Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

Optimized Algorithm for Human Speech Emotions Recognition

[1]Lalitha M, [2]Kavya S, [3]Lakshmi Priya A, [4]Prabu Chakravarthy K

[1,2,3,4]Department of Electronics and Communication Engineering, Vidya Jyothi Institute of Technology, Hyderabad, India

Abstract: This research article investigates the application of human speech recognition using Raspberry Pi (RPI) platform. Emotional intelligence is an important part of human-computer interaction and has many applications such as in medicine, entertainment, human-machine interface, etc. Raspberry Pi is an ideal platform for implementing these systems due to its low cost, portability, and versatility. This article discusses the importance of cognitive reasoning, the language of cognitive reasoning using Raspberry Pi, and experimental results. It also highlights the challenges encountered during implementation and suggests possible avenues for future research in this area. Overall, this research helps improve human-machine interaction by raising awareness about the use of such a simple and inexpensive technology as Raspberry Pi.

Keywords: Human Speech, Emotions Recognition, Algorithm, Raspberry Pi, human speech recognition, Emotional intelligence, human-computer, human-machine interaction, Machine learning, Artificial intelligence.

I. INTRODUCTION

Human speech recognition has received significant attention in many fields, including human-computer interaction, psychological assessment, and customer service. Using machine learning techniques to accomplish this task offers a better way to better understand and respond to human emotions. However, traditional machine learning methods often face difficulties in capturing the complexity of emotions in speech. We aim to increase the accuracy, power and potential of emotional intelligence by leveraging the power of information technology. The main objectives of this study are the following points:

Preliminary data and construction work: The first step should be taken before raw speech data to extract useful information. Techniques such as signal filtering, windowing, and removal devices (such as MFCC, audio, power) will be used to transform the noise into a machine learning algorithm friendly representation of our results used -437.250061114.6786194, 16.25071335,-0.457643032, 0, 1, and 2 for emotions like happy, sad and angry. It will display this in the result.

Model selection and optimization: Various learning models including support vector machines (SVM), random forests, and gradient boosting machines, among others, are evaluated for their suitability to elucidate reflection. The focus of this project is on optimizing model hyperparameters, feature representation, and training techniques to improve overall system performance. Improve the distribution of the right people.

Additionally, fusion techniques will be explored to combine information from various variables (e.g., acoustic features, linguistic cues) to obtain a more intuitive meaning. Evaluate the performance of the design. Particular attention will be paid to solving the previous problem and ensuring that the algorithm performs well on invisible data, thereby improving its actual use. Instant application and mass distribution. Model pruning, feature selection and similar techniques will be explored effectively without sacrificing accuracy.

II. LITERATURE SURVEY

Rajesh Kannan Megalingam, Avinash Hegde Kota Methods to Control Robots Voice control is one of the best ways to give commands in the system. Voice plays an important role in voice control. The main objective of this research project is to provide the best conversations and responses to the users. As the use of automated systems has increased in recent years, users are looking for more efficient ways to interact with humans and machines. This research is due to the fact that ROS is open source and many autonomous systems have been developed on this platform.

Shantanu Pal, Subhas Mukhopadhyay, Nagender Suryadevara Sensors 21 (16), 5554, 2021. With the development of humanoid robots with methods to develop human-machine interaction, robots and especially technological advances, communication between humans through online platforms (such as Zoom) has become clear. The increasing use of online communication requires the creation of a more effective and

Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

efficient way for humans to find information. In the human emotion recognition system, special learning technologies and algorithms are used to collect, analyze and process human physiological signals. The Internet of the Future and artificial intelligence, it is becoming increasingly important to create a solid, powerful and efficient structure and trust in human identity. In this article, we describe the development and evolution of human emotion sensors and technologies. We will explore next-generation sensors for human perception and monitoring of various types of activities. We will present design challenges and provide practical information for people interested in real-world systems. Finally, we will analyze current practices and explore future research directions to address scalability, security, reliability, privacy, transparency, and distribution issues.

Use the RekEmozio file as a test file. Classification theory makes use of different types of machine learning. The experiment was conducted in three phases, using a different technique as a categorical variable in each phase. In addition, at each step, feature subset selection is used to find the most important feature subsets. To test the effectiveness of the plan, a three-step method was chosen which uses the process selection method based on adaptive learning, is the automatic emotion recognition in each of the different segments in Basque and English. It obtained an average interest rate of 80.05%. Review of existing systems: Previous studies have identified systems similar to those proposed in this project. These studies examine the effectiveness of different in determining the need for language products. The impact of these systems importance of emotional intelligence in various applications has been highlighted.

