Available online at http://journalijces.com

INTERNATIONAL JOURNAL OF CURRENT ENGINEERING SCIENCES

RESEARCH ARTICLE

ISSN: 2320- 9046

International Journal of Current Engineering Sciences- Vol. 6, Issue, 4, pp. 27-29, April, 2016

SPEAKER IDENTIFICATION SYSTEM USING PITCH AND FORMANT ANALYSIS

Vinod Patil[1*]., Ankit Anand[2] and Rahul Goyal[3]

Department of Electronics and Telecommunication, BVDU College of Engineering, Pune-Satara Road, Pune - 411043

A R T I C L E I N F O

Article History:

Received 26th January, 2016
Received in revised form 25th February, 2016
Accepted 24th March, 2016
Published online 26th April,2016

Key words:

Feature Extraction, Speaker Identification, Formant, Pitch frequency, Cepstrum, LPC, Emotion

A B S T R A C T

The aim of this paper is to design a Speaker Identification system using speech features like formant and pitch analysis. The pitch and formants are first extracted from the speech signal and then their analysis is carried out to recognize different emotional states of the person like happy, sad and neutral. Here we have used Cepstral Analysis Method for pitch extraction and LPC Method for formant extraction.

INTRODUCTION

Speech is a means of communication and exchange of thoughts between individuals. Speaker Identification (SI) refers to the process of identifying an individual by extracting and processing information from his/her speech. It is a task of finding the best-matching speaker for unknown speaker from a database of known speakers. The speaker characteristics are identified from speech data and are analyzed using suitable analysis techniques. The analysis technique aims at selecting proper frame size along with some overlap and extracting the relevant features from speech. Speech signal are partitioned into voiced speech segment (pitch of voiced speech) and unvoiced speech segments (random noise with no periodicity). Some parts of speech which are neither voiced nor unvoiced are called transition segments. The speech signal can be analysed either by

- Time Domain Method
- Frequency Domain Method

There are many features used to identify the speaker from a set of known speakers like intensity, duration, fundamental frequency, pitch, formant, spectral variations and wavelet based features. In this paper we basically use two features namely pitch and formant to identify speaker and its different states of emotions like happy, sad and neutral.

Speaker Identification System

It is the process of determining which registered speaker provides a given utterance. In the speaker identification task a voice print of an unknown speaker is analyzed and then compared with speech samples of known speakers. The unknown speaker is identified as the speaker whose model best matches with input model.

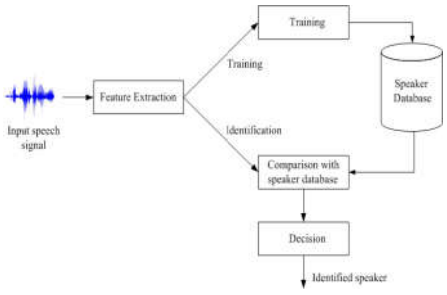

Figure 1 Speaker Identification System

In the Speaker identification systems generally operate without the user's knowledge. It can be applied in routing users to the correct mailbox, identifying talkers in a discussion, alerting speech recognition systems of speaker changes, checking if a user is already enrolled in a system, etc. Speaker identification is also used for police purposes. For instance, a criminal's voice can be cross checked against a database of criminals' voices looking for a match, and ergo the identity. Speaker identification problems consist in

Corresponding author:* **Vinod Patil
Department of Electronics and Telecommunication, BVDU College of Engineering, Pune-Satara Road, Pune - 411043

- ❖ Verification distinguishing multiple speakers during a conversation.
- ❖ Identifying an individual's voice based upon previously supplied data regarding that individual's voice.

Feature Extraction

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behaviour. Before this process we perform pre processing task which involves truncation, frame blocking, windowing and short term Fourier transform. Now we extract two features namely pitch and formant to identify the speaker and its different states of emotions.

LPC Based Formants Estimation Technique

Formants are defined as the resonance of vocal tract and estimation of their location and frequencies at their location which is important in emotion (disgust, neutral, happy, sad) recognition. Formant is the spectral peak of sound spectrum, of the voice, of a person. The peaks of the vocal tract response in each configuration correspond roughly its formant frequencies [1, 2, 3]. For voice and periodic speech the vocal tract can be modelled by a stable all-pole model. The resonance or peaks of the vocal tract transfer function correspond roughly to the formant frequencies of particular sound.

