

Voice Identification Secure System by Statistical Model of Speech Signal Using Normalization Technique

Jitendra Jangir¹, Bablu Kumar Singh², Mohd.Insaf Ali³

^{1,3}Jodhpur National University, Rajasthan,India.

²JIET Jodhpur,Rajasthan,India.

Abstract

This paper is based on the characteristic, analysis and processing on human speech signal for generation of voice identification system using spectra correlation and real time normalization method. Speech spectrogram depicts short-term variation in intensity, frequency and magnitude in graphical form thus these contains provide much useful information about voice identification. When two user speak the word, their pronunciation is similar but not identical thus spectrogram of their speech will show similarities and differences. Many method are available for processing and recognition of speech like Hidden Markov Model, multi-space distribution (MSD) based tone modelling and using quantization method but this proposed normalisation method is simple, less time consuming, highly accurate process for real time speech signal. The performance of robust speech recognition mainly degraded whenever the speech signal is effected by any noise. It is required to improve the stability of speech against noise for robust recognition, which is focusing on all major levels of speech recognition: feature extraction, feature enhancement, and speech modelling.

Keywords: Speech Recognition Technique, Feature Extraction, Normalization Technique, Cepstral Analysis, Spectrogram.

1. Introduction

The identity claim of a speaker in the case Speech Recognition is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program[1]. Speech is produced by human when air is forced from the lungs which is works as power supply and flow through the vocal cords and along the vocal tract. It introduces short-term correlations into the speech signal, and can be thought of as a

filter with broad resonances called formants. The frequencies of these formants are controlled by varying the shape of the tract, for example by moving the position of the tongue.

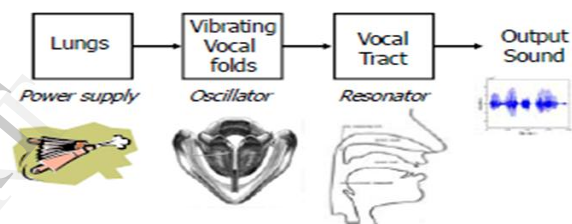


Figure 1. Speech is produced by human

Speech signals captured by microphones are corrupted by various noise sources. Speech enhancement, i.e., improving the quality of degraded speech, has many applications such as speech communications and man-machine interaction. Despite more than three decades of research, speech enhancement algorithms are not robust to different operating conditions [5]. the various characteristics of speech

Acoustic: Use spectral feature conveying vocal tract information. *Prosodic:* Use feature derive from prosody (pitch, energy tracks) to characterize speaker specific prosodic pattern. *Phonetic:* Use phone sequence to characterize speaker specific pronunciation and speaking patterns. *Idiolect:* Use word sequences to characterize speaker specific use to word pattern. *Linguistic:* Use linguistic pattern to characterize speaker specific conversion style [7]

Speech processing still covers an extremely broad area, which relates to the following three engineering applications:

- i) Speech Coding and transmission that is mainly concerned with man-to man voice comm.
- ii) Speech Synthesis which deals with machine-to-man communications.

iii) Speech Recognition relating to man-to-machine communication [3]

2. Spectrogram and Cepstrum Analysis

Very useful method to recognize and finding parameter of voice using spectrogram. Spectrogram is a graph with three geometric dimensions the horizontal axis represents time, the vertical axis is frequency; a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity or colour of each point in the image[4].

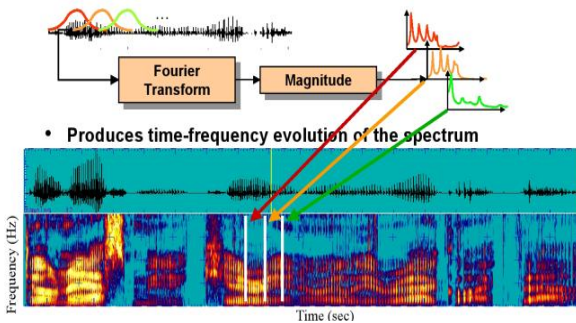


Figure 2. Representation of spectrogram

Pitch estimation "Cepstrum" is a play on the word spectrum as one might suspect and is simply a spectrum of a spectrum. The original time signal is transformed using a Fast Fourier Transform (FFT) algorithm and the resulting spectrum is converted to a logarithmic scale. This log scale spectrum is then transformed using the same FFT algorithm to obtain the power cepstrum. The power cepstrum reverts to the time domain and exhibits peaks corresponding to the period of the frequency spacing's common in the spectrum[2].

