Decoding Sentiment: An Enhanced PCA in Tagalog Speech Using Prosodic Suprasegmental Features

[1] Ailen B. Garcia
Technological Institute of the Philippines
Aurora Blvd., Cubao, Quezon City
Philippines
qabgarcia01@tip.edu.ph

[2] Bobby D. Gerardo, D.Eng
Northern Iloilo State University
Estancia, Iloilo, Philippines
bgerardo@nisu.edu.ph

[3] Ruji P. Medina, Ph.D
Technological Institute of the Philippines
Aurora Blvd., Cubao, Quezon City
Philippines
ruji.medina@tip.edu.ph

***Abstract*—Prosodic Suprasegmental Features (PSF) and Principal Component Analysis (PCA) are techniques used in speech sentiment analysis that are capable of capturing the emotional content of speech sentiment and providing increased accuracy. However, PSF is prone to high dimensional features of data and PCA is sensitive to outliers. This study explores the effectiveness of enhanced PCA sentiment analysis in Tagalog speech, focusing on prosodic suprasegmental features such as pitch, duration, energy/intensity, and intonation. Using the optimized PCA, the retrieved prosodic features' dimensionality was decreased, and it was then used to train a Support Vector Machine (SVM) classifier, which divided speech into three categories: positive, negative, and neutral. The model achieved an accuracy of 82%, with an f1-score of 81% and recall of 82% as compared to no PCA enhancement and the simple speech sentiment analysis having an accuracy of 64% and 14% respectively. The results demonstrate the effectiveness of prosodic suprasegmental features and enhanced PCA in capturing emotional nuances within spoken Tagalog and provide a solid foundation for sentiment analysis in speech, providing insightful information on the emotional content of Tagalog speech.**

Keywords—prosodic suprasegmental features, PCA, Tagalog speech, SVM, sentiment analysis

I. INTRODUCTION

An important component of a student's well-being is their emotional health, which has a big impact on their relationships, academic achievement, and quality of life[1]. Understanding the role of students' emotional intelligence has become more crucial as researchers and practitioners start looking into potential interventions, given the rise in mental health issues and the impact of psychosocial factors on students. One way to measure the student's emotional status is to get their sentiments. Sentiment analysis automates the sentiment extraction or classification process from sentiment reviews by utilizing Natural Language Processing (NLP), text analysis, and computational techniques [2][3].

In the Philippines, online meetings and communication have been implemented since the pandemic in many organizations including the educational sector [4][5]. Students expressed their reactions sentiments and ideas virtually during online or synchronous classes [6]. Investigating students' opinions, emotions using techniques of sentiment analysis [7] can understand the students' feelings that students experience in academic, personal, and societal environments.

Recently, sentiment analysis has gained significant attention due to its wide-ranging applications in various domains such as student/customer feedback analysis, social media monitoring, market research, etc. in learning what do people think and want [5][8]. Sentiment analysis is the process of categorizing the emotional tone that speakers express through spoken words. This process can provide important information about the attitudes, beliefs, and sentiments of [9][10] those who are speaking. With the increasing availability of speech data and people exchange information through speech, there is a growing interest in sentiment analysis of speech or audio[11]. However, analyzing sentiment using audio signals is a significant challenge [12][13] due to the difficulty of accurately determining the robust feature set needed to detect sentiments expressed within the audio signal [14][15]. Also, [12] some models failed to make accurate predictions in emotion recognition task and sentiment analysis tasks with higher numbers of classes.

One of the techniques used in data reduction is Principal Component Analysis or PCA. It is a commonly used technique for data reduction or compression that preserves the important features of the data [16]. It also transforms high-dimensional data into a lower-dimensional subspace for feature reduction. PCA is able to reduce the number of features and is observed to improve accuracy rate [17].

