International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 7; Issue 1(I); January 2018; Page No. 9311-9313
DOI: http://dx.doi.org/10.24327/ijcar.2018.9313.1534

Research Article

ACOUSTIC MODELLING ALGORITHMS FOR SOUND RECOGNITION: NLP

Suman P. Gouda., Mallamma V. Reddy., Gayatri P. Patil and Nishali R. Nandarage

Department of Computer Science, Rani Channamma University, Belagavi-591156, India

A R T I C L E I N F O

Article History:
Received 11th October, 2017
Received in revised form 10th November, 2017
Accepted 26th December, 2017
Published online 28th January, 2018

Key words:

Acoustic model, Phoneme, Phonetics

A B S T R A C T

Speech recognition has become an essential and considered as the exciting field in natural language processing, where the accuracy and acceptance of speech recognition has come a long way in the last few years. Speech recognition is the method; machine identifies word or phases in spoken language. Speech recognition focus on phonetics is a branch of linguistics. Phonetics is the study of phoneme or sound system of any language and it has three different aspects such as Articulator phonetics, Acoustic phonetics, and auditory phonetics. This paper presents acoustic modeling algorithms for sound identification along with the results particularly for Kannada, English and Telugu.

INTRODUCTION

Natural language processing (NLP) [1] is special area of research in computer Science which is used to identify and understand the languages such as text or speech to perform useful tasks. NLP gathers the information how humans speak so that certain tools are developed to make machine understand. A human express their thoughts orally to one another by moments that alter basic tones created by voice into specific sound. Speech can be produced by any muscular action produced by head, neck, chest and abdomen. Generally, there are three categories in automatic speech recognition [2]:

- Articulator approach: Describes how vowels and consonants are produced in lips, tongue and throat.
- Acoustic phonetics approach: Describes physical sound transmission from speaker to listener.
- Auditory phonetic approach: Describes the perception of speech through brain

Speech recognition we focus on acoustic phonetics approach [3], where it uses waveform technique for spoken words (audio), are recognized and then mapped to the stored database to identify language by analyzing sound waves signals such as finding frequencies, amplitude, and duration of speech. One way to examine the acoustic [4] property is by considering the waveform produced during the speech. Speaking the air particles are compressed and rarefied producing sound waves outwards which can be plotted in the graph as shown in Figure.1. The system architecture of the speech recognition as shown in Figure.2.

Corresponding author:* **Suman P. Gouda
Department of Computer Science, Rani Channamma University, Belagavi-591156, India

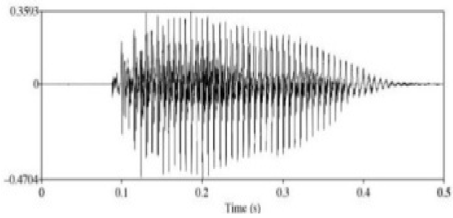

Fig.1 sound waves with respect to time

The system architecture of the speech recognition

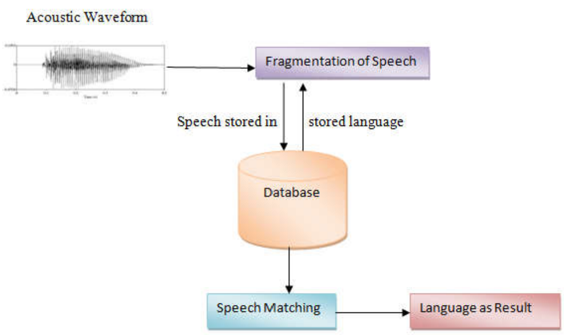

Fig.2 The system architecture of the speech recognition

Algorithm

Step1: Speech is produced by the articulators.
Step2: create the waveform of speech using microphones.
Step3: Store this waveform in the phonetic transcription Dictionary which consists of respective languages.
Step4: Waveforms are mapped with the database.
Step5: If mapping is successful then name of the language produced else it is considered to be invalid.

Speech recognition process

Speech Recognition consists of various steps which are involved in the speech recognition. Figure.3. Shows various steps involved in Speech Recognition.

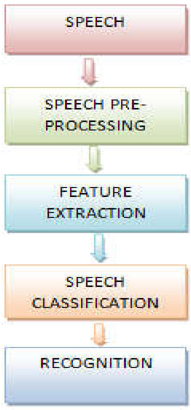

Fig. 3 Process of Speech Recognition

Speech

Speech is the vocalized form of human communication. Each word in the speech is created by the combination of vowels and consonant (phonemes).In this step, the speech of human is received in the form waveform. To record this waveform there are various software available. The acoustic environments have great effect on the speech generated.

Speech Pre-processing

Speech Pre-processing helps to eleminate the variation in the speech and also improves the accuracy of the speech recognition. Speech Pre-processing involves noise filtering, smoothing, end point detection, framing, windowing, echo removal.

Feature Extraction

Speech of every person varies from one person to another because each person has different characteristics and utterance of the voice varies. Theoretically, speech can be recognized using digitized waveform. But due to large variance in the speech there is a need to perform some feature Extraction. There are various feature extraction technologies in speech processing such as MFCC (Mel Frequency Cepstrum Coefficients), LPC (Linear Predictive Coding).