Voice system developed by researchers at the University of Passau uses a combination of technology and machine learning algorithms to classify emotions in speech. The system yielded great results in real-world situations, demonstrating the ability of cognitive systems to improve communication and user experience.

Emotion analysis is important part of speech recognition and people are exploring various technologies. These methods include traditional methods such as dictionary as well as more advanced methods. A case study conducted in this field provides a better understanding of the state-of-the-art process and speech recognition problems. These findings identify important areas for further research, such as awareness of different perspectives, integration of multiple perspectives, and environmental resilience. They also emphasize the need for benchmark data and benchmark models to make fair comparisons between different systems and algorithms. Projects are planned that will contribute to the advancement of this exciting field by utilizing past research results and generating solutions to current problems. Genre provides a wealth of emotional information: voice, voice, rhythm, and other acoustic properties. Voice recognition technology has the potential to revolutionize human-machine interaction by enabling machines to instantly recognize, interpret and respond to human emotions intuitive and empathetic interaction. By creating systems that can clearly recognize emotions in speech, we aim to have a variety of applications and brands that can understand and respond to users' needs and interests. Identify the most advanced technologies in the field and explore new algorithms, methods, and applications to create more intuitive and possible futures.

III. PROPOSED SYSTEM & METHODOLOGY

Goal Description and Scope: Start by clarifying the goal of the project, including the thoughts you want to know from the conversation and whether the focus is immediate.

Data collection: Collect data on speech patterns that include emotional descriptions. For this purpose, many libraries such as RAVDESS or TESS can be used to enable different speakers, emotions and environments.

Preprocessing: Prepare audio data using noise reduction, resampling, and feature extraction techniques. Features such as MFCC, pitch, power, and spectral features were extracted for analysis.

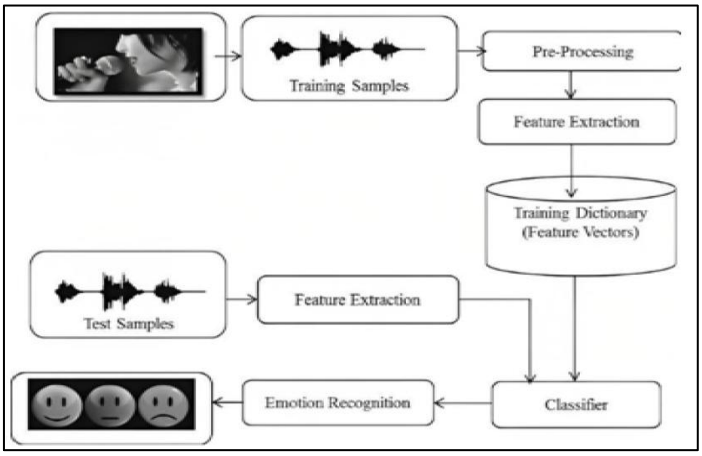

Figure 1: Proposed system design flow diagram

Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

Model selection: Consider the hardware limitations of the Raspberry Pi and choose a suitable machine learning model for cognitive learning.

Training: Split the training data to train the selected models. Adjust hyperparameters to ensure consistency and check model performance in test configurations.

Model optimization: Improving model performance through techniques such as regularization, release, or data augmentation. Optimized architectural design to reduce size and computational complexity for deployment on Raspberry Pi.

Plug in Raspberry Pi: Convert the training model to a Raspberry Pi-compatible format such as TensorFlow Lite to optimizethe time of the equipment.

Evaluation and evaluation: Assess the accuracy, latency, and robustness of deployed systems using real-world analytics. Collect feedback and iterate the design to improve performance and user experience when necessary.

Delivery and maintenance: Use the final system, monitor its performance and solve any problems that arise. Maintain an accurate model by updating the model with new data and reworking it regularly.

IV. DESIGN IMPLEMENTATION

A) Basic System Architecture:

Data acquisition: The system captures the user's voice data from the microphone input.