Developing strong sentiment analysis algorithms faces significant challenges in languages like Tagalog which have a dearth of annotated speech data and variation in the intonation patterns across the different parts of the Philippines. Filipino or Tagalog is the national language of the Philippines. It belongs to the Malayo-Polynesian group of Austronesian languages, and it is a member of the Central Philippine subgroup of Philippine languages [18]. It is one of the most spoken language forms in the Philippines, [19] wherein it can be used for communication among people orally or virtually.

Tagalog, as one of the widely spoken languages in the Philippines with 45 million speakers. It presents an interesting subject for sentiment analysis. However, existing sentiment analysis models primarily focus on English or other widely studied languages, leaving a gap in sentiment analysis research specific to Tagalog speech which is the spoken language of the students. To address these challenges, this study focuses on a novel approach for developing a Tagalog speech sentiment model. The proposed approach leverages prosody suprasegmental features using enhanced Principal Component Analysis (PCA) to effectively classify the sentiment expressed in Tagalog speech among the students.

II. RELATED WORKS

Sentiment analysis is the task of automatically determining the sentiment or emotional tone of text or speech [20]. The rapid growth of Internet-based applications, such as social media platforms and blogs, has resulted in comments and reviews concerning day-to-day activities [21]. Sentiment analysis is a systematic study to identify and extract information present in the source materials using natural

language processing, computational linguistics, and text analytics [22][23].

Recent advancement of social media which is an enormous ever-growing source has led people to share their views through various modalities such as audio, text and video [24]. Furthermore, in this last few years, sentiment analysis (SA) has attracted increasing interest in the text mining area. It increasingly [25] becomes a popular research area for opinion mining in education that analyses and understands students' opinions toward their institutions for improving the quality of decision-making.

Audio sentiment analysis using automatic speech recognition is an emerging research area and, in its infancy, where opinion or sentiment exhibited by a speaker is detected from natural audio. It is relatively underexplored when compared to text-based sentiment detection. Extracting speaker sentiment from natural audio sources is a challenging problem. Generic methods for sentiment extraction generally use transcripts from a speech recognition system, and process the transcript using text-based sentiment classifiers. [26][12][27].

Prosody, the suprasegmental aspects of speech including intonation, rhythm, stress, and pitch, plays a crucial role in communication across languages and cultures [28]. The suprasegmental characteristics of Tagalog are important for communication since they can alter the meaning of words, phrases, and sentences [29]. Examples are: 1.) Stress *(Diin)*, 2.) Pitch *(Tono)*, 3.) Intonation *(Himig)*, and 4.) Pause *(Antala)*. These characteristics enable speakers to express nuances, intentions, and feelings that aren't always possible to express with just one phrase. The essential role of prosody in shaping the richness and complexity of human communication, emphasizing its relevance across diverse linguistic contexts and communicative settings [30]. Prosodic features on the other hand, like (pitch, intensity and speech rate) [12] are generally the most commonly implemented features for Speech Emotion Recognition (SER) as they are considered highly correlated with speaker emotion. Also, [31] prosody research has had a significant impact in improving the naturalness of speech synthesis, and has found some successes improving information extraction from speech, speech assessment and extracting affect.

In speech emotion recognition, there are many methodologies that can be used in order to reduce the dimensionality of data such as principal component analysis (PCA), Linear discriminant analysis (LDA), Random forests, etc. PCA seems to be one of the most popular methodologies. PCA is a preprocessing linear transformation technique [32].

Analyzing sentiment using audio signals is a significant challenge due to the difficulty of accurately determining the robust feature set needed to detect sentiments expressed within the audio signal [14]. The following are the development and constant improvement of speech sentiment analysis: [12],[14],[33], [34], [35],[36],[37]

Table 1. Methods and Techniques in Speech Sentiment Analysis

Methods/Techniques	Dataset	Accuracy
• Spectrogram • STFT • Random Forest	• bag-of-visual-words • 2D image	76%
• universal speech • transformer	• CMU-MOSEI dataset	81%
• CNN/ RNN • BERT	• XD-Violence dataset	85.63%
• LSTM • Transformer	• RAVDESS • Emo-DB	RAVDESS: 75.62%, Emo-DB: 85.55%,
• Prosodic spectral • DNN	• RAVDESS datasets	Accuracy: 78.83%
• MFCC • Mel Spectrogram • Chroma	• TESS and RAVDESS • Custom datasets	• TESS 99.4% • RAVDESS 89.62% • Custom dataset 78.28%

III. METHODS

The following figure 1 is the proposed Tagalog speech sentiment architecture or model.

