e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

ARTIFICIAL SPEAKER FOR DEAF AND DUMB PEOPLE USING THEIR MURMURING SOUND

Manoj V*1

*1PES Institute Of Technology And Management, Shimoga, Karnataka, India.

ABSTRACT

This paper presents an innovative approach utilizing Information and Communication Technology to enhance communication for the deaf & dumb community through the application of murmuring sounds. The proposed system leverages advanced Speech-to-Text (STT) and Text-to-Speech (TTS) techniques, translating subtle vocalizations into readable text and audible speech. By combining Hidden Markov Models (HMM) and Deep Neural Networks (DNN), the system ensures accurate and real-time translation. This solution aims to bridge the communication gap between the deaf & dumb individuals and the general population, facilitating better expression and understanding. The advancement in assistive technology promises To boost the standard of those who are deaf and speech impairments, highlighting With the potential for widespread deployment and future development.

Keywords: Speech-To-Text (STT), Text-To-Speech (TTS), Speech Recognition, Murmuring Sounds.

I. INTRODUCTION

Creating an artificial speaker for the deaf & mute using their murmuring sounds represents a groundbreaking advancement in assistive technology. This innovative project leverages the capacity of advanced speech-to-text (STT) and text-to-speech (TTS) technologies to translate non-verbal sounds into coherent speech, thus enabling efficient interaction between those who are deaf and speech impairments. By capturing murmuring sounds, which are often overlooked in traditional communication methods, and converting them into understandable speech, the system is aimed to build a bridge the communication gap between the deaf & mute community and the rest of society. The technology employs sophisticated algorithms, such as Hidden Markov Models (HMM) for STT and Deep Neural Networks (DNN) for TTS, ensuring high accuracy and natural-sounding speech output. This innovation not only empowers users by giving them a voice but also promotes inclusivity, allowing them to express their thoughts and engage in conversations seamlessly. Additionally, the system's ability to interpret and vocalize murmurs provides a unique and personalized communication tool, enhancing the social interaction & standard of living for those with communication challenges. The development of this technology highlights the significance of leveraging machine learning & signal processing to create solutions that enhance accessibility and foster a more acknowledging society. By integrating these advanced technologies, the artificial speaker stands as a testament to the potential of modern science to transform lives and promote equality in communication.

II. LITERATURE REVIEW

"Hybrid- continuous speech- recognition systems by HMM, MLP and SVM: a comparative study," [1] E. Zarrouk et al (2014) evaluated hybrid- continuous- speech- recognition systems using HMM , Multi-Layer Perceptrons (MLP), and SVM. They found that HMM-SVM systems offered higher accuracy, while HMMMLP models were more computationally efficient presenting important future insights speech recognition advancements.

"Acoustic modeling and feature selection for speech recognition" [2] Y. Zheng's (2005) focused on improving the accuracy of systems for speech- recognition by optimizing acoustic models and selecting the most effective features. This research contributed to enhancing model performance and robustness in various acoustic environments.

"A Novel Model for Speech to Text Conversion" [3] Deepa V et al. (2014) proposed a novel model for speechto-text conversion that uses advanced technology machine learning methods to enhance the accuracy and efficiency of the transcription process. Their work focused on achieving the conversion algorithm to handle diverse speech patterns and accents, significantly improving the performance and reliability various systems for voice recognition.

"Comparison & combination of features in hybrid HMM / MLP and a HMM / GMM voice recognition system," [4]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

P. Pujol et al. (2005) conducted a study comparing and combining features in hybrid HMM/MLP (Hidden Markov Model/Multi-Layer Perceptron) and HMM/GMM (Hidden Markov Model/Gaussian Mixture Model) systems for recognizing speech. Their research highlighted the complementary strengths of these models, demonstrating that combining features from both approaches can enhance overall recognition accuracy and robustness in diverse acoustic environments.

"An investigation of DNNs for noise - robust speech- recognition," M. L. Seltzer et al.(2013) investigated the utilliztion of the DNNs for improving noise robustness in voice recognition systems. Their research demonstrated that DNNs significantly enhance the capacity to precisely identify speech in noisy environments, outperforming traditional models by effectively learning and compensating for noise distortions.