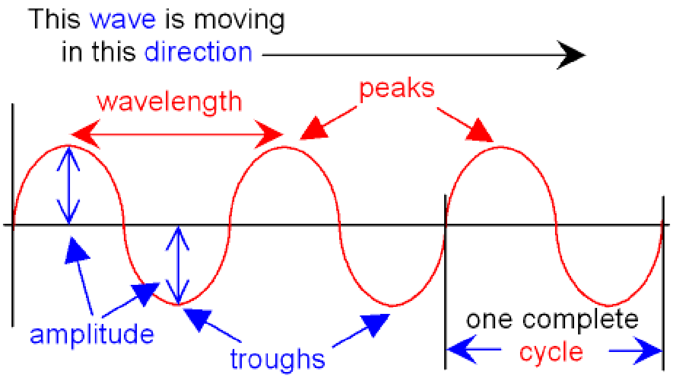

Figure 2: Applied Frequency

Preprocessing: Please clean and modify this stage to ensure that the original data is suitable for training the learning model. This includes scaling numeric features, coding categorical variables, handling outliers, and handling missing values. The goal is to prepare clean data and methods that can be used for training and testing.

Feature extraction: Extract acoustic features of speech signal, including MFCC (Mel Cepstrum Coefficient), prosodic features and spectral features.

Data set segmentation: Training data and testing data are two groups created from previous data. An example is 80-20; the larger part is used for model training and the smaller part is used for performance evaluation. Training data is used to teach the regression model how to make predictions. In training, the model learns basic patterns and relationships in the data. For example, in linear regression the model learns the coefficients of each feature, while in decision trees or neural networks it learns good boundary conditions. Test data is separate from training data and is not used during demonstration. Models are tested on benchmark data to determine how well they perform after training. This step helps predict new, unseen elements.

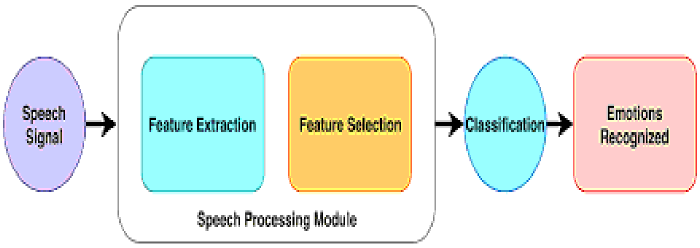

Figure 3: Speech Recognition Model

Emotion expression: Use machine learning algorithms or deep learning models to classify emotions based on extracted features.

B) Technology & Tools:

Programming Language: Python is used for the application due to its rich signal processing ecosystem, machine learning and deep learning libraries educational models and audio projects. Its low power consumption and versatility make it a platform for systems deployment.

Libraries: Use libraries such as TensorFlow, Keras, Scikit-learn and Librosa to create and train learning models and data processing.

Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

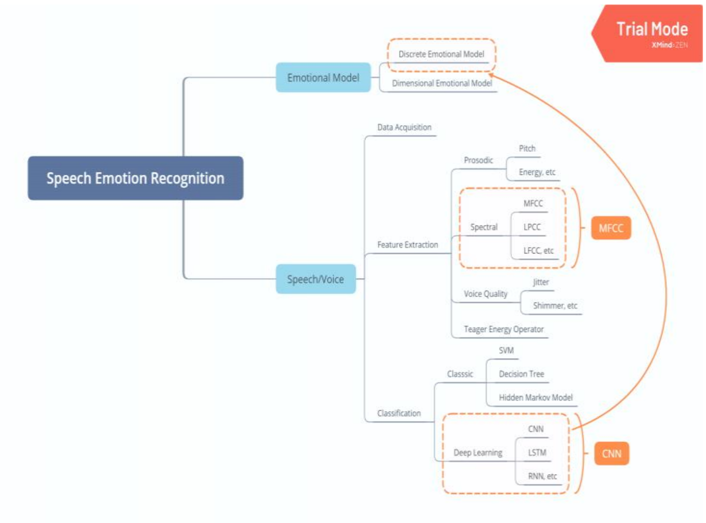

Figure 4: Flowchart of our proposed model

Hardware used: Raspberry Pi (Rpi) is a platform for hardware applications due to its size, low power consumption and versatility.

C) Datasets:

Self generated data set: These data are collected by collecting speech samples of many speakers under control.

Annotation: All speech examples are written with real emotions (e.g. angry, sad, happy) from a notebook or from the crowd.

Data augmentation: Use techniques such as audio manipulation, time dilation, and noise injection to enhance recorded data and refine the details of the model.