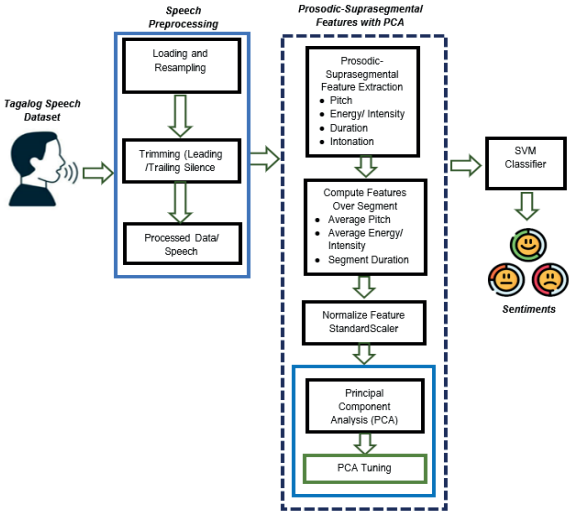

Fig. 1. The Proposed Tagalog Speech Sentiment Model.

A. Datasets

The datasets were from the survey and interviews of the students. There were 470 audio samples from different speakers (students). The recording was done in a closed room with a condenser microphone was utilized. An informed consent was distributed and filled out by the respondents as part of the research ethics standards.

B. Preprocessing

In the preprocessing, removing any irrelevant or noisy parts of the speech data, such as background noise, non-speech sounds, or interruptions. Normalize the voice data to guarantee that the amplitude and scale are the same in all of the recordings. This may entail modifying the audio samples' gain or volume levels. It also involves splitting the speech data into smaller segments, like words or phrases, to make feature extraction and analysis easier through Audacity application. The compilation of Tagalog speech was performed with proper annotation in the respective directories.

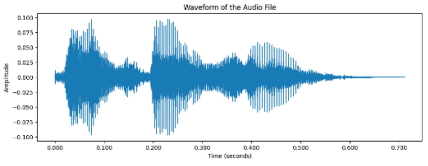

Fig. 2 Sample Waveform of the Audio File

B.1 Loading and Resampling

The loading part is to converts audio files into numerical array (time series) and a sampling rate. The function loads the audio file, returning the time-domain signal *'y'* and the sampling rate *'sr'*. The signal *'y'* is an array representing the amplitude of the audio signal at each time step. Resampling alters the audio signal's sample rate to maintain consistency between several audio files. When working with datasets that have different sample rates, resampling is frequently required. All audio files should be converted to the same sample rate to guarantee consistency in feature extraction and further analysis. The mathematical representation is as follows:

If *x(t)* represents the continuous audio signal, it is sampled at discrete time intervals to create a time series *y[n]*

The sampling rate *sr* (samples per second) determines the interval Δt between samples, where $\Delta t = 1/sr$.

The time series y[n] is a sequence of sample from the audio signal [38] [39]:

$$y[n] = x(n \cdot \Delta t) \tag{1}$$

If the original sampling rate is sr_{orig} and the target sampling rate is sr_{new}, librosa performs resampling. The relationship between the original and resampled signals is determined by [33] [39]:

$$y_{new}[n] = \text{Resample}(y_{orig}[n], sr_{orig}, sr_{new}) \tag{2}$$

B.2 Trimming

Trimming eliminates the audio signal's silent segments at the start and finish in order to concentrate on its non-silent parts [39] [40]. This can be applied to audio data preparation to remove unnecessary quiet and condense the dataset. It removes leading and trailing silence from an audio signal using the librosa library.

