**Speech Emotion Recognisation**

**Team\_4019**

**J. Chanakya**

P. Lokesh

B. Venkata Siva Rama Krishna

K. Taraka Sathya Bala Ujwal

K. Guru Vishnu

Team Leader Email-Id:

**Parul Institute of Engineering and Technology**

****

**Abstract:**

Communication is the key to express one’s thoughts and ideas clearly. Amongst all forms of communication, speech is the most preferred and powerful form of communications in human.

Speech emotion recognition (SER) is an important field of research in the domain of human-computer interaction (HCI). The goal of SER is to identify the emotional state of a speaker from their speech signals. The importance of SER arises from the fact that humans use emotions as a primary means of communication. Therefore, the ability of machines to recognize emotions from speech can significantly improve the effectiveness of human-machine interaction.

The aim of this project is to develop a system that can automatically recognize the emotional state of a speaker from their speech signals. The system will be based on machine learning techniques, which involves training a model on a dataset of speech signals labeled with corresponding emotions. The model will then be used to classify the emotions of new speech signals.

The dataset used in this project will contain speech signals from a diverse range of speakers, with varying emotions such as happy, sad, angry, and neutral. The dataset will also include various types of speech, such as spontaneous speech, scripted speech, and speech in different languages.

The system developed in this project will have various applications in real-world scenarios, such as call centers, where the system can be used to detect the emotional state of customers and direct them to appropriate services. The system can also be used in mental health diagnosis and therapy, where it can help identify emotional states and provide appropriate support.

**Introduction:**

For several years now, the growth in the field of Artificial Intelligence (AI) has been accelerated. AI, which was once a subject understood by computer scientists only, has now reached the house of a common man in the form of intelligent systems.

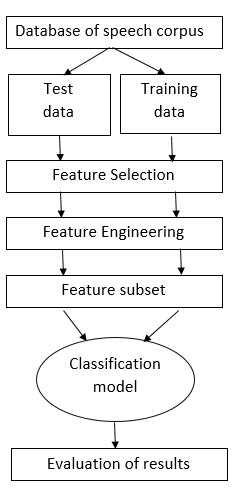
Speech Emotion Recognition (SER) is the process of identifying the emotional state of a speaker from their speech signals. The ability to recognize emotions from speech signals has many potential applications in various fields such as mental health diagnosis, call centers, and human-computer interaction. The problem of SER has been the subject of research for several decades, and machine learning techniques have emerged as a powerful tool for solving this problem.

The main challenge in SER is the high variability of speech signals due to factors such as accent, dialect, intonation, and context. Furthermore, emotions are complex and often subjective, making it difficult to identify them accurately from speech signals. However, recent advances in machine learning, particularly in deep learning, have shown promising results in addressing these challenges.

The solution to the problem of SER involves developing a system that can learn to recognize emotions from speech signals using machine learning techniques. The system will be based on a dataset of speech signals labeled with corresponding emotions, and the model will be trained to identify patterns in the data that correspond to different emotional states. The system will then be used to classify the emotions of new speech signals.

The project will focus on developing a robust and accurate SER system that can handle a wide range of emotional states and speech signals. The system will also be evaluated on its ability to generalize to new speakers and languages. The ultimate goal is to develop a system that can be used in real-world applications, such as call centers and mental health diagnosis, to improve the effectiveness of human-machine interaction and provide better support for individuals.

**Block Diagram:**



**Fig:1 Flow of Implementation**

**Technologies:**

Speech Emotion Recognition (SER) projects typically use a combination of technologies to develop an accurate and robust system. Some of the key technologies used in an SER project include:

**Digital Signal Processing (DSP):** DSP techniques are used for preprocessing audio signals to remove noise, silence, and other unwanted components that can affect emotion recognition accuracy.

**Feature Extraction:** Feature extraction techniques are used to extract relevant features from the audio signal that are indicative of different emotional states. Common feature extraction techniques include Mel-frequency cepstral coefficients (MFCCs), pitch, and energy.

**Machine Learning:** Machine learning techniques are used to train a model to classify different emotional states based on the input features. Common machine learning algorithms used in SER projects include support vector machines (SVMs), random forests, and neural networks.

**Deep Learning:** Deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promising results in SER projects. These techniques can automatically learn relevant features from the audio signal and can handle complex relationships between features.

**Natural Language Processing (NLP):** NLP techniques can be used to process text data such as transcripts of speech signals. These techniques can extract relevant features such as sentiment and emotion from the text data, which can be used in conjunction with audio features for improved emotion recognition accuracy.

**User Interface (UI) and Human-Computer Interaction (HCI):** UI and HCI technologies are used to develop user-friendly interfaces that allow users to interact with the SER system. These interfaces can include graphical user interfaces (GUIs), voice response systems, and chatbots.

Overall, an SER project involves a combination of technologies, including DSP, feature extraction, machine learning, deep learning, NLP, UI, and HCI, to develop an accurate and robust system for recognizing emotions from speech signals.

**Results:**

**Conclusion:**

The emerging growth and development in the field of AI and machine learning have led to the new era of automation. Most of these automated devices work based on voice commands from the user. Many advantages can be built over the existing systems if besides recognizing the words, the machines could comprehend the emotion of the speaker (user).

Some applications of a speech emotion detection system are computer-based tutorial applications, automated call center conversations, a diagnostic tool used for therapy and automatic translation system.

In this project, Initially the limited number of publically available speech database made it challenging to implement a well-trained model. Next, several novel approaches to feature extraction had been proposed in the earlier works, and selecting the best approach included performing many experiments. Finally, the classifier selection involved learning about the strength and weakness of each classifying algorithm with respect to emotion recognition. At the end of the experimentation, it can be concluded that an integrated feature space will produce a better recognition rate when compared to a single feature.