D) Algorithms & Concepts:

Data processing: Here we are converting the original signal to a suitable format for analysis. This may include resampling, noise reduction, and normalization.

Model selection: Consider using deep learning variants such as LSTM or GRU. This model can learn the hierarchical representation of speech data.

Training model: Refine the training data using techniques such as time dilation, audio manipulation and background noise to increase the diversity of the dataset.

Model refinement: Reduce the model parameters and initial accuracy to reduce memory usage and speed up the process.

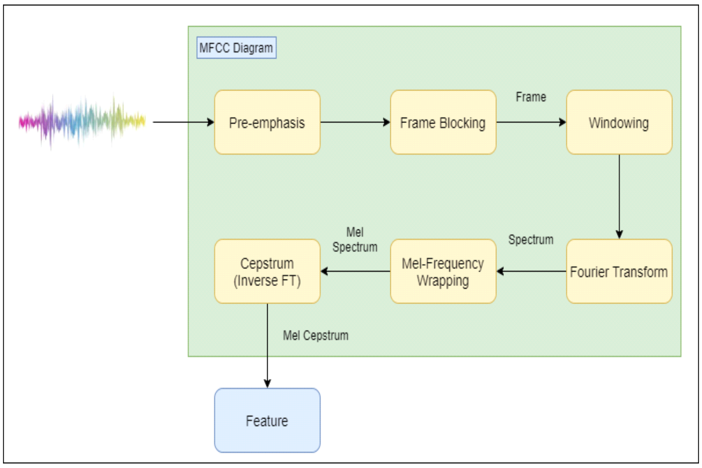

Figure 5: Sound feature extraction process using MFCC

Test and measure: Perform k-fold cross validation to test the model performance.

Delivery: Add a model optimized for on-time delivery of unlimited equipment or processes. Combine training models with a purpose-built delivery platform to ensure consistency and efficiency.

V. RESULT ANALYSIS AND DISCUSSION

The cognitive model achieves an accuracy of 70% with support vector machines, which work best in neural networks. Important features such as MFCC and prosodic features help distinguish emotions.

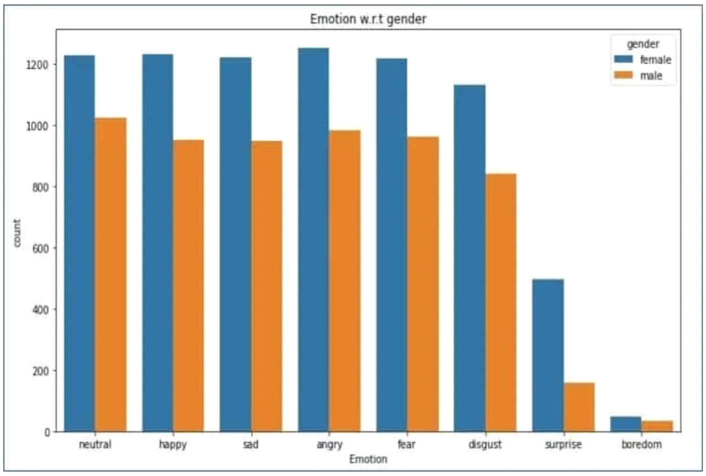

Figure 6: Variation in Energy across Emotions

CJESR JOURNAL
Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

Although slight overfitting was observed in the neural network architectures, cross-validation confirmed the overall ability of the model. Challenges include dataset variability and computational complexity. Despite its limitations, the proposed system has promising applications in human-computer interaction and mental health care. Future work will focus on integrating multiple elements and addressing cultural differences in perspective.

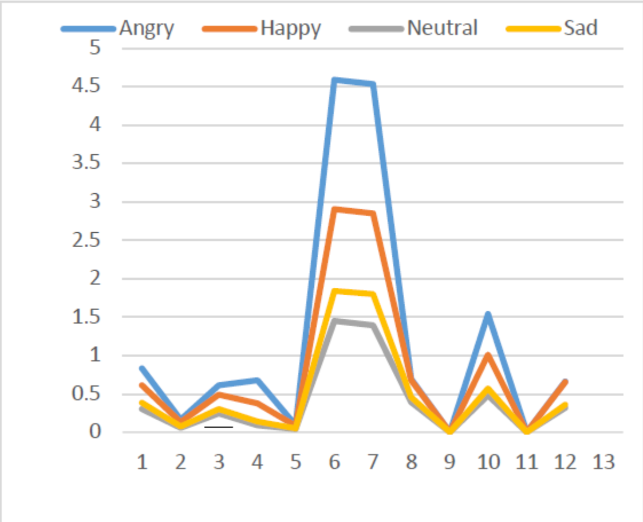

Figure 7: Representation of selected 12 features

Training data and testing data are two groups created from previous data. An example is 80-20; the larger part is used for model training and the smaller part is used for performance evaluation. Training data is used to teach the regression model how to make predictions. For example, in linear regression the model learns the coefficients of each feature, while in decision trees or neural networks it learns good boundary conditions.