Speech Classification

There are various classification techniques used to classify the speech which helps in taking out the hidden information from the input signals. Few of these techniques are as follows HMM (Hidden Markov Modeling), DTW (Dynamic Time Warping), VQ (Vector Quantization).

Using Models and algorithms of the speech recognition

Hidden Markov Model

Hidden Markov Model [5] provides an effective frame work for modeling vector sequence. As a result, almost all speech recognition systems are based on HMM. It is easy and sensible to use. In speech recognition, the Hidden Markov Model gives the sequence of n-dimensional vectors as output. These vectors consists of cepstral analysis of speech, which is obtained by taking Fourier transform of a signal and decorrelating spectrum using cosine transform.

Dynamic Time Warping (DTW)

Dynamic Time Warping is well-known algorithm used to find optimal alignment between two sequences of series which is time and speed dependent. These sequences are warped in a non-linear mode to match one another. This can be applied to audio, video and any other data that can be changed to linear representation can be examined with DTW [6]. DTW is used to compare different speech patterns using automatic speech recognition. The sequence alignment method is often used in Hidden Markov Model. Time alignment of two sequences which are time-dependent Aligned points are indicated by arrow is shown in Figure.4. Structure of recognizer is shown in Figure.5.

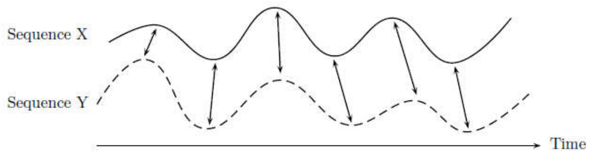

Fig. 4 Time alignment of two sequences which are time-dependent. Aligned points are indicated by arrow.

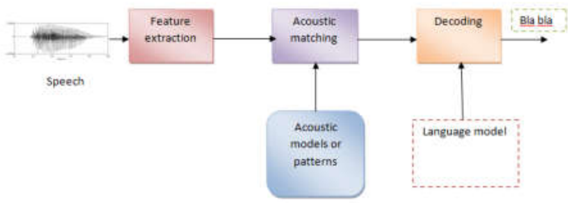

Fig. 5 Structure of recognizer.

Neural network

Neural network [7] appears as an acoustic modeling approach. It can be used in many aspects of speech recognition such as speaker adaption, isolated words recognition, phoneme classification. Neural network doesn't make any assumptions about feature statistical properties and they have certain special properties which helps them to make recognition model for speech recognition.

Sampling Theory

When a person's analog speech signal is recorded using microphone through the computer, the quality of speech signal will directly decide quality of speech recognition. Produced data quality will be decided by sampling [8] frequency. This analog signal consists of lot of different frequencies. Higher sampling frequency gives the better sampled result.

Cross-correlation Algorithm

When a person's speech is recorded, each word spoken by the speaker has different frequency bands due to different vibrations of vocal cord. And shape of spectrum also differs, so to find speech recognition, we need to compare spectrum between first two signals and third recorded signal. Then we can find out which two signals better matches the recorded signal, with the help of system we can judge which reference word is recorded for the third time. The Cross-correlation [9] between two signals can be found using cross correlation.

Auto-correlation Algorithm

The technique of correlation is used to find the similarity between two signals produced by vocal cord of speaker. The Auto-correlation [10] is also correlation technique which tells us how much similar the voice are. Example: If a person has cold. So his voice will be bit different from normal voice. With the help of auto-correlation we will find similarity between two voices even if there is something wrong with current voice of same person. It is because correlation is integration and the result is the area which lies under the curve. So while checking with cold voice few patterns will be similar to original voice and we get a high value of area. Similarity between two voices depends on the value of integration

CONCLUSION

Speech recognition is the most integrated areas of machine intelligence, since humans depend on speech for daily activity. This paper carries out the attempt to provide a review of how much this technology of Speech Recognition has progressed in previous years. This paper presents list of techniques with their properties for future extraction. The work will be continued for implementing the above mentioned algorithms and further same will be applied.

References

1. https://en.wikipedia.org/wiki/Natural_language_processing
2. http://support.docsoft.com/help/whitepaper-asr.pdf
3. http://www.phonetik.unimuenchen.de/~jmh/research/papers/acoustics.pdf
4. https://en.wikipedia.org/wiki/Acoustic_model
5. Dr.V.Ajantha Devi" An Analysis on Types of Speech Recognition and Algorithms"
6. Dr.V.Ajantha Devi" An Analysis on Types of Speech Recognition and Algorithms"; https://www.coursehero.com/file/17693945/9783540740476-c1-3pdf/
7. http://andrew.gibiansky.com/blog/machine-learning/speech-recognition-neural-networks/
8. www.springer.com/cda/content/document/cda.../9783642195853-c2.pdf?SGWID...
9. https://www.researchgate.net/publication/255656750_Speaker_Recognition_Using_Spectral_Cross-correlation_A_Fast_Algorithm
10. https://www.quora.com/what-is-the-use-of-auto-correlation-in-speech-recognition

How to cite this article:

Suman P. Gouda *et al* (2018) 'Acoustic Modelling Algorithms For Sound Recognition: NLP', *International Journal of Current Advanced Research,* 07(1), pp. 9311-9313. DOI: http://dx.doi.org/10.24327/ijcar.2018.9313.1534

****