Review On Speech Signal Processing & Its Techniques

Kiran[1], Assistant Professor, Department of Forensic Science, School of Bio Engineering & Biosciences, Lovely Professional University, Phagwara, Punjab, India
Aastha Pangotra, M.Sc. Forensic Science, Department of Forensic Science, School of Bio Engineering & Biosciences, Lovely Professional University, Phagwara, Punjab, India
Corresponding author: Kiran, Assistant Professor, Department of Forensic Science, School of Bio Engineering & Biosciences, Lovely Professional University, Phagwara, Punjab, India

ABSTRACT:
Immense advancement in technology has enabled an easy speech & speaker recognition and signal processing. This review discusses the formation of speech, various mechanism and techniques used in identifying, extracting and deciphering the speech signals and its processing. The rapid pace of the field demands active efforts to ensure that this breakthrough technology is used responsibly for effective signal recognition and processing.
This technology has a major importance in Forensics as it assists in investigation process, by this they can repair, enhance, and analyse audio recordings using a variety of scientific tools and techniques.

INTRODUCTION:
Speech is defined as the means of communication used by people through articulating vocal sounds which does not refers only to the noises but to the pieces of the linguistic codes.
Text is a written form of communication. In speech the fundamental analog form of message is an acoustic waveform which is known as the Speech signal. The process of speech processing involves the use of various branches of science like physics, computer science, pattern recognition, and linguistics [1]. Speech processing simply involves the examination of digital speech signals so as to manipulate, store, transfer and provide an output of signals. The input includes the speech recognition and the output is speech synthesis. Speech recognition is simply defined as the process of converting it into text form. Some aspects of speech processing are perceptual coding of speech and audio, speech recognition, enhancement, modification, speech-to-text synthesis etc. Speech processing has various applications in various fields [2].
In forensic science it helps in investigation as it can serve for Evidence enhancement as it can clarify the sound in a audio & video recordings. It is used for analysis, interpretation and identification. The audio could tell about the environment where the recording took place. The volume and the tone of the audio voice can tell about the spatial relationship within scene. Technical details like presence of an unnatural waveform may indicate that an edit has been made [3].

FORMATION OF SPEECH:
The formation of speech involves conceptualization links to intention to create speech, formulation; links to the creation of linguistic form of the message , and articulation which results in production of soundwaves. For the production of speech sounds air stream is used. For the articulation of speech sounds lung-air is used. When we breath in & out the vocal are drawn apart and glottis is open [4]. When some speech sound is produced when the glottis is open such sounds are known as Voiceless sounds. During production of certain speech

sounds the vocal cords are loosely held and the pressure of the air from the lungs make them open & close rapidly, this is called the Vibration of vocal cords and the sound produced so is known as Voiced sounds.
The rate of vibration of vocal codes is called as Frequency and this determines the pitch of the voice [5].

PREPROCESSING OF SPEECH SIGNAL:
This is the first phase and involves the segregation of voiced and unvoiced regions of the captured signal. It helps in adjusting and modifying the speech signal for further computer processing & feature extraction analysis [6]. It involves;

- BACKGROUND NOISE REMOVAL, in this the ambient noise is removed by using Signal-to-noise ratio.
- SPEECH WORD DETECTION, in this voice activity detectors (VAD) are used to separate speech & non-speech segments.
- ZERO CROSSING RATE(ZCR) is the rate of signal changes of signal during frame
- ENERGY NORMALIZATION
- WINDOWING, in this the segmented waveform is multiplied by a time window to reduce discontinuity of speech.

FEATURE EXTRACTION OF SPEECH SIGNAL:
It is the process of converting the soundwave it a digital signal and then obtaining the various features like pitch, energy, power and vocal tract configuration from the speech signal [7].
Following techniques are commonly used for feature extraction from speech signals:

- LINEAR PREDICTIVE CODING: in this the values of signals is expressed as a linear combination of preceding values which describes the time varying linear system which represents the vocal tract. It provides auto-regression based speech features. It is a formant estimation technique. It is used for encoding speech at low bit rate although it cannot differentiate between words with similar vowel sound [8].
- MEL-FREQUENCY CEPSTRUM: used for speech processing tasks & mimics the human auditory system. The Mel frequency scale is equal to the linear frequency spacing below 1000Hz & a log spacing above 1000Hz [9].
- RELATIVE SPECTRAL FILTERING: it is a band pass filtering technique, designed to lessen the effect of noises and enhance the speech.
- PROBABILISTIC LINEAR DISCRIMINATE ANALYSIS: this is based on the i-vector having full information, it is a low dimensional vector having fixed length. It is flexible acoustic model which uses variable number of interrelated input frames [10].

SPEECH RECOGNITION TECHNIQUES:
It is the process of synthesis of text form from speech. It helps in recognising isolated words, connected words, continuous words and spontaneous words.
There are mainly 3 techniques employed for this purpose:

1. HIDDEN MARKOV MODEL(HMM):
   This is used for the hidden or unobserved states. It provides a simple and effective framework for modelling time-varying spectral vector sequences.
   It uses forward algorithm for isolated word recognition, viterbri algorithm for continuous speech recognition and forward-backward algorithm for training an HMM. In this the states are represented as acoustic models, each discrete time step,

a transition is taken into new state, and then the output symbol is generated.
This model uses algorithm for estimation of hidden variables out of given list of observations. The transitions and output symbols chosen are governed by probability distribution [11].