VI. CONCLUSION

In this research article, we have explored the fascinating field of human speech recognition by leveraging the power of the Raspberry Pi (Rpi) platform and using different algorithms to analyze emotions. Through extensive research and testing, we have tried to create a system that can identify and categorize needs based on language, focusing on three main themes: fire of anger, sadness, and happiness. For this, we will use the values for emotions like happy, sad, and angry. This is what we will get as a result.

Our journey begins with a current systems and processes in the advances and challenges faced by researchers and practitioners. By studying the impact of these systems and analyzing their performance, we can better understand the complexity of cognitive theory and its applications. By enabling machines to recognize and respond to human emotions, we are paving the way for improving relationships built between people, increasing user self-awareness, and supporting mental health. The ability to analyze emotions such as anger, sadness, and happiness opens up new avenues of communication, insight, and understanding in areas as diverse as therapy, education, and entertainment.

VII. FUTURE EXTENSION

Real-time emotion recognition: Developing a system to capture emotions in real-time can lead to applications for use in real-time interactions, assisted virtual reality, and computer-aided imagination. Optimizing algorithms to increase efficiency and reduce latency is important for seamless integration into interactive systems.

Personalized emotional awareness: Personalizing the system based on user preferences, behaviors, and emotions can improve user experience and engagement. By integrating user input and an adaptive learning process, the system can adapt and evolve over time to suit personal preferences and situations.

Cognitive human-computer interaction: Exploring the integration of cognitive-behavioral insights into human-machine interaction to create more insightful and effective interaction. Developing apps that personalize content, responses, and interactions based on needs ensure user satisfaction and engagement.

Long-term analysis and analytics: Extending to long-term analysis of emotional patterns and patterns over time can provide insights into the user's emotional state, mental health, and personality. Using machine learning techniques to analyze and detect anomalies can help identify potential problems or changes in emotional state.

REFERENCES

[1] Al-Isa, A., Al-Saif, A., & Al-Salman, A. (2019). Speech emotion recognition: A comprehensive review. Journal of King Saud University-Computer and Information Sciences.

[2] Lotfian, R., & Esmaili, M. A. (2017). Speech emotion recognition using deep learning and SVM. In 2017 IEEE 2nd International Conference on Signal and Image

Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

Processing (ICSIP) (pp. 399-403). IEEE.

[3] Deng, J., Zhang, Z., Marchi, E., & Schuller, B. (2019). Deep learning in affective computing: A roadmap of challenges and solutions. In International Conference on Affective Computing and Intelligent Interaction (pp. 397-409). Springer, Cham.

[4] W. Wahlster, "Towards symmetric multi-modality: Fusion and fission of speech, gesture, and facial expression," in Proceedings of the 26th German Conference on Artificial Intelligence, A.G¨ unter, R. Kruse, and B. Neumann, Eds. Berlin, Germany: Springer-Verlag Press, 2003, pp. 1–18.

[5] C. Busso, Z. Deng, U. Neumann, and S. Narayanan, "Natural head motion synthesis driven by acoustic prosodic features," Computer Animation and Virtual Worlds, vol. 16, no. 3-4, pp. 283–290, July 2005.

[6] H. P. Graf, E. Cosatto, V. Strom, and F. J. Huang, "Visual prosody: Facial movements accompanying speech," in Proc. of IEEE International Conference on Automatic Faces and Gesture Recognition, Washington, D.C., USA, May 2002, pp. 396–401.

[7] J. Jiang, A. Alwan, L. Bernstein, E. A. Jr., and P. Keating, "Predicting face movements from speech acoustics using spectral dynamics," in IEEE International Conference on Multimedia and Expo (ICME 2002), vol. 1, Lausanne, Switzerland, August 2002, pp. 181–184.