III. METHODOLOGY

The methodology for developing an artificial speaker for deaf & dumb people using their murmuring sound involves several key steps. Initially, the murmuring sounds are captured and preprocessed to filter out noise and enhance the signal quality. This is followed by feature extraction, where specific characteristics of the murmuring sounds, such as frequency and amplitude, are identified and analyzed. The following action involves using a Speech-to-Text (STT) system to translate these sounds into textual data. HMM are engaged during this process for their efficiency in recognizing and processing speech patterns.

For converting the text back into audible speech, a Text-to-Speech (TTS) system is implemented. This system leverages DNN to ensure accurate and natural-sounding speech synthesis. The TTS mechanism is executed in several stages: first, a grapheme-to-phoneme model converts text into phonetic representation; then, a segmentation model determines the boundaries of each phoneme. Following this, the phoneme duration model predicts how long each phoneme, & the fundamental frequency model forecasts the pitch. Finally, the audio synthesis model integrates these elements to produce the final speech output.

This dual-process approach of STT and TTS facilitates real-time communication by translating murmuring sounds into readable text and then converting this text into spoken words, thereby bridging the communication gap for the deaf & dumb community.

1. TTS

The workflow of a Text-to-Speech (TTS) system, divided into two main stages: Training and Inference. Every step includes particular procedures. that collectively enable the system to convert text into natural-sounding speech.

(a) Training Stage

The training stage begins with the input of audio data, which is segmented into phonemes, the basic units of sound. This segmentation process identifies the boundaries and durations of each phoneme within the audio stream. Concurrently, text data is converted into phonemes through a grapheme-tophoneme conversion process, utilizing a phoneme dictionary to map written characters (graphemes) to their corresponding phonetic sounds (phonemes). Once the phonemes are identified, a duration prediction model is trained to accurately predict the how long each phoneme based on the segmented audio. Additionally, the fundamental frequency (F0), which represents the pitch of the sound, is extracted from the audio. A model is then trained to predict this frequency for each phoneme. The final step during the training stage involves audio synthesis, where the phoneme sequence, along with their predicted durations and fundamental frequencies, is utilized to generate the final speech output. This comprehensive training process ensures that the TTS system can effectively learn the nuances of speech, preparing it for accurate text-to-speech conversion during inference.

(b) Inference Stage

In the inference stage, the process begins with the input of new text data. This text undergoes grapheme-to-phoneme conversion, similar to the training stage, where a phoneme dictionary is utilized to translate the text into phonetic sounds. The trained duration prediction model is then employed to predict how long each phoneme for the given text. Concurrently, the trained fundamental frequency (F0) prediction model forecasts the pitch for each phoneme. Finally, the audio synthesis model uses the sequence of phonemes, along with their predicted durations and fundamental frequencies, to generate the corresponding speech output. This inference stage leverages the models trained in the prior stage to convert text into natural-sounding speech efficiently

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

and accurately. By following this structured methodology, the TTS system ensures high-quality speech synthesis, enabling effective communication for users.