2. DYNAMIC TIME WARPING (DTW):
   It is used for comparison between two temporal sequences based on their difference between speed. It measures the similarity between the sequences which vary in time and speed, it also finds the optimal nonlinear alignment between them [12].

3. ARTIFICIAL NEURAL NETWORKS(ANN):
   This model is based upon the reasoning of neurons of human brain. ANN is made up of number of processing units also known as neurons. The neurons are connected by passing signals and output is transmitted through outgoing neuron which have branches transmitting the same signal. The single unit of this system is known as perceptron. This system helps in character recognition [13].
   Various other studies have been done in this regard with successful findings [14-23].

CONCLUSION:

Speech signal processing can be used in field of forensic science for identification of a criminal. In order to achieve effective feature extraction, we can use techniques such as linear predictive coding, Mel-frequency cestrum, Relative spectral filtering and probabilistic linear discriminate analysis.

REFERENCES:

[1] Kamble, B. C. (2016). Speech Recognition Using Artificial Neural Network–A Review. Int. J. Comput. Commun. Instrum. Eng, 3(1), 61-64.
[2] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "Neural Network used for Speech Recognition", Journal of Automatic Control, University of Belgrade, Vol. 20, pp. 1-7, 2010. http://dx.doi.org/10.2298/JAC1001001G
[3] Ganesh Tiwari,"Text Prompted Remote Speaker Authentication: Joint Speech & Speaker Recognition/Verification System.
[4] http://www.guidogybels.eu/asrp4.html
[5] Yashwanth H, Harish Mahendrakar and Suman Davia, " Automatic Speech recognition Using Audio Visual Cues", IEEE India Annual Conferencec pp. 166-169, 2004.
[6] G. Saha, Sandipan Chakroborty, Suman Senapati, "A New Silence Removal and Endpoint Deletion Algorithm for Speech and Speaker Recognition Applications.
[7] Urmila Shrawankar, Dr. Vilas Thakare, "Techniques for Feature Extraction in Speech Recognition System: A Comparative Study.
[8] Lei Xie, Zhi-Qiang Liu, "A Comparative Study of Audio Feature for Audio Visual Conversion in MPEG-4 Compliant Facial Animation", Proc. of ICMLC Dalian, 13-16, August 2006. http://dx.doi.org/10.1109/icmlc.2006.259085
[9] Honig, Florian Stemmer, George Hacker, Christian Brugnara, Fabio, "Revising Perceptual Linear Prediction", In interspeech – 2005, pp. 2997-3000.
[10] Vimal Krishnan VR, Athulya Jayakumar, Babu Anto P, "Speech Recognition of Isolated Malyalam Words Using Wavelet Feature and Artificial Neural Networks", 4th IEEE International Symposium on Electronic Design, Test and Application, 2008.

[11] Santosh K. Gaikwad, Bharti W. Gawali, Pravin Yennawar, "A Review on Speech Recognition Techniques", IJCA Vol. 10, No. 3, pp. 16-24, November 2010 http://dx.doi.org/10.5120/1462-1976.
[12] Lindasalva Muda, "Voice Recognition Algorithm Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Vol. 2, Issue 3, March 2010.
[13] Singh Satyanand, Dr. E. G. Rajan, "Vector Quantization Using MFCC and Inverted MFCC", International Journal of Computer Applications, Vol. 17, No. 1, pp. 1-7, March 2011.
[14] Sonali B. Maind, Priyanka Wankar, "Research Paper on Basic of Artificial Neural Network ", International Journal on Recent & Innovation Trends in Computing & Communication, Vol. 1, Issue 1, pp. 96-100.
[15] Robison, A. J. Cook, G. D. Ellis, D. P. W. Fosteruissier, E., Renals, S. J., Williams, D. A. G., "Connectionist Speech Recognition of Broadcast News", Speech Commnication 37: 27-45, 2000.
[16] James Martens, Ilya Sutskever, "Learning Recurrent Neural Network with Hessian-Free Optimization", University of Toronto, Canada.
[17] Gasser Auda, Mohamed Kamel, "Modular Neural Network: A Survey", International Journal of Neural System, Vol. 9, No. 2, pp. 129-151, April 1999. http://dx.doi.org/10.1142/S0129065799000125
[18] Shyam M. Guthikonda, "Kohonen Self-Optimizing Maps", Wittensberg University, December 2005. Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE)
[19] Gulati S. Comprehensive review of various speech enhancement techniques. Advances in Intelligent Systems and Computing,1108,2020.
[20] Bansalwal K., Sharma K., Jain A. Speech recognition implementation. International Journal of Innovative Technology and Exploring Engineering,8(9),2019.
[21] Bachate R.P., Sharma A. Automatic speech recognition systems for regional languages in India. International Journal of Recent Technology and Engineering,8(2),2019.
[22] Priya, Gahier A.KText to speech conversion in Punjabi-a review. International Journal of Control Theory and Applications,9(41),2016.
[23] Singh K. Part-of-speech tagging using genetic algorithms. International Journal of Simulation: Systems, Science and Technology,16(6),2015.