C. Prosodic Suprasegmental Feature Extraction

One of the key components of a sentiment analysis is feature extraction. The speech signal contains a variety of characteristics. This technique extracts and chooses the best characteristics from speech signals, which include a wealth of information about speech. Choosing the appropriate elements that allow the vocal signal to convey additional emotional information which speech characteristics are most effective at differentiating between emotions is unclear. In this study, we focused on the four types of features, the pitch, energy/intensity, duration, and intonation based on the prosodic suprasegmental features.

C.1 Pitch.

Pitch is one of the fundamental prosodic features used in speech sentiment analysis. It is vital to the expression of emotions through speech and represents the frequency of vibration of the vocal cords. Emotions such as happiness or enthusiasm are frequently correlated with higher pitches, whereas calmness or sad might be indicated by lower pitches. Below is the mathematical representation of pitch extraction process [39] [40].

- Compute the pitch matrix *Pij* where:

$$P_{ij} = \text{piptrack}(y, sr) \tag{3}$$

Pij represents the pitch probability at the *i-th* frequency bin and the *j-th* frame

- Extract the pitch [38] for each frame *j:*

$$\text{pitch}[j] = \max_i P_{ij} \tag{4}$$

This gives the maximum pitch probability for each frame *j.*

- Filter out zero values [39] [40]:

$$\text{pitch\_filtered} = \{\text{pitch}[j] \mid \text{pitch}[j] > 0\} \tag{5}$$

Through the process of filtering out frames where no pitch was observed, the most likely basic frequency, or pitch, for each frame of the audio file is efficiently extracted.

C.2 Energy or Intensity

In audio processing, Root Mean Square (RMS) Energy is a metric used to estimate the strength or volume of an audio signal within a given time window. It is also helpful for examining the signal's intensity over time, which may indicate various characteristics including speech stress patterns, music dynamics, or the presence of speech.

The RMS energy for a frame of audio is calculated using the following formula [38] [39] [40]:

$$\text{RMS Energy} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x[n]^2} \tag{6}$$

Where:

N is the number of samples in the frame

x[n] represents the amplitude of the *n-th* sample in the frame.

C.3 Duration

An audio signal's duration, which is typically expressed in seconds, is the entire amount of time the audio file has been recorded. Knowing the duration is important for many tasks in audio processing, including segmentation, time-based analysis, and just figuring out how long the audio content is.

The duration of the audio signal can be calculated using the following formula:

$$\text{Duration} = \frac{\text{Number of Samples}}{\text{Sampling Rate}} = \frac{N}{sr} \tag{7}$$

Where:

N is the total number of samples in the audio signal

sr is the sampling rate in samples per second (Hz)

C.4 Intonation

Intonation refers to the variation in pitch while speaking, which can convey different meanings, emotions, or emphasis. It is how the pitch of the voice rises and falls throughout speech. Emotions like surprise, delight, grief, and sarcasm can all be expressed through intonation patterns. A sentence's finality or confidence, for instance, may be indicated by a falling intonation, whereas an inquiry or uncertainty may be shown by a rising intonation. These pitch changes can be captured in sentiment analysis to aid in differentiating between neutral, positive, and negative feelings.

This is the formula for intonation particularly the pitch slope:

$$\text{Pitch Slope} = \frac{F_0(t_2) - F_0(t_1)}{t_2 - t_1} \quad (8)$$

Where:

- **Pitch Slope** measures the rate of change in pitch between two time points t_1 and t_2. A positive slope indicates rising intonation, while a negative slope indicates falling intonation.