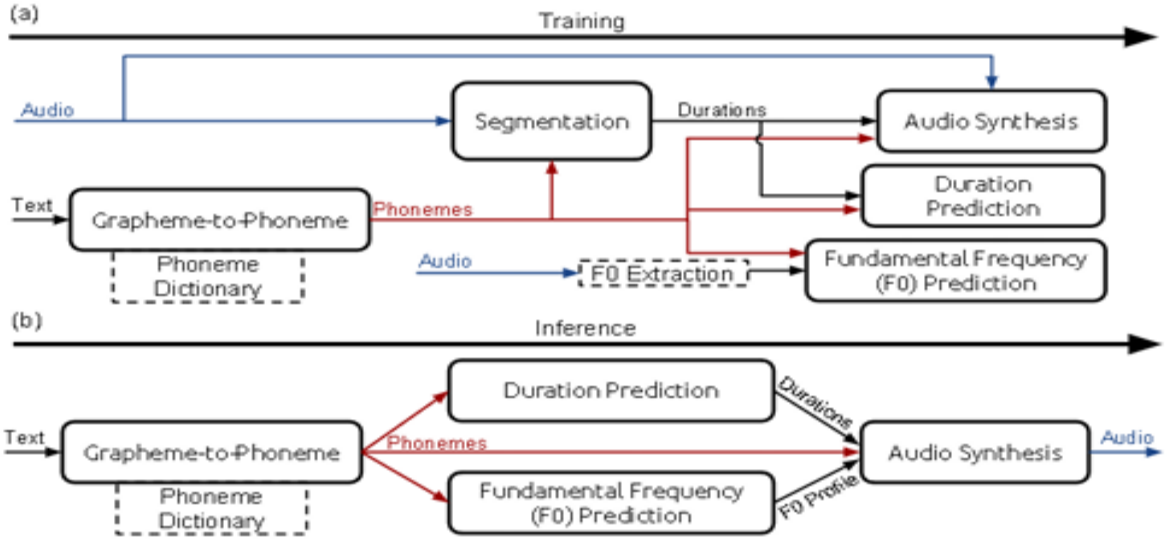

Figure: Flow chart

2. STT

Speech-to-Text (STT) technology is a transformative tool that transforms verbal communication into written text. This technology has significantly advancements in recent years, achieving a high degree of accuracy and efficiency. STT systems typically rely on complex algorithms and models such as HMMs and Deep Neural Networks (DNNs) to process and interpret speech. These models break down speech into smaller components, analyze phonemes, and use statistical methods to predict the most likely corresponding text. The accurateness of STT can be affected by various factors, including the speaker's accent, background noise, and the best quality of recording equipment.

For deaf & dumb community, STT technology offers a substantial improvement in communication. By converting spoken communication or language into text, STT allows individuals who cannot hear or speak to read and understand conversations in real-time. This technology can bridge the communication gap between the deaf & dumb community and those who use spoken language. Implementing STT in mobile devices and applications provides these individuals with greater accessibility and independence. Although STT systems have made great strides, achieving 100% accuracy remains challenging due to variability in human speech and environmental conditions. However, continuous improvements in machine learning and natural language processing are expected to enhance the performance and reliability of STT systems.

IV. RESULTS AND DISCUSSION

Artificial speaker systems designed for deaf and mute individuals utilize advanced training techniques to convert murmuring sounds into comprehensible speech. In the training phase, these systems learn to analyze audio features and map them to phonetic units, predict the durations of phonemes, and estimate fundamental frequency (F0). This comprehensive training ensures that the system can generate natural-sounding speech by learning from paired audio and text data. During inference, when a user inputs text, the system leverages its learned model to predict phoneme durations and F0 values, synthesizing the text into real-time speech output. This capability allows the artificial speaker to effectively interpret murmuring sounds and produce intelligible speech, thereby enhancing communication accessibility for the deaf and mute community.

V. CONCLUSION

The improvement of an artificial speaker for deaf and dumb individuals using their murmuring sounds represents a significant breakthrough in assistive technology. By leveraging advanced Speech-to-Text (STT) and Text-to-Speech (TTS) systems, this technology can effectively translate non-verbal sounds into coherent speech, thereby bridging communication gaps and promoting inclusivity. Utilizing sophisticated algorithms

such as Hidden Markov Models (HMM) and Deep Neural Networks (DNN), the system ensures high accuracy and natural-sounding speech output. This innovation not only empowers individuals by providing them with a voice but also improves their potential to engage in social interactions and express themselves. As machine learning and signal processing technologies continue to evolve, the capabilities of such assistive devices will only improve, fostering a more inclusive society where communication barriers are significantly reduced.

VI. REFERENCE

[1] E. Zarrouk, Y. B. Ayed, and F. Gargouri, "Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study," International Journal of Speech Technology, vol. 17, pp. 223-233, 2014.

[2] Y. Zheng, "Acoustic modeling and feature selection for speech recognition," Citeseer, 2005.

[3] Deepa V.Jose, Alfateh Mustafa, Sharan R,"A Novel Model for Speech to Text Conversion" International Refereed Journal of Engineering and Science (IRJES)ISSN (Online) 2319-183X, Volume 3, Issue 1 (January 2014)

[4] P. Pujol, S. Pol, C. Nadeu, A. Hagen, and H. Bourlard, "Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system," IEEE Transactions on Speech and Audio processing, vol. 13, pp. 14-22, 2005.

[5] M. L. Seltzer, D. Yu, and Y. Wang, "An investigation of deep neural networks for noise robust speech recognition," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 7398-7402.