. Decision boundaryfor QDA algorithm in the (F_0, σ) plane:(a) training phase, (b) testing phase

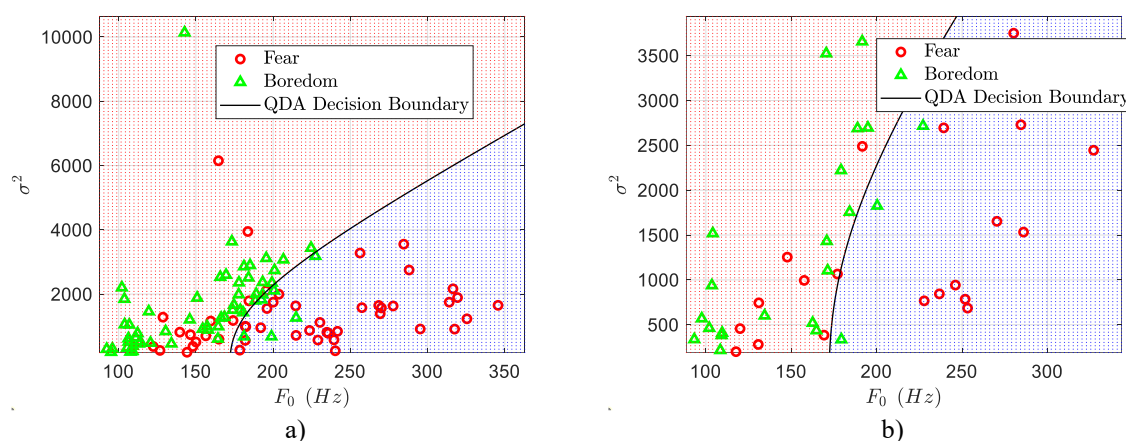
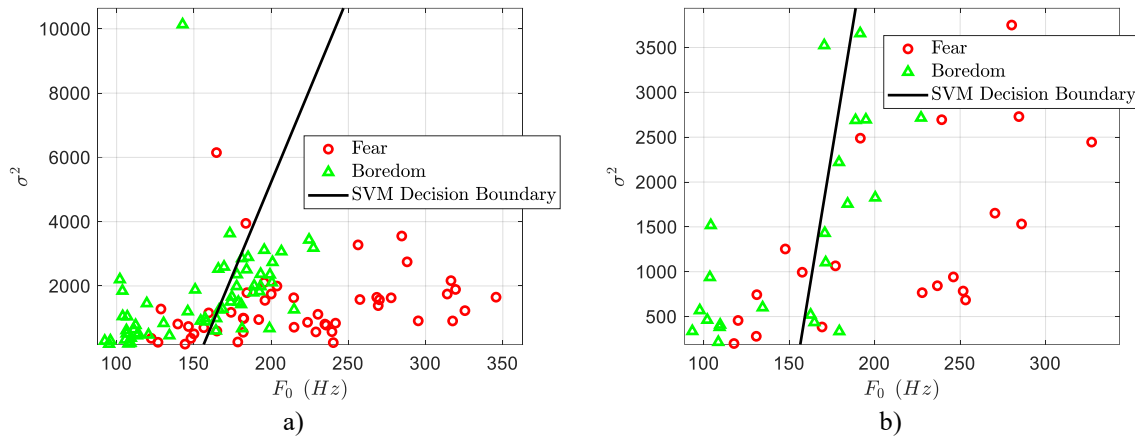
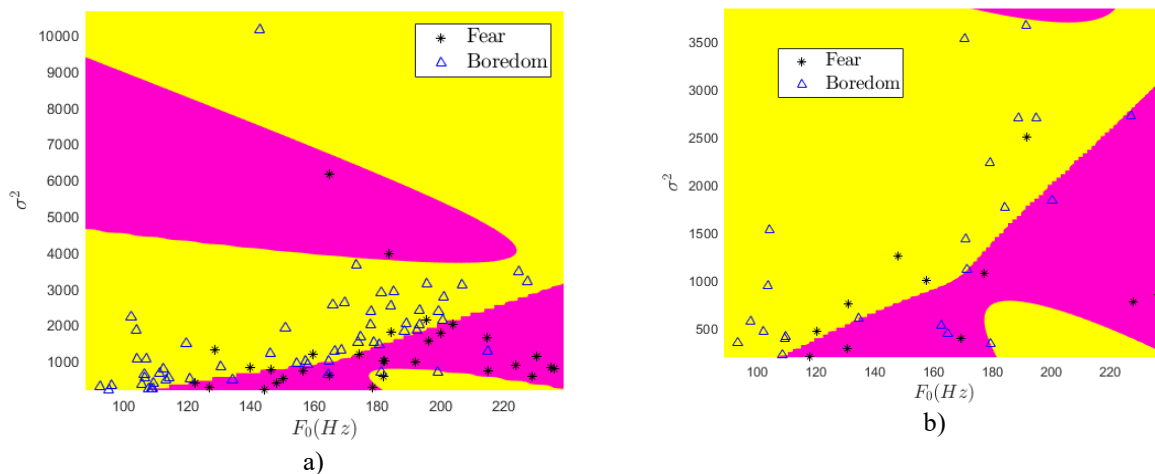


Figure 4. Decision boundary for SVM algorithm in the (F_0, σ) plane: (a) training phase, (b) testing phase.**Figure 5.** Decision boundary for ANN algorithm (Architecture [2 3 2 2]) in the (F_0, σ) plane: (a) training phase, (b) testing phase**Table 1.** Confusion matrix parameters for the (F_{0i}, σ_i) plane during training and testing.