D. Compute Feature Over Segments

To better collect prosodic information, analyze the attributes across specific audio segments. Divide the audio into smaller segments and compute the mean features for each segment. The following are the formula for the computation of features:

D.1 Pitch

$$\text{Average\_Pitch} = \frac{1}{N}\sum_{i=1}^{N}\text{Pitch}_i \quad (9)$$

Where N is the number of pitch values in the segment

D.2 *Energy or Intensity*

$$\text{Average\_Intensity} = \frac{1}{M}\sum_{j=1}^{M}\text{RMS}_{t_j} \quad (10)$$

Where M is the number of segments or frames, and the RMS_{tj} is the RMS value for the *j-th* segment.

D.3 Duration

$$\text{Segment\_Duration} = \frac{\text{Number\_of\_Samples}}{\text{Sampling\_Rate}} \quad (11)$$

Where: Number_of_Samples is the number of samples in the segment and Sampling_Rate is the number of samples per second (Hz)

D.4 *Intonation*

$$\text{Mean Pitch} = \frac{1}{N}\sum_{i=1}^{N}F_0(t_i) \quad (12)$$

Where N is the number of pitch samples

Mean pitch gives the average pitch over a segment of speech, which can help identify the overall intonation pattern (e.g., whether the speech is generally high-pitched or low-pitched).

E. Normalize Features

A vital part of data preprocessing, particularly for machine learning applications, is normalizing features. The convergence and performance of learning algorithms are enhanced by ensuring that features have similar scales.

Feature scaling, or feature normalization, is the process of putting feature values on a same scale by modifying their range. This is significant because features on comparable scales enable many machine learning algorithms to operate more efficiently or to converge more quickly. Standardize the features to have a mean of 0 and a variance of 1, ensuring comparability. This technique transforms the features to have a mean of 0 and a standard deviation of 1, having this formula:

$$x' = \frac{x - \mu}{\sigma} \quad (13)$$

F. Principal Component Analysis

One statistical method for reducing the dimensionality of data while retaining as much variance as possible is principal component analysis, or PCA. It converts the data into a new coordinate system in which the first axes (principal components) are where the largest deviations by any projection of the data end up.

With principal component analysis (PCA), a group of related variables is converted into a new, uncorrelated set of variables known as prime elements. On the other hand, the PCA is useless if the data is already irrelevant. The primary elements, in conjunction with the irrelevant data, are arranged orthogonally according to the variability they reflect. That is, the optimal amount of variability within the original information set is represented by the primary principal part for a single dimension.

The algorithm and formula of PCA are the following:

- Center the Data: Subtract the mean of each feature from the data to center it around the origin:

$$\tilde{X} = X - \bar{X} \quad (14)$$

- Compute the Covariance Matrix – Calculate the covariance matrix of the centered data, where n is the number of samples.

$$C = \frac{1}{n-1}\tilde{X}^T\tilde{X} \quad (15)$$

- Compute the Eigenvalues and Eigenvectors – Solve the eigenvalue problem for the covariance matrix:

$$Cv_i = \lambda_i v_i \quad (16)$$

- Sort and select principal components
- Project the original data into the principal components where W is the matrix of the top k eigenvectors.

$$X_{\text{pca}} = \tilde{X}W \quad (17)$$

where λ_i are the eigenvalues and v_i are the eigenvectors

G. PCA Tuning

To achieve the best result of the machine learning model performance, an optimize PCA or PCA tuning was applied in the study. It involves finding the optimal number of components that provides the best adjustment data reduction and preserving information.

H. Support Vector Machines

In this work, three different classes of sentiments such as positive, negative and neutral are classified using a multiclass support vector machine (SVM) classifier. This is a multiclass SVM classifier, which is trained using three distinct classes at first. It is then further trained using speech signal features that have been retrieved, selecting just the best features from the speech signals. Test input is received after a multiclass SVM has been trained, and it is compared to three distinct classes and trained speech features.

IV. RESULTS AND DISCUSSIONS

The simulation of the study includes the following: python programming using the Google Colab, Audacity for editing the audio or speech samples, Librosa library in python for speech analysis and manipulation, tensorflow, scikitLearn, and an Intel (R) Core i7 65000U CPU @ 2.50 GHz with 8 GB of RAM.

