

Zigbee Based Wireless Voice To Text Translator

Pooja Dinde, Ruchita Hanchina, Sakshi Tale, Milind Kadlag
ENTC Dept, MET Institute of Engineering,
Nasik

Abstract:- Speech is the first important primary need, and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. This paper basically gives an overview of major technological perspective and appreciation of the fundamental progress of speech to text conversion and also gives complete set of speech to text conversion based on Zigbee wireless communication. A comparative study of different technique is done as per stages. This paper concludes with the decision on future direction for developing technique in human computer interface system in mother tongue and it also discusses the various techniques used in each step of a speech recognition process and attempts to analyze an approach for designing an efficient system for speech recognition. However, with modern processes, algorithms, and methods we can process speech signals easily and recognize the text. In this system, we are going to develop an speech-to-text engine. However, the transfer of speech into written language in real time requires special techniques as it must be very fast and almost 100% correct to be understandable. The objective of this paper we use a speech recognition system which recognizes the spoken command by the user and compares it with already existing database and also Zigbee wireless communication between two systems.

Keywords: *Speech Recognition, Communication, Algorithm*

INTRODUCTION:

Human interact with each other in several ways such as facial expression, eye contact, gesture, mainly speech. The speech is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. Speech-to-text conversion (STT) system is widely used in many application areas. Text-to-speech (TTS) convention transforms linguistic information stored as data or text into speech. It is widely used in audio reading devices for blind people now a day's. In the last few years however; the use of text-to-speech conversion technology has grown far beyond the disabled community to become a major adjunct to the rapidly growing use of digital voice storage for voice mail and voice response systems. Also developments in Speech synthesis technology for various languages have already taken place. In this project we use a speech recognition system which recognizes the spoken command by the user and compares it with already existing database and also Zigbee wireless communication between two systems. Speech is the primary and most convenient means of communication between humans. Whether due to technological curiosity to build machines that mimic human's or desire to automate work with machine, research in speech recognition as a first step towards human-machine communication. Speech recognition is the process of recognizing the spoken word to take necessary actions

accordingly. ZigBee is a wireless technology developed as an open global standard to address the unique needs of low-cost, low-power, wireless sensor networks. Zigbee is the set of specs built around the IEEE802.15.4 wireless protocol. As Zigbee is the upcoming technology in wireless field, we had tried to demonstrate its way of functionality and various aspects like kinds, advantages and disadvantages using a small application of controlling the any kind of electronic devices and machines. The Zigbee technology is broadly adopted for bulk and fast data transmission over a dedicated channel.

LITERATURE SURVEY:

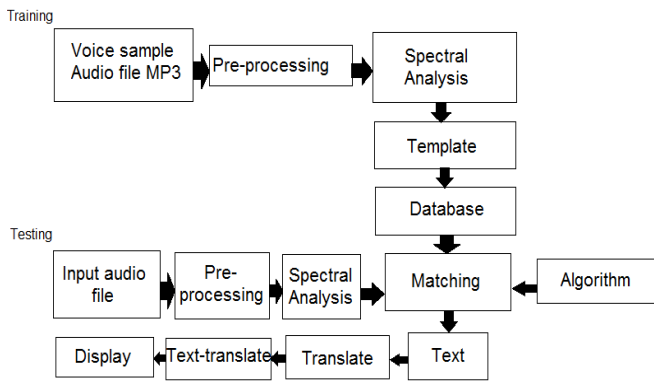
The conversion of speech signal to words in a orderly manner by means of algorithm applied as a machine program is termed as speech recognition. The objective of speech recognition area is to make changes and develop a system for speech input to a machine based on progress of statistic modeling of speech. [1] the most extensive and prominent way to extract spectral features is determining Mel-frequency cepstral coefficients (MFCC).MFCC used in speech recognition depend on frequency domain using Mel scale that intern depends on human ear scale. MFCC exhibits real cepstral of a windowed short time signal which is evolved from fast Fourier transform (FFT). Audio feature extraction technique based on MFCC extracts characteristics from identical speech to once that are preowned by human for hearing speech whereas parallely all other information's are neglected.[2] Hermansky developed a model called he perceptual linear prediction (PLP).In PLP the concept of psychophysics of hearing is used that models human speech.PLP enhances speech recognition rate by removing unjustified information of the speech .to equate characteristics of human auditory system changes have been brought in spectral characteristics which is the only difference between PLP and LPC technique.[2][3] PNCC, a front end technique which is slightly different than MFCC,which uses gamma tone filters instead of Mel-scale transformation [5] Imitating the performance of cochlea. To increase robustness, term called *medium time power bias removal* is included further. The evaluate the quality reduction of speech due to noise the arithmetic to geometric mean ratio is calculated by bias vector.[4] Speech to text engine transforms speech to text from an instant voice input, complementing users giving better ideas for a different choice of data entry. HMM are used to perform speech to text conversion. HMM generates stochastic models from a known statement and in contrast with the possibility that the unknown statement was generated by each model. A speech signal