	Training				Testing			
	LDA	QDA	SVM	ANN	LDA	QDA	SVM	ANN
P	49	49	49	49	20	20	20	20
N	59	59	59	59	22	22	22	22
TP	33	33	37	43	12	11	14	15
FP	7	6	28	3	3	3	12	5
TN	52	53	31	56	19	19	10	17
FN	16	16	12	6	8	9	6	5
PPV	0.8250	0.8462	0.5692	0.9348	0.800	0.7857	0.5385	0.750
NPV	0.7647	0.7681	0.7209	0.9032	0.7037	0.6786	0.6250	0.7727
ACC	0.7870	0.7963	0.6296	0.9167	0.731	0.7143	0.5714	0.7619
TPR	0.6735	0.6735	0.7551	0.8776	0.600	0.550	0.700	0.750
FPR	0.1186	0.1017	0.4746	0.0508	0.1364	0.1364	0.5455	0.2273
FNR	0.3265	0.3265	0.2449	0.1224	0.400	0.450	0.300	0.250
TNR	0.8814	0.8983	0.5254	0.9492	0.8636	0.8636	0.4545	0.7727

3.2. Analysis of the results

Analyzing the diagrams in Figures 2.-5., it can be seen that the decision line is correctly positioned for plane $P(F_{0i}, \sigma_i)$.

Examining the results presented in Table 1, conclusions are drawn regarding the correct and incorrect classification rate of emotions in frequency domain in the plane $P(F_{0i}, \sigma_i)$:

Based on the results presented in Table 1, a detailed performance analysis of the four algorithms (LDA, QDA, SVM, and ANN) during both training and testing phases can be conducted, with particular emphasis on their ability to discriminate between the emotions states fear and boredom.

Overall Accuracy (ACC)

- LDA achieves moderate results: ACC = 78.7% (training) and 73.1% (testing). A decrease in accuracy of ~5.6%, indicates mild overfitting.
- QDA shows similar performance: ACC = 79.63% (training) and 71.43% (testing), with a slightly decrease of ~8.2%.
- SVM yields the lowest values: ACC = 62.96% (training) and 57.14% (testing), suggesting poor generalization.
- ANN significantly outperforms the others: ACC = 91.67% (training) and 76.19% (testing). Although a decrease in accuracy of ~15.5% is observed, ANN still achieves the highest overall accuracy.

Precision (PPV) and Specificity (TNR)

- ANN demonstrates the highest precision (PPV) during both training (93.48%) and testing (75%), indicating the lowest rate of false positives for fear detection.
- The specificity (TNR) for ANN is also the highest: 94.92% (training) and 77.27% (testing), confirming its reliability in correctly excluding boredom.

Sensitivity (TPR) and False Positive Rate (FPR)

- ANN achieves the highest sensitivity (TPR) during training (87.76%) and testing (75%), reflecting its strong capability to detect true cases of fear.
- SVM shows high sensitivity (75.51% during training) but an excessively high FPR (47.46%), indicating a tendency toward over-detection of fear.

Comparison Between Phases: Training vs. Testing

All algorithms exhibit some degree of performance degradation during testing, which is expected due to dataset differences. However:

- ANN shows the smallest decline in key metrics (TPR, PPV, TNR) compared to the other algorithms.
- SVM demonstrates the poorest generalization, as evidenced by a sharp increase in FPR during testing (54.55%).

Analysis shows to as that th ANN emerges as the most effective algorithm due to:

- Achieving the highest overall accuracy (ACC) during both training and testing.
- Demonstrating the best precision (PPV) and specificity (TNR).
- Exhibiting the highest sensitivity (TPR) while maintaining a controlled FPR.

Although all algorithms show some degree of overfitting, ANN best balances model complexity and generalization capability. This aligns with contemporary research highlighting the advantage of deep models for tasks with high uncertainty, such as emotion recognition. The decrease in accuracy during testing is common but least pronounced in ANN. SVM performed the weakest, likely due to the linear inseparability of the data in the feature space.

4. Conclusion

Based on experimental results and performance analysis of different classification algorithms for recognizing fear and boredom emotions from speech, artificial neural networks (ANN) have shown superiority in accuracy, sensitivity and specificity compared to LDA, QDA and SVM. The use of the fundamental frequency (F_0) and its first derivative as input features, combined with the construction of decision boundaries in the (F_0 , σ) plane, enabled the efficient detection and separation of emotional states. Although the neural network showed some signs of overfitting, it still performed the best when tested on new data, making it a strong candidate for use in real-world emotion recognition systems. These results suggest that future improvements could come from using more complex audio features and deeper learning models.

References

- Ayadi, M. E., 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>
- Berlin Database of Emotional Speech. (2005). *Technische Universität Berlin*. <https://www.tu.berlin/en/kw/research/projects/emotional-speech>
- Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.3.09) [Computer software]. <https://www.praat.org>
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., & Weiss, B. (2005). A database of German emotional speech. In *Proceedings of Interspeech* (pp. 1517–1520).
- Cairns, D., & Hansen, J. H. L. (1994). Nonlinear analysis and detection of speech under stressed conditions. *Journal of the Acoustical Society of America*, 96(6), 3392–3400. <https://doi.org/10.1121/1.410601>
- 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>
- Prinčević, B., Milivojević, Z., & Simović, V. (2023). Estimation of emotions normal/anxiety by fundamental frequency trajectory analysis. *KNOWLEDGE – International Journal*, 58(3), 495–500.

- 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>
- 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, A. P., & Dandare, S. N. (2014). Human emotion recognition from speech. *International Journal of Engineering Research and Applications*, 4(7), 74–78.