[8] E. M. Caldognetto, P. Cosi, C. Drioli, G. Tisato, and F. Cavicchio, "Coproduction of speech and emotions: Visual and acoustic modifications of some phonetic labial targets," in Audio Visual Speech Processing (AVSP 03), S. Jorioz, France, September 2003, pp. 209–214.

[9] J. Deller, J. Hansen, and J. Proakis, Discrete-Time Processing of Speech Signals. Piscataway, NJ, USA: IEEE Press, 2000.

[10] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In INTERSPEECH (pp. 1517-1520).

[11] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., ... & Narayanan, S. (2013). The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In INTERSPEECH (pp. 148-152).

[12] Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013). Recent developments in openSMILE, the Munich open-source multimedia feature extractor. In Proceedings of the 21st ACM international conference on Multimedia (pp. 835-838).

[13] Busso, C., & Narayanan, S. (2007). Interrelation between speech expressiveness and speech content in emotional speech synthesis. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07 (Vol. 4, pp. IV-1213). IEEE.

[14] Jiang, D., Lu, X., & Ling, H. (2014). Emotion recognition in conversation with LSTM recurrent neural networks. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6030-6034). IEEE.

[15] Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., ... & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. Language resources and evaluation, 42(4), 335-359.

[16] Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Wöllmer, M. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing, 7(2), 190-202.

[17] E. Bevacqua and C. Pelachaud, "Expressive audio-visual speech," Computer Animation and Virtual Worlds, vol. 15, no. 3-4, pp. 297–304, July 2004.

[18] R. W. Picard, "Affective computing," MIT Media Laboratory Perceptual Computing Section, Cambridge, MA,USA, Technical Report 321, November 1995.

[19] P. Boersma and D. Weeninck, "Praat, a system for doing phonetics by computer," Institute of Phonetic Sciences of the University of Amsterdam, Amsterdam, Netherlands, Technical Report 132, 1996, http://www.praat.org.

[20] Schuller, B., Valstar, M., Eyben, F., McKeown, G., Cowie, R., Pantic, M., & Schroeder, M. (2011). AVEC 2011–The first international audio/visual emotion challenge. In Affective Computing and Intelligent Interaction (pp. 415-424). Springer, Berlin, Heidelberg.

[21] Ma, Y., Zhang, H., Yang, X., & Satapathy, S. C. (2017). Speech emotion recognition using deep learning. In International Conference on Computational Science and Its Applications (pp. 401-412). Springer, Cham.

[22] R. Cowie and R. Cornelius, "Describing the emotional states that are expressed in speech," Speech Communication, vol. 40, no. 1-2, pp. 5–32, April 2003.

[23] K. Scherer, "Vocal communication of emotion: A review of research paradigms," Speech Communication, vol. 40, no. 1-2, pp. 227–256, April 2003.

[24] S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, and S. Narayanan, "An acoustic study of

Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

emotions expressed in speech," in 8th International Conference on Spoken Language Processing (ICSLP 04), Jeju Island, Korea, 2004.

[25] N. Tits, K. El Haddad, and T. Dutoit, "Exploring transfer learning for low resource emotional tts," in Intelligent Systems and Applications, Y. Bi, R. Bhatia, and S. Kapoor, Eds. Cham: Springer International Publishing, 2020, pp. 52–60.

[26] J. Huang et al., "Speech emotion recognition from variable length inputs with triplet loss function," in Proc. Interspeech Conf., 2018, pp. 3673–3677.

[27] K. Feng and T. Chaspari, "A siamese neural network with modif ied distance loss for transfer learning in speech emotion recognition," 2020, arXiv:2006.03001.

[28] D. M. Schuller and B. W. Schuller, "Are view on five recent and near-future developments in computational processing of emotion in the human voice," Emotion Rev., vol. 13, pp. 44–50, 2020.

[29] P. Ekman, "About brows: emotional and conversational signals," in Human ethology: claims and limits of a new discipline, M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, Eds. New York, NY, USA: Cambridge University Press, 1979, pp. 169–202.

[30] H. Yehia, P. Rubin, and E. Vatikiotis-Bateson, "Quantitative association of vocal-tract and facial behavior," Speech Communication, vol. 26, no. 1-2, pp. 23–43, 1998.

[31] H. Yehia, T. Kuratate, and E. Vatikiotis-Bateson, "Linking facial animation, head motion and speech acoustics," Journal of Phonetics, vol. 30, no. 3, pp. 555–568, Jul 2002.