can be observed as a piece wise stationary signal or a short time stationary signal, hence HMM are thereby used in speech to text conversion.[6]

A parametric density function that indicates weighted sum of Gaussian components densities is referred as Gaussian Mixture Model (GMM).To compare the feature extracted from the model with stored template, Gaussian Mixture Model is used. Representation of Gaussian Mixture model is done with the help of Gaussian distribution which is thereby calculated by its mean, variance and weight of the same.[7]

Machine translation or MT comes under computational linguistic that examines use of software to translate text from source language to target language. A translation from an intermediate representation that imitates the meaning of original sentence is created in transfer based machine translation which is similar to interlingual machine translation. In this knowledge of the source and target languages are used to evaluate its grammatical structure, transferring that to a structure appropriate for developing text in a target language, and thereby obtaining the desired text.[8] Dictionary based MT uses dictionary entries, similar to that of a normal dictionary-word by word, generally with not much correlation of meaning amongst them. Morphological analysis of lemmatization may or may not be used in dictionary lookup. Whereas this approach to machine translation is least polished, but to translate long lists of phrases on a significant level, dictionary based MT is Ideal. [9]

SYSTEM IMPLEMENTATION:



Speech Recognition is the process of automatically recognizing a certain word spoken by a particular speaker based on individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify his/her identity and provide controlled access to services like voice based biometrics, database access services, voice based dialing, voice mail and remote access to computers. Speech recognition basically means talking to a computer, having it recognize what one is saying. There are many types of features, which are derived differently and have good impact on the recognition rate. This project presents one of the techniques to extract the feature set from a speech signal, which can be used in speech recognition systems. Speech recognition system

performs two fundamental operations: signal modeling and pattern matching.

(1)Pre-processing:

Before recognition, first the speech signal is preprocessed. It consists of pre-emphasis, end point detection, framing and windowing. Speech recognition is performed in 4 stages:

- Analysis
- Feature extraction
- Modeling
- Matching

(2) Analysis

It deals with the stage with suitable frame size for segmenting speech signal for further analysis and extraction

There are 3 techniques of perform analysis of speech signal: Segmentation analysis, Sub segmental analysis, Supra segmental analysis

Supra segmental analysis:

In this analysis speech signal is analyzed using behavior character of speaker. This paper refers this technique for analyzing the speech signal as impulse speech signal is provided as input. The analysis takes place using spectrum analysis of two parameters:

- i. Amplitude
- ii. Frequency

(3) Feature Extraction:

Converting the sound waves into a parametric representation is a major part of any speech recognition approach. Here both static and dynamic features of speech used for speech recognition task because the vocal track is not completely characterized only by static parameters. For this various algorithm are available such as MFCC, PLP, and PNCC.