3. Basic Model of Speech recognition

All In fundamental method of speech recognition using pattern matching in fig1 it converts the analog speech waveform to digital form using an A-to-D converter and then the feature analysis module which is converts the sampled speech signal to a set of feature vectors. This technique is used for speech coding and derive the feature vectors[6].The segment of block in the system, namely the pattern matching, dynamically time aligns the set of feature vectors representing the speech signal with a concatenated set of stored patterns, and chooses the identity associated with the pattern which is the closest match to the time-aligned set of feature vectors of

the speech signal. The symbolic output consists of a set of recognized words, in the case of speech recognition, or the identity of the best matching talker, in the case of speaker recognition, or a decision as whether to accept or reject of speaker verification.

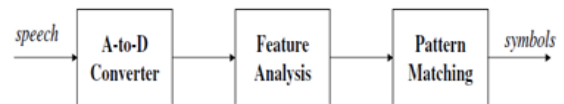


Figure 3. Basic Speech recognition model using Pattern matching

4. Voice Identification Secure System using normalization method

In this voice identification secure system, initially number of speech sample of particular user is stored using microphone into memory and it is created the reference data base for speech identification modal after that the run time of same user sample is applied through microphone. It extracts the feature from run time recorded audio signal using normalization of absolute value of Fourier transform, this test signal compared with feature's of data base and it creates a error during the pattern matching. Pattern matching will give the array of error and it compared with reference or threshold value. Error which is satisfied the rule will be accepted else rejected.

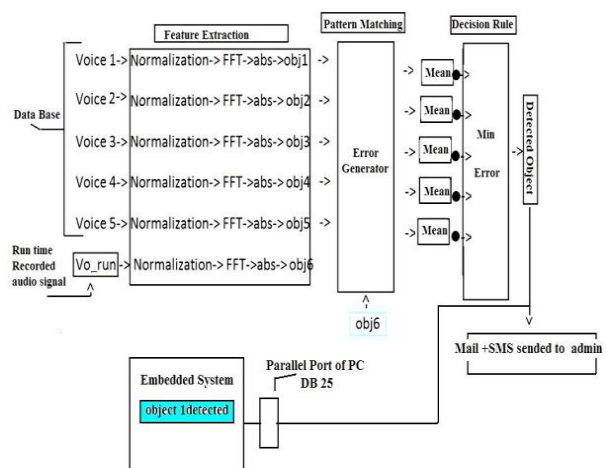


Figure 4.Voice identification system using normalization method

User interface of VIS contains the following GUI panel

- i.) Password protected initialization.
- ii.) Speech identification.
- iii.) Peripheral testing and measurement.
- iv.) Error comparison of speech signal.

4.1 Password protected initialization

Password protected initialization panel of access also have a ability to report the administrator when invalid password appear more than five times It also provides the information about number of invalid entry identification system provided a security for user

4.2 Speech Identification

Audio visual facility available for creating the data base means during the recording of sound signal from the microphone simultaneously it is captured as in figure 5

After the data base generation using the number of recorded speech sample and now it record the run time user speech sample for identification from the data base.

- a.) After the successful recognition it sends the mail and SMS of detected object to administrator.
- b.) It provides the interfacing with embedded system using parallel port for monitoring
- c.) It generates the following plots of speech signal
 - i.) Wave form of recorded signal.
 - ii.) Magnitude curve of recorded signal.
 - iii.) Spectrogram of recorded signal.
 - iv.) Cepstrum analysis of recorded signal.
 - v.) Curve of Fourier coefficient in complex plain.

4.3 Peripheral testing and measurement

It provides a information about the connected peripherals like microphone and webcam. It have a measurement section which is provides following information

- a. Sampling rate of recorded speech signal.
- b. Number of bit for encoding.
- c. Channel information (Dual/Mono).
- d. Pitch of the recorded Speech signal.
- e. Peak value in cestrum analysis.

4.4 Error comparison of speech signal

It provides the error difference of all samples which is in data base w.r.t. reference sample and selected the best recognition using decision making device using threshold

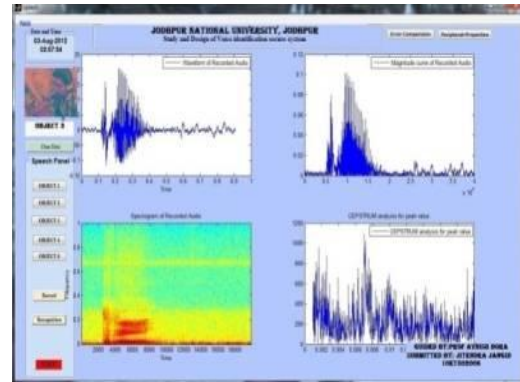


Figure 5. Speech signal GUI panel

5. Result

Recognition of predefine recorded sample from data base using correlation method having 100% matching but this process not valid for dynamic detection.