Following the proposed Tagalog Speech Sentiment architecture or model the following results were obtained:

A. Prosodic Suprasegmetal Features with PCA Results

Figure 3 shows the result of prosodic suprasegmental features with Principal Component Analysis (PCA). The dataset can be made simpler, underlying patterns can be found, and the performance of further analysis or models can be enhanced by using PCA on prosodic suprasegmental characteristics.

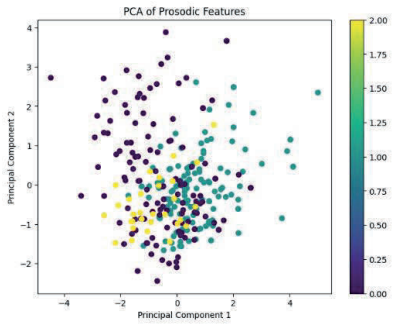

Fig. 3. PCA of Prosodic Suprasegmental Features

B. SVM Classification Results

The Support Vector Machines or SVM's model performance was evaluated based on the following evaluation metrics: accuracy, precision, recall, and F1 score. Figure 4 show the results of the Tagalog speech sentiment analysis. The model performs well overall with the accuracy rate of 82%. The highest performance observed for positive sentiment, with an F1 score of 81%.

	precision	recall	f1-score	support
Negative	0.77	0.90	0.83	41
Neutral	0.80	0.40	0.53	10
Positive	0.88	0.83	0.85	42
accuracy			0.82	93
macro avg	0.82	0.71	0.74	93
weighted avg	0.82	0.82	0.81	93

Fig. 4. Tagalog Speech Sentiment Accuracy Results

C. Comparison of Tagalog Speech Sentiment Analysis with PCA and without PCA based on Prosodic Suprasegmental Features

In comparing the results of enhanced PCA Tagalog speech sentiment utilizing prosodic suprasegmental features and without PCA enhancement and the original model of speech sentiment, the proposed model got the highest accuracy results of 82% compared to 64% and 14.75% respectively. This means that the proposed model performs well in the Tagalog speech sentiment analysis.

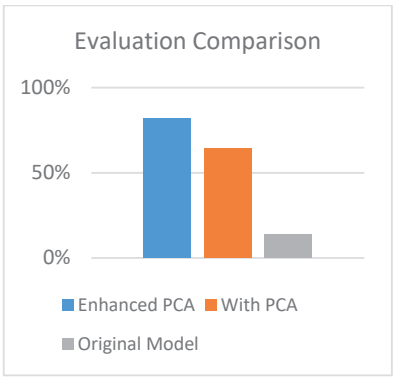

Fig. 5 Accuracy Rate Comparison of the Enhanced PCA, With PCA and the Original Model

V. CONCLUSION

The enhanced PCA sentiment analysis of Tagalog speech utilizing prosodic suprasegmental features such as pitch, duration, and intensity, achieved a notable accuracy of 82% using a Support Vector Machine (SVM) classifier. The prosodic suprasegmental feature set's dimensionality was successfully decreased by using an optimized PCA, which allowed the SVM to concentrate on the most important elements and improve the model's capacity to differentiate between positive, negative, and neutral sentiments. The 82% accuracy shows how prosodic elements can be used to capture refined emotional differences in Tagalog speech. This result is especially encouraging considering how difficult it is to do sentiment analysis in a language as context- and tone-sensitive as Tagalog. Overall, the study provides a strong foundation for further investigation and application in this field by demonstrating that enhanced PCA analysis of prosodic suprasegmental characteristics is a feasible method for classification of sentiment in Tagalog speech. The extraction of suprasegmental features is crucial in sentiment analysis and the selection of deep learning models are the considerations for future research.