**Future Scope:**

For future advancements, the proposed project can be further modeled in terms of efficiency, accuracy, and usability. Additional to the emotions, the model can be extended to recognize feelings such as depression and mood changes. Such systems can be used by therapists to monitor the mood swings of the patients. A challenging product of creating machines with emotion is to incorporate a sarcasm detection system. Sarcasm detection is a more complex problem of emotion detection since sarcasm cannot be easily identified using only the words or tone of the speaker. A sentiment detection using vocabulary, can be integrated with speech emotion detection to identify a possible sarcasm. Therefore, in the future, there would emerge many applications of a speech-based emotion recognition system.

Overall, the future scope of SER technology is vast and has the potential for significant advancements in various fields, including human-computer interaction, mental health, and social sciences.

**References:**

1. Soegaard, M. and Friis Dam, R. (2013). The Encyclopedia of Human-Computer Interaction. 2nd ed.
2. Developer.amazon.com. (2018). Amazon Alexa. [online] Available at: https://developer.amazon.com/alexa
3. Store.google.com. (2018). Google Home Tips & Tricks – Google Store. [online] Available at: <https://store.google.com/product/google_home_learn>
4. Apple. (2018). iOS - Siri. [online] Available at: https://www.apple.com/ios/siri/
5. The Official Samsung Galaxy Site. (2018). What is S Voice?. [online] Available at:

http://www.samsung.com/global/galaxy/what-is/s-voice/ [Accessed 2 May 2018].

1. Gartner.com. (2018). Gartner Says 8.4 Billion Connected. [online] Available at: https://www.gartner.com/newsroom/id/3598917.
2. H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186–202, Jan. 2015.
3. L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012.
4. T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden Markov models,” Speech Commun., vol. 41, no. 4, pp. 603–623, Nov. 2003.
5. J. Rong, G. Li, and Y.-P. P. Chen, “Acoustic feature selection for automatic emotion recognition from speech,” Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, May 2009.
6. S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.
7. Dupuis, K. and Pichora-Fuller, M. (2010). [Collection] University of Toronto, Psychology Department, Toronto emotional speech set (TESS). Toronto.
8. J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera. KEEL DataMining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. Journal of Multiple-Valued Logic and Soft Computing 17:2-3 (2011) 255-287.
9. S, Khalid, T, Khalil and S, Nasreen. (2014). 2014 Science and Information Conference, A survey of feature selection and feature extraction techniques in machine learning. PP.372-378.
10. Giannakopoulos, T. (2018). pyAudioAnalysis. [online] GitHub. Available at:

https://github.com/tyiannak/pyAudioAnalysis.

1. Practicalcryptography.com. (2018). Practical Cryptography. [online] Available at:

http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstralcoefficients-mfccs/.

1. Dsp.stackexchange.com. (2018). Framing an audio signal. [online] Available at: https://dsp.stackexchange.com/questions/27243/framing-an-audio-signal.
2. Dataminingblog.com. (2018). Standardization vs. normalization | Data Mining Blog - www.dataminingblog.com. [online] Available at: [http://www.dataminingblog.com/standardization-vsnormalization/.](http://www.dataminingblog.com/standardization-vs-normalization/)
3. Statistics.laerd.com. (2018). Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association.. [online] Available at: [https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statisticalguide.php.](https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php)
4. Trevino, A. (2018). Introduction to K-means Clustering. [online] Datascience.com. Available at:

[https://www.datascience.com/blog/k-means-clustering.](https://www.datascience.com/blog/k-means-clustering)

1. DeZyre. (2018). Principal Component Analysis Tutorial. [online] Available at:

[https://www.dezyre.com/data-science-in-python-tutorial/principal-component-analysis-tutorial.](https://www.dezyre.com/data-science-in-python-tutorial/principal-component-analysis-tutorial)

1. George Dallas. (2018). Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction. [online] Available at[: https://georgemdallas.wordpress.com/2013/10/30/principalcomponent-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/.](https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/)
2. Analytics Vidhya. (2018). Simple Guide to Logistic Regression in R. [online] Available at:

[https://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/.](https://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/)

1. En.wikipedia.org. (2018). Logistic regression. [online] Available at:

https://en.wikipedia.org/wiki/Logistic\_regression.

1. Ray, S. (2018). 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python). [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/.
2. En.wikipedia.org. (2018). Bayes' theorem. [online] Available at:

https://en.wikipedia.org/wiki/Bayes%27\_theorem.

1. Ray, S. (2018). Understanding Support Vector Machine algorithm from examples (along with

code). [online] Analytics Vidhya. Available at:

https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/.

1. Srivastava, T. (2018). Introduction to KNN, K-Nearest Neighbors : Simplified. [online] Analytics Vidhya. Available at: [https://www.analyticsvidhya.com/blog/2014/10/introduction-k-neighboursalgorithm-clustering/.](https://www.analyticsvidhya.com/blog/2014/10/introduction-k-neighbours-algorithm-clustering/)
2. Ray, S. (2018). Decision Tree | Predictive Analytics. [online] Analytics Vidhya. Available at:

https://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/2/.

1. (2018). How Random Forest Algorithm Works in Machine Learning. [online] Available at:

[https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning3c0fe15b6674.](https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning-3c0fe15b6674)

1. Brownlee, J. (2018). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning - Machine Learning Mastery. [online] Machine Learning Mastery. Available at:

[https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machinelearning/.](https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/)

1. Exsilio Blog. (2018). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog. [online] Available at: http://blog.exsilio.com/all/accuracy-precision-recall-f1score-interpretation-of-performance-measures/.