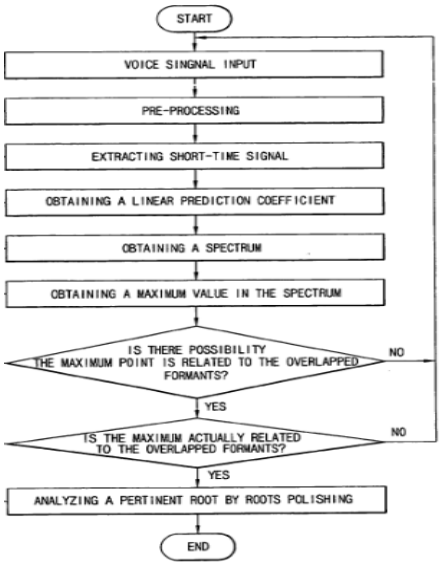

Figure 2 Flowchart For Finding Formant Frequency

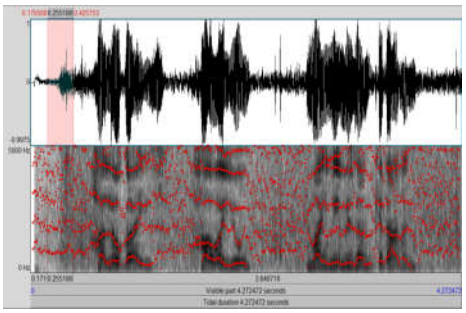

Figure 3 Speech signal and its Formants

Above figure shows the formant frequencies plot along with the original speech signal. The four formant frequencies obtained are 994.36 Hz, 1991.22 Hz, 3000.38 Hz and 4286.48 Hz respectively for a speech signal having sampling frequency 44100 Hz.

Cepstrum Based Pitch Extraction Technique

Basically pitch extraction allows ordering of sounds on frequency related scale. It determines state of speech of signal. It has done on (.wav) format sound file that are recorded in our database of different speaker.

Algorithm for Pitch Extraction

We firstly take input voice signal of (.wav) format. Now this analog signal is sampled with a suitable rate and then it is quantized to convert it into a digital signal. The digital signal is then hamming windowed to convert it into a suitable frame size. The signal is converted into frequency domain by using Fast Fourier Transform. Signal's absolute values are then considered to obtain logarithm of signal. The signal is then transformed into Cepstral domain by taking its IFFT. The very first signal peak represents the pitch frequency.

In speech processing, pitch extraction using Cepstral method is used to determine who is talking, for speaker separation and for phase based speech reconstruction. Once in the cepstral domain, the pitch can be estimated by picking the peak of the resulting signal within a certain range. The cepstrum is given in terms of "quefrency" which, besides being a terrible name, represents pitch lag. Therefore the lag at which there is the most energy represents the dominant frequency in the log magnitude spectrum thereby giving you the pitch.

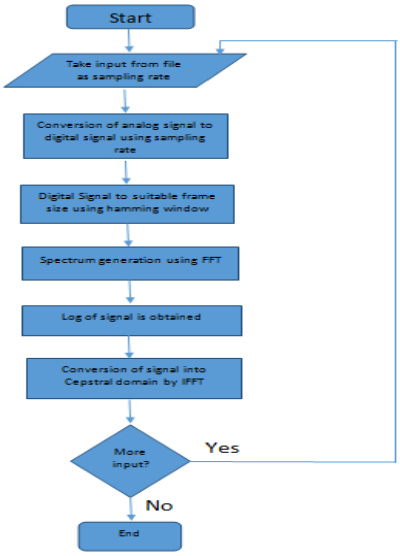

Figure 4 Flowchart for Pitch Extraction

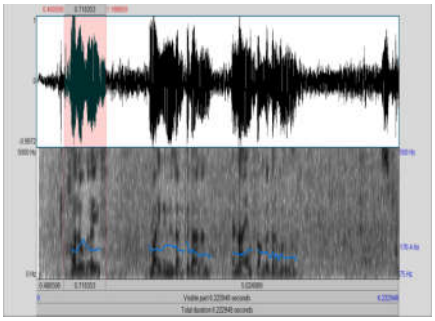

Figure 5 Speech Signal and its Pitch Frequency

Different States of Emotions

There are different states of emotions like anger, sadness, frustration, happiness, surprise and many more. Depending upon diferent age groups, the value of pitch frequency and formant frequency differ for diferent people depending upon their state of emotion.

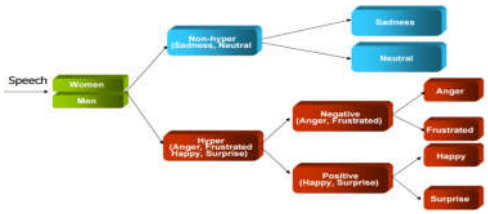

Figure 6 States of Emotions

RESULT

Analysis of speech signal to identify the 3 emotional state Neutral, Angry and Happy from the speech signal has been carried out using Pitch and Formant Frequencies as the basic features. From the results obtained it can be clearly seen that Pitch is the best feature to identify the two emotions Neutral state and Angry. This is shown in table 2 below. The pitch frequency is much lesser for Neutral speech than angry speech. The mean of Pitch frequency for Neutral emotion is 144.2 Hz where as it is 292.25 Hz for Angry emotion. Thus pitch can be considered as a clear indicator to identify the emotions Neutral and Angry from the speech of a person.