ACKNOWLEDGMENT

The author would like to acknowledge the support provided by the people behind this undertaking. To Occidental Mindoro State College (OMSC) students and teachers for assisting with the collection of data. To her family, friends, professors, and classmates for the encouragement and guidance. Above all, to our almighty God, for wisdom, strength, and courage to pursue this endeavor.

REFERENCES

[1] Hsieh, W. J., Powell, T., Tan, K., & Chen, J. H. (2021). Kidcope and the COVID-19 pandemic: understanding high school students' coping and emotional well-being. *International journal of environmental research and public health*, *18*(19), 10207.

[2] Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, *30*(4), 330-338.

[3] Dolianiti, F. S., Iakovakis, D., Dias, S. B., Hadjileontiadou, S., Diniz, J. A., & Hadjileontiadis, L. (2018, June). Sentiment analysis techniques and applications in education: A survey. In *International conference on technology and innovation in learning, teaching and education* (pp. 412-427). Cham: Springer International Publishing.

[4] Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Applied Sciences*, *11*(18), 8438.

[5] Osmanoğlu, U. Ö., Atak, O. N., Çağlar, K., Kayhan, H., & Can, T. (2020). Sentiment analysis for distance education course materials: A machine learning approach. *Journal of Educational Technology and Online Learning*, *3*(1), 31-48.

[6] Dolianiti, F. S., Iakovakis, D., Dias, S. B., Hadjileontiadou, S., Diniz, J. A., & Hadjileontiadis, L. (2018, June). Sentiment analysis techniques and applications in education: A survey. In *International conference on technology and innovation in learning, teaching and education* (pp. 412-427). Cham: Springer International Publishing.

[7] Shanthi, I. (2022). Role of educational data mining in student learning processes with sentiment analysis: A survey. In *Research Anthology on Interventions in Student Behavior and Misconduct* (pp. 412-427). IGI Global.

[8] Baragash, R., & Aldowah, H. (2021, March). Sentiment analysis in higher education: a systematic mapping review. In *Journal of Physics: Conference Series* (Vol. 1860, No. 1, p. 012002). IOP Publishing.

[9] Kaushik, L., Sangwan, A., & Hansen, J. (2017). Automatic Sentiment Detection in Naturalistic Audio. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25, 1668-1679.

[10] Boukabous, M., & Azizi, M. (2022, March). Multimodal sentiment analysis using audio and text for crime detection. In *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-5). IEEE.

[11] Liu, G., Cai, S., & Wang, C. (2023). Speech emotion recognition based on emotion perception. *EURASIP Journal on Audio, Speech, and Music Processing*, *2023*(1), 22.

[12] Atmaja, B. T., & Sasou, A. (2022). Sentiment analysis and emotion recognition from speech using universal speech representations. *Sensors*, *22*(17), 6369.

[13] Bansal, M., Yadav, S., & Vishwakarma, D. K. (2021, April). A language-independent speech sentiment analysis using prosodic features. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1210-1216). IEEE.

[14] Luitel, S., & Anwar, M. (2022, August). Audio Sentiment Analysis using Spectrogram and Bag-of-Visual-Words. In *2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 200-205). IEEE.

[15] Andayani, F., Theng, L. B., Tsun, M. T., & Chua, C. (2022). Hybrid LSTM-transformer model for emotion recognition from speech audio files. *IEEE Access*, *10*, 36018-36027.

[16] Jagtap, S., & Desai, K. R. (2019). REAL-TIME SPEECH BASED SENTIMENT RECOGNITION.

[17] Patil, S., & Kharate, G. K. (2022, January). PCA-Based Random Forest Classifier for Speech Emotion Recognition Using FFTF Features, Jitter, and Shimmer. In *International Conference on Electrical and Electronics Engineering* (pp. 194-205). Singapore: Springer Singapore.

[18] Schachter, P., & Reid, L. A. (2018). Tagalog. In *The world's major languages* (pp. 852-876). Routledge.

[19] Boquiren, A. J., Garcia, R., Hungria, C. J., & de Goma, J. (2022). Tagalog Sentiment Analysis Using Deep Learning Approach With Backward Slang Inclusion.

