International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

# Speech Signal Processing for Speaker Recognition

**Yudhvir Singh Sidhu**[1], **Rupinder Kaur**[2]

[1, 2] Doaba Institute of Engineering and Technology

**Abstract:** ***In this paper we provide a brief overview of the area of speaker recognition, describing applications, underlying techniques and some indications of performance. Following this overview we will discuss some of the strengths and weaknesses of current speaker recognition technologies and outline some potential future trends in research, development and applications***

**Keywords:** speaker, speech, sound, transducer, microphone

## 1. Introduction

The voice signal conveys many levels of information to the users. At the primary level, voice conveys a message via words, but at other levels voice conveys information about the language being spoken and the emotions, gender and the identity of the speaker. The voice signal contains rich messages and there are three main recognition fields of speech, which are of most interest and have been studied for several decades. These three fields are speech recognition, language recognition and speaker recognition.

**Speech Recognition System ($S_p$RS)** refers to the ability of a machine or program to recognize or identify spoken words and carry out voice. The spoken words are digitized into sequence of numbers, and matched against coded dictionaries so as to identify the words. The area of the speech recognition is one which a large no. of options must be specified before the problem.

**Speaker Recognition System (SRS)** is dynamic biometric task, which stems from the more general speech processing area. Similarly to most of the other speech-related recognition activities (speech recognition, language recognition, etc.), speaker recognition is a multidisciplinary problem.

**Speaker Verification System (SVS)** is the process of determining the speaker identity that who is the person is speaking. Different terms which have the same definition as speaker verification could be found in literature, such as voice verification, voice authentication, speaker/talker authentication, talker verification.

## 2. Literature Review

The first attempts for speaker recognition system was done in the 1960s. Pruzansky at Bell Labs was among the first to initiate research by using filter banks and correlating two digital spectrograms for a similarity measure. Filter-bank is a classical spectral analysis technique which consists in representing the signal spectrum by the log-energies at the output of a filter bank, where the filters are overlapping band-pass filters spread along the frequency axis .This representation gives a rough approximation of the signal spectral shape while smoothing out the harmonic structure if any. Pruzansky and Mathews improved upon this technique and, Li et al. further developed it by using linear discriminators.

Doddington at Texas Instruments (TI) replaced filter banks by formant analysis. Formant frequency is determined from the frequency representation of a speech signal. When a speaker pronounces a vowel, the vocal tract forms a certain shape. The formant frequencies are the resonant frequencies that correspond to that particular shape. There is a set of frequencies that associate with one shape. Formant Analysis involves directly solving for the parameters of a vocal tract function, which can be modeled by a all-pole transfer function. The coefficients of the transfer function are solved using the Auto-Regressive algorithm. Intra-speaker variability of features, one of the most serious problems in speaker recognition, was intensively investigated by Endres et al and Furui. Inter-speaker variability, the variation of speech between different speakers is the effect that makes speaker verification possible. The greater the inter-speaker variability between true speaker and impostor, the more accurate a system is likely to be. Another variation known as intra speaker variability refers to the variation of speech from a single speaker.

It has also been found that adding long-range features can provide a larger relative gain in performance when larger amounts of training data are available. Second, unlike frame-based features, longer-range features reflect voluntary behaviour, and as such could potentially be usefull not only for recognizing speakers, but also for recognizing characteristics of the speech, such as the speaking style (e.g., casual chit-chat versus argumentation versus event planning). Finally, regardless of the applied task, research on long-range features should be of fundamental scientific interest to researchers interested in understanding speaking behaviour.

## 3. Problem Formulation

The focus in the proposed work is given on the speaker verification task of the voice recognition. The work includes verifying a person by extracting the features of his/her voice. In this work, the concentration is on the text- dependent speaker verification. The commands that are to be used for the completion of will be based on some predefined words such as Open, Close, Stop Bye, Go, Come etc.

## 4. Methodology

As the proposed work is based on the speaker identification, it is required to record the voice sample of different persons and then test it for the display identity of the person. Digital speech acquisition is used for acquiring the voice signal which converts the analog speech signal of different pressure waves into equivalent digital signal. The signal will be captured using microphone will be used as transducer. The signal is then fed to a processing unit which will filter the signal and process for feature extraction. The features that has been reported by different researcher in the field of speaker identifications, will be extracted and fed through the pattern matching technique, identity of the person will be identified.

## 5. Conclusion

It is clear that speaker verification technology is indeed ready for use. But, as stated before, it is not the universal solution. The main strength of speaker verification technology is that it relieson a signal that is natural and unobtrusive to produce and can be obtained easily from almost anywhere using the familiar telephone network (or internet) with no special user equipment or training. This technology has prime utility for applications with remote users and applications already employing a speech interface. Additionally, speaker verification is easy to use, has low computation requirements (can be ported to cards and handhelds) and, given appropriate constraints, has high accuracy. Some of the flexibility of speech actually lends to its weaknesses. First, speech is a behavioral signal that may not be consistently reproduced by a speaker and can be affected by a speaker's health (cold or laryngitis). Second, the varied microphones and channels that people use can cause difficulties since most speaker verification systems rely on low-level spectrum features susceptible to transducer/channel effects. Also, the mobility of telephones means that people are using verification systems from more uncontrolled and harsh acoustic environments (cars, crowded airports), which can stress accuracy. Robustness to channel variability is the biggest challenge to current systems. Spoofing of systems is often cited as a weakness, but there have been may approaches developed to thwart such attempts (prompted phrases, knowledge verification). There is current effort underway to address these known weaknesses. Some of these weaknesses may be overcome by combination with a complementary biometric, like face recognition.

## References

[1] J. Kua, J. Epps, E. Ambikairajah, and E. Choi, 2009 "LS regularization of group delay features for speaker recognition," in Proc. Inter speech, pp. 2887–2890.

[2] S. Nakagawa, W. Zhang, and M. Takahashi, 2006 "Text-independent/text prompted speaker recognition by combining speaker-specific GMM with speaker adapted syllable-based HMM," IEICE Trans., vol. E89-D, no. 3, pp. 1058–1064.

[3] S. Nakagawa, K. Asakawa, and L.Wang, 2007 "Speaker recognition by combining MFCC and phase information," in Proc. Inter speech, pp. 2005–2008.

[4] L.Wang, N. Kitaoka, and S. Nakagawa, 2007 "Robust distant speaker recognition based on position-dependent CMN by combining speaker-specific GMM with speaker-adapted HMM," Speech Communication., vol. 49, no. 6, pp. 501–513.

[5] L. Zhao, 2003 "speech signal processing", Machine Press.

[6] H. HU, 2000 "speech signal processing", Harbin University of Industry Press.

[7] J.H Xie, 1995 "HMM and its application to speech processing ",Huazhong University of Science and Technology Press.

[8] S. Furui, 1997 "Recent advances in speaker recognition". AVBPA97, pp 237—251.

[9] J. P. Campbell, 1997 ``Speaker recognition: A tutorial,'' Proceedings of the IEEE, vol. 85, pp. 1437--1462.

[10] Special Issue on Speaker Recognition, 2000 Digital Signal Processing, vol. 10.