The emotion Happy is recognized satisfactorily using the Formant Frequency estimation method. Table 1. This shows the range of formant frequencies f1, f2 and f3 for the emotion Happy for five different speakers Also the mean formant frequency for different speakers have been shown in table. Thus the emotion Happy can be easily identified from the formant frequency analysis. Such a trend is not indicated by the emotions Neutral and Angry with respect to the formant frequencies.

Table 1 Formant Frequencies and its Mean for Five Speakers in Happy State.

Emotion Happiness	Formant Frequency F1 in Hz	Formant Frequency F2 in Hz	Formant Frequency F3 in Hz	Mean Frequency F in Hz
Ankit Anand	256.4	771.5	1734.2	920.7
Vinay	544.3	880.6	1795.7	1073.5
Rahul	265.3	741.1	1616.1	874.2
Rajeev	319.4	728.4	1502.9	848.7
Ankush	290.2	891.4	1491.8	891.1

Table 2 Pitch Frequency for 5 persons in Neutral and Anger State

Speaker	Pitch Frequency for Neutral Emotion in Hz	Pitch Frequency for Anger Emotion in Hz
Ankit	144.86	367.14
Vinay	199.54	250.19
Rahul	103.58	278.63
Rajeev	166.81	345.58
Ankush	105.23	219.73
Mean	144.12	292.25

CONCLUSION

It is noticed that different emotions spreads over different frequency bands. Pitch and Formant Frequencies have been used to recognize three states of emotions. Pitch estimation using Cepstral analysis is carried out for Neutral and Anger emotions for five speakers. It is seen that pitch has higher frequency range for Anger emotion than speech produced under Neutral condition. Formant frequency estimation is done using LPC analysis which extracts the first three formant frequencies. Thus it is seen that Pitch calculation for speech signal gives better results for Neutral and Angry emotions than the formant frequency estimation method. Happy emotion is recognized satisfactorily using LPC based formant frequency estimation method.

References

B. Atal “Automatic Recognition of Speakers from their Voices”, Proceedings of the IEEE, vol. 64, April 1976, pp. 460-475.

Vinay K. Ingle, John G. Proakis, “Digital Signal Processing Using Matlab V4”, PWS Publishing Company, 1997.

Todor Dim itrov Ganchev, “Speaker Recognition”, PhD Thesis, Wire Communication Laboratory, Dept. of Computer Science and Engineering, University of Patras, Greece, November 2005.

S. S. Nidhyananthan and R. S. Kumari, “Language and Text-Independent Speaker Identification Sys- tem using GMM”, WSEAS Transactions on Signal Processing, Vol.9, pp. 185-194, 2013.

“A Comparative Study of Formant Frequencies Estimation Techniques” Dorra Gargouri, Med Ali Kammoun and Ahmed Ben Hamida, ENIS, University of Sfax

“A Comparative Performance Study of Several Pitch Detection Algorithms”, IEEE Transactions on acoustics, speech, and signal processing, VOL. ASSP-24, NO. 5, October 1976.

“Recognition of emotions in speech by a hierarchical approach”, Affective Computing and Intelligent Interaction and Workshops, ACII 2009. 3rd International Conference on 10-12 Sept. 2009 page(s): 1 - 8 Amsterdam 2009.

A Framework for Automatic Human Emotion Classification using Emotion Profiles” Emily Mower, IEEE Transactions on Audio, Speech and Languare Processing, Vol 19, July 2011.

T.L. Nwe, Analysis and Detection of Human Stress and Emotion from Speech Signals, Ph. D. Thesis, National University of Singapore, 2003.

Matlab V4”, PWS Publishing Company Vinay K. Ingle, John G. Proakis, “Digital Signal Processing Using, 1997.

H. Poor, “An Introduction to Signal Detection and Estimation”, New York: Springer-Verlag, 1985. [7]

Royce Chan and Michael Ko, Speaker Identification by MATLAB, June 14, 2000.

Adaptation of pitch and spectrum for HMM-based speech synthesis using MLLR,” by M. Tamura, T. Masuko, K. Tokuda, and T. Kobayashi, Proc of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 805–808, May 2010.

D. A. Reynolds and R. C. Rose, “Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models”, IEEE Transactions on Speech and Audio Processing, Vol. 3, pp. 72-83,1995.