Table 1. Recognition form correlation method

S/No.	No. of Speech Sample	No. Matching process	Percentage of Successful Identification
1	2	30	100
2	3	30	100
3	4	30	100
4	5	30	100

Recognition of run time recorded sample from data base using normalization method.

Table 2. Recognition using normalization method

S/No.	No. of Speech Sample	No. of process	Percentage of Successful Identification
1	2	30	98
2	3	30	92
3	4	30	89
4	5	30	85

When we used the filter during the process of recognition we observe the detection is improved as compared to without filter

Table 3. Using normalization method using filter

S/No.	No. of Speech Sample	No. of process	Percentage of Successful Identification
1	2	30	98
2	3	30	94
3	4	30	96
4	5	30	94

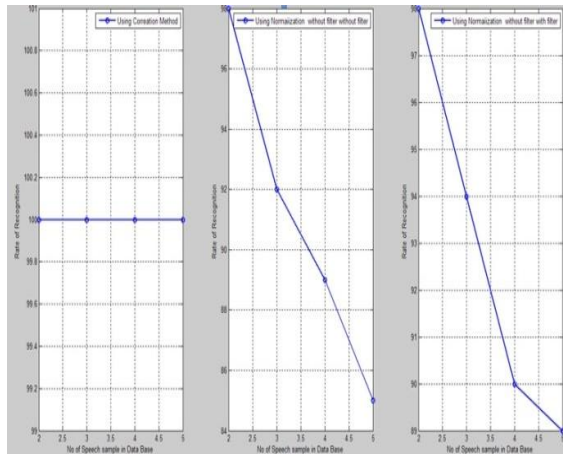


Figure 6. Recognition rate v/s no. of sample.

When we are using the filter for processing it gives the improvement in identification for efficiency using normalization method.

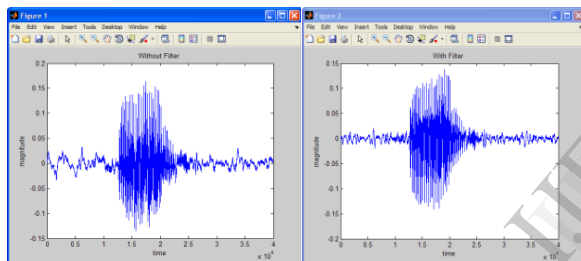


Figure 7. Recording using without and with filter enhanced signal

6. Conclusion

DSP used in a wide range of everyday applications are speech coding, speech synthesis and recognition, Image security system, Adaptive filtering for voice signal. By Fourier Transform we can easily switch between time-space domain and frequency domain so applicable in many other areas. correlation method have higher recognition rate but static process represents its limitation which is overcome by normalization method. Normalization method provide a dynamic performance at the time of recognition .Which can further improved by filter process for noise reduction during recording of sound using microphone.

7. References

- [1] S S. Bhabad, Gajanan K. Kharate, " An Overview of Technical Progress in Speech Recognition" ,*International Journal of Advanced Research in Computer Science and Software Engineering*, Pune university India , Volume 3, Issue 3, March 2013 ISSN: 2277 128X.
- [2] Robert B Randall, " A History of Cepstrum Analysis and its Application to Mechanical Problems", *International Conference* at Institute of Technology of Chartres, France, October 29-30, 2013,pp 11-16
- [3] A.Girish Kumar et al, " A New Technique for Perceptual Distortion Measure on A Spectro Temporal Auditory Model,"*IOSR Journal of Electronics and Communication Engineering*", Volume 8, Issue 5, Nov. - Dec. 2013,pp-10-16.
- [4] Rohini R. Mergu and Dr.Shantanu K. Dixit, " Multi-Resolution Speech Spectrogram ", *International Journal of Computer Applications* (0975–8887), Volume 15–No.4, February 2011,pp 28-32
- [5] M. Benzeghiba et al, "Automatic speech recognition and speech variability: A review", *Journal speech communication*,netherands,Oct,2007, pp 763-786
- [6] L. R. Rabiner and R. W. Schafer, "Introduction to Digital Speech Processing", now publishers Inc., USA, 2007
- [7]http://wwwobile.ecs.soton.ac.uk/speech_codec/speech_properties.html