[20] Raja, J. G. J. S., & Juliet, S. (2023, May). Deep learning-based sentiment analysis of Trip Advisor reviews. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 560-565). IEEE.

[21] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780.

[22] Anand, S., & Patra, S. R. (2022). Voice and Text Based Sentiment Analysis Using Natural Language Processing. In *Cognitive Informatics and Soft Computing: Proceeding of CISC 2021* (pp. 517-529). Singapore: Springer Nature Singapore.

[23] Swetha, B. C., Divya, S., Kavipriya, J., Kavya, R., & Rasheed, A. A. (2017). A novel voice based sentimental analysis technique to mine the user driven reviews. *International Research Journal of Engineering and Technology*.

[24] Mohanty, M. D., & Mohanty, M. N. (2022). Verbal sentiment analysis and detection using recurrent neural network. In *Advanced Data*

[25] *Mining Tools and Methods for Social Computing* (pp. 85-106). Academic Press.

[26] Baragash, R., & Aldowah, H. (2021, March). Sentiment analysis in higher education: a systematic mapping review. In *Journal of Physics: Conference Series* (Vol. 1860, No. 1, p. 012002). IOP Publishing.

[27] Boukabous, M., & Azizi, M. (2022, March). Multimodal sentiment analysis using audio and text for crime detection. In *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-5). IEEE

[28] Kaushik, L., Sangwan, A., & Hansen, J. (2017). Automatic Sentiment Detection in Naturalistic Audio. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25, 1668-1679.

[29] Dimzon, F. D., & Pascual, R. M. (2023). Prosodic characterisation of children's Filipino read speech for oral reading fluency assessment. *International Journal of Technology Enhanced Learning*, *15*(1), 74-94.

[30] Liscombe, J. J. (2007). *Prosody and speaker state: paralinguistics, pragmatics, and proficiency*. Columbia University.

[31] Bektosheva, D. (2024). THE ROLE OF PROSODY IN COMMUNICATION. *Академические исследования в современной науке*, *3*(12), 102-104.

[32] Rosenberg, A. (2018, June). Speech, prosody, and machines: Nine challenges for prosody research. In *Proc. Speech Prosody* (pp. 784-793).

[33] Patel, N., Patel, S., & Mankad, S. H. (2022). Impact of autoencoder based compact representation on emotion detection from audio. *Journal of Ambient Intelligence and Humanized Computing*, *13*(2), 867-885.

[34] Bansal, M., Yadav, S., & Vishwakarma, D. K. (2021, April). A language-independent speech sentiment analysis using prosodic features. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1210-1216). IEEE.

[35] Mohanty, M. D., & Mohanty, M. N. (2022). Verbal sentiment analysis and detection using recurrent neural network. In *Advanced Data Mining Tools and Methods for Social Computing* (pp. 85-106). Academic Press.

[36] Idris, I., Salam, M. S. H., & Sunar, M. S. (2016). Speech emotion classification using SVM and MLP on prosodic and voice quality features. *Jurnal Teknologi*, *78*(2-2).

[37] Gumelar, A. B., Kurniawan, A., Sooai, A. G., Purnomo, M. H., Yuniarno, E. M., Sugiarto, I., ... & Fahrudin, T. M. (2019, August). Human voice emotion identification using prosodic and spectral feature extraction based on deep neural networks. In *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)* (pp. 1-8). IEEE.

[38] Charoendee, M., Suchato, A., & Punyabukkana, P. (2017, July). Speech emotion recognition using derived features from speech segment and kernel principal component analysis. In *2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1-6). IEEE.

[39] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015, July). librosa: Audio and music signal analysis in python. In *SciPy* (pp. 18-24).

Koffi, E. (2020). A tutorial on acoustic phonetic feature extraction for automatic speech recognition (ASR) and text-to-speech (TTS) applications in African languages. *Linguistic Portfolios*, *9*(1)