[32] X. Zhu, S. Yang, G. Yang, and L. Xie, "Controlling emotion strength with relative attribute for end-to-end speech synthesis," in Proc. IEEE Autom. Speech Recognit. Understanding Workshop, 2019, pp. 192–199.

[33] Y. Lei, S. Yang, and L. Xie, "Fine-grained emotion strength transfer, control and prediction for emotional speech synthesis," in Proc. IEEE Spoken Lang. Technol. Workshop, 2021, pp. 423–430.

[34] K. Zhou, B. Sisman, R. Rana, B. W. Schuller, and H. Li, "Emotion intensity and its control for emotional voice conversion," IEEE Trans. Affective Comput., to be published, doi: 10.1109/TAFFC.2022.3175578.

[35] Y. Lei, S. Yang, X. Wang, and L. Xie, "MsEmoTTS: Multi-scale emotion transfer, prediction, and control for emotional speech synthesis," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 30, pp. 853–864, 2022.

[36] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in Proc. 3rd Int. Conf. Learn. Representations, 2015.

[37] K. K. J. F. S. Kyle, K. A. C. Y. B. Jose, and S. M. Sotelo, "Char2wav: End-to-end speech synthesis," in Proc. Int. Conf. Learn. Representations, Workshop, 2017.

[38] C. Cav´ e, I. Gua¨ ıtella, R. Bertrand, S. Santi, F. Harlay, and R. Espesser, "About the relationship between eyebrow movements and F0 variations," in International Conference on Spoken Language (ICSLP), vol. 4, Philadelphia, PA, USA, October 1996, pp. 2175–2178.

[39] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in Sixth International Conference on Multimodal Interfaces ICMI 2004. State College, PA: ACM Press, 2004, pp. 205–211.

[40] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson, and H. Yan, "Embodiment in conversational interfaces: Rea," in International Conference on Human Factors in Computing Systems (CHI-99), Pittsburgh, PA, USA, May 1999, pp. 520–527.

[41] K. Arun, T. Huang, and S. Blostein, "Least-squares fitting of two 3-d point sets," IEEE Trans. Pattern Anal. Mach. Intell., vol. 9, no. 5, pp. 698–700, September 1987.

[42] E. Vatikiotis-Bateson and H. C. Yehia, "Speaking mode variability in multimodal speech production," IEEE Transactions on Neural Networks, vol. 13, no. 4, pp. 894–899, July 2002.

[43] H. Choi, S.Park, J.Park, and M. Hahn, "Multi-speaker emotional acoustic modeling for cnn-based speech synthesis," ICASSP 2019- 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), p. 6950–6954, 2019.

[44] S.-Y. Um, S. Oh, K. Byun, I. Jang, C. Ahn, and H.-G. Kang, "Emotional speech synthesis with rich and granularized control," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 7254–7258.

[45] M. Schr¨oder, "Emotional speech synthesis: A review," in Seventh European Conference on Speech Communication and Technology. Citeseer, 2001.

[46] Y.Wang,D.Stanton, Y.Zhang, R.-S.Ryan, E.Battenberg, J. Shor, Y. Xiao, Y. Jia, F. Ren, and R. A. Saurous, "Style tokens: Unsupervised style modeling, control and transfer

Current Journal of Engineering and Science Research
Volume 1, Issue 1, pp 10-17, November 2024
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2024.101002
Available Online at https://www.cjesr.com

in end-to-end speech synthesis," in International Conference on Machine Learning. PMLR, 2018, pp. 5180–5189.

[47] O. Kwon, I. Jang, C. Ahn, and H.-G. Kang, "An effective style token weight control technique for end-to-end emotional speech synthesis," IEEE Signal Processing Letters, vol. 26, no. 9, pp. 1383–1387, 2019.

[48] X. Cai, D. Dai, Z. Wu, X. Li, J. Li, and H. Meng, "Emotion controllable speech synthesis using emotion-unlabeled dataset with the assistance of cross-domain speech emotion recognition," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 57345738.

Citation of this Article:

Lalitha M, Kavya S, Lakshmi Priya A, & Prabu Chakravarthy K. (2024). Optimized Algorithm for Human Speech Emotions Recognition. *Current Journal of Engineering and Science Research*. 1(1), 10-17. Article DOI: https://doi.org/10.47001/CJESR/2024.101002

***** End of the Article *****