This paper refers PNCC algorithm of feature extraction PNCC:

Power Normalized Cepstral Coefficient is a front end technique which is slightly different than MFCC. It uses gamma tone filters instead of Mel-scale transformation. This algorithm provides Substantial improvement in accuracy as compared to MFCC and PLP.PNCC processing requires only about 33 percent more computation compared to MFCC.

(4)Modeling:

Modeling refers to Generating a speaker model using speaker specific feature vector. This also includes various types of approaches such as Acoustic phonetic Pattern recognition, Template based, Dynamic time wrapping, Knowledge based, Statistical based, Learning based.

In this paper template based approach for modeling is used In Template based approach Unknown speech is compared against a set of pre-recorded words (Templates) in order to find the best Match. This has the advantage of using perfectly accurate word models.

(5) Matching:

For quick and accurate automatic voice recognition technology, the digital processing of speech and voice recognition algorithms is considered to be very essential. A study of speech recognition shows that the speaker recognition is categorized as: speaker recognition and speaker identification.

Consider a sequence of feature vector $\{x_1, x_2, \dots, x_i\}$ that represents voice of an unknown speaker in the speaker recognition phase. These sequences are compared with codes from the pre-defined database. Euclidean distance is there distortion distance measured between two vector set which aids in identifying the unknown speaker. The formula used to calculate the Euclidean distance can be defined as following:

The Euclidean distance between two points

$$P = (p_1, p_2, \dots, p_n) \text{ and } Q = (q_1, q_2, \dots, q_n),$$

$$= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

The unknown speaker is identified by choosing the speaker with lowest distortion distance.

CONCLUSION:

In this paper basically speech to text conversion and then text translation is done. Speech-to-Text conversion system is implemented by using the PNCC for feature extraction as it provides substantial improvement in accuracy and for recognition Template based approach is adopted which gives easiness. In speech database, few audio files are recorded and these are analyzed to get feature vectors. These features are initially modeled in template based approach and after that the test spoken word is compared with the forward algorithm of template. Here for wireless communication purpose ZIGBEE technology is referred as it is broadly adopted for bulk and fast transmission.

REFERENCES:

- [1] Santosh K Gaikwad, Bharti W Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique; University Aurangabad, Maharashtra."
- [2] Namrata Dave, "Feature Extraction Methods LPC, PLP, MFCC in Speech Recognition; G.H. Patel College of Engineering, Gujrat".
- [3] h. Hermansky, "Perceptually linear predictive analysis of speech," Acoustical Society of America Journal, Vol 87, PP 1788-1752, Apr 1990
- [4] Gellert Sarosi, Mihaly Mozsary, Peter Mihajlik, and Tibor Fegyó, "Comparison of Feature Extraction Methods for Speech Recognition in noise free and traffic noise environment, Dept of Telecommunication and Media Informatics Budapest University of Technology and Economics Budapest, Hungary, THINKTech Research centre non profit LLC, Aitia International Inc".
- [5] R.D. Patterson, K. Robinson, J. Holdsworth, D. MCKrown, C. Zhang & M.H. Allerhand, "Complex Sounds and auditory images," in Pergamon Press, Oxford PP. 429-446, 1992
- [6] D.B. Paul, "Speech Recognition using HMM", The Lincoln Laboratory Journal, Vol 3, Number 1 (1990)
- [7] Virendra Chauhan, Shobhana Dwivedi, Pooja Karale, Prof. S.M. Potdar, "Speech To Text Converter using Gaussian Mixture Model; Sinhgad Academy of Engineering - Pune"
- [8] Jurafsky Danie; Martin, James H (2009). "Speech and Language Processing. PP. 906-908"
- [9] Uwe Muegge (2006), "AN Excellent Application for Crummy Machine Translation; Automatic Translation of a large Database in Elizabeth Grafe, Proceedings of the annual conference of the German society of Technical Commutator, Stuttgart; Telkom, 18-21