International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 07 | July-2016 www.irjet.net p-ISSN: 2395-0072

Different Methods Used In Voice Recognition Techniques

Vipul C. Rajyaguru[1]

[1]Assistant Professor, Department of Instrumentation & Control Engineering, Government Engineering College, Rajkot, Gujarat, India.

Abstract - *In this paper voice and speech detection and recognition techniques are introduced. Also different types of voice detection and recognition techniques have been discussed. According to raw data and need of conversion of data, the data processing is examined. As per the application which technique is to be used whether to develop document/record or to deal with hardware interface with outside environment is also discussed.*

Key Words: *VTTF, METUbet, taxonomy, AVSD, PMVDR, ASV & ASI, VoiceDirect 364, DIVE/TCL program, VR Stamp.*

1. INTRODUCTION

Speech recognition is an alternative to traditional methods of interacting with a computer, such as textual input through a keyboard. An effective system can replace, or reduce the reliability on, standard keyboard and mouse input. A speech recognition system consists of:

(a) A microphone, for the person to speak into.
(b) Speech recognition software.
(c) A computer to take and interpret the speech.
(d) A good quality soundcard for input and/or output. [7]

Capturing the vocal tract transfer function (VTTF) from the speech signal while eliminating other extraneous speaker dependent information such as pitch harmonics is a key requirement for accurate speech recognition. It is well known that the vocal tract transfer function is mainly encoded in the short-term spectral envelope. Therefore, extracting the short term spectral envelope accurately and robustly (especially in additive noise) is crucial for robust speech recognition. It is also widely accepted within the speech recognition community that incorporating perceptual considerations, such as the Mel and Bark scales, into the feature extraction process leads to improved accuracy. [3]

2. UNDERSTANDING WORD RECOGNITION OF THE LANGUAGE

Some vowels and consonants have variants depending on the place they are produced in the vocal tract. For example in Turkish, the letter a in the word laf [laf] is pre dorsal, while in almak [α mαk], a's are post dorsal. Therefore, 29 letters in the Turkish alphabet are represented by 45 phonetic symbols in. It may also be true to say that there is nearly one-to-one mapping between written text and its pronunciation. [6]

A mapping of the SAMPA characters to the METUbet characters is shown in Table. METUbet has 39 phonetic representations compared to the 45 phonetic SAMPA representations. The reason is that the open-short and closed-long forms of the letters u, ü, o, ö, and i are represented by the same phonetic symbol in METUbet. The closed-long forms of those letters appear when they are preceding soft g, which causes only the lengthening of those letters. This does not need to be considered for the phonetic alignment and phoneme recognition using Hidden Markov Models. [6]

3. DIFFERENT TYPES OF TECHNIQUES

Following are the Different Techniques used in the Detection of Voice.

3.1 Call – Type Classification

The CU Call Center Corpus consists of conversations collected from the University of Colorado Information Technology Services (ITS) telephone Help Desk, also known as the IT Service Center. The IT Service Center is a small contact center consisting of approximately 20 agents who support the trouble-shooting of a wide range of computer and telecommunications issues. The call centers are staffed daily by three to five agents who typically field 200 calls per day with an average call duration of 5 minutes. [5]

They have examined the transcribed conversations between callers and agents and have developed a hierarchical taxonomy of call-types for this task domain. Their taxonomy is based on analysis of 743 dual-channel transcribed conversations (1486 call-sides). They began their analysis by providing a summary of each call and later grouped similar calls into levels representing broad classes of call types. A total of 98 detailed call types have been determined through inspection of the data. A subset of the taxonomy of calls is shown in Figure 1. [5]

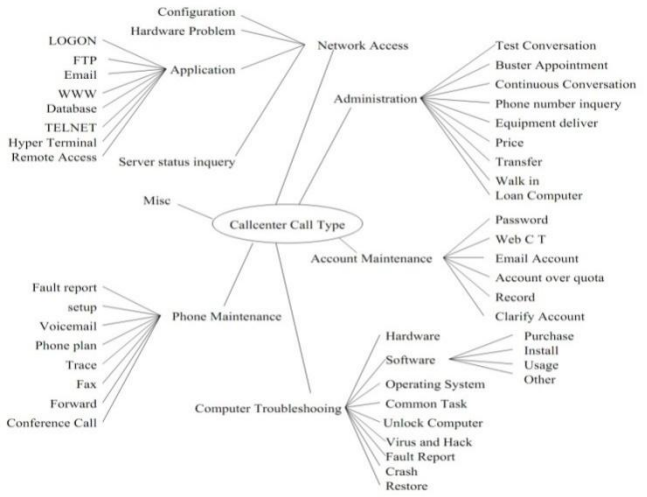

Fig -1: The Hierarchical Taxonomy of Call Types.

3.2 A Distribute Architecture

The University of Colorado has previously participated in both SPINE-I and SPINE-II evaluations. Our efforts towards the evaluation systems have focused on:

(a) The development of new features for robust speech recognition,
(b) Improved model adaptation methods and
(c) An efficient, integrated approach to joint speech detection and recognition for noisy environments. [2]

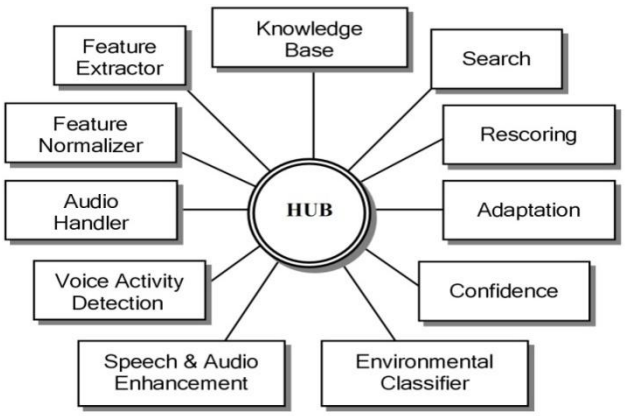

Fig -2: A Distributed Architecture ForA Speech Recognition System That Incorporates A Hub and Eleven Servers.

3.3 Automatic Voice Signal Detection (AVSD)

In the well-known Fourier analysis, signal is broken down into constituent sinusoids of different frequencies. Another way to think of Fourier analysis is as a mathematical technique for transforming our view of the signal from time-based to frequency-based. A simple Fourier transform is illustrated in Fig. 3(a).Taking the Fourier transform of a signal can be viewed as a rotation in the function space of the signal from the time domain to the frequency domain. [4]

Fig -3(a): Fourier Transform from Time Domain to Frequency

Similarly, the wavelet transforms can be viewed as transforming signal from the time domain to wavelet domain. This new domain contains more complicated basis functions called wavelets, mother wavelets or analyzing wavelets. A simple Fourier transform is illustrated in Fig. 3(b). [4]

Fig -3(b): Wavelet Transform from Time Domain to Wavelet Domain.

3.4 PMVDR (Perceptual Minimum Variance Distortion less Response) Algorithm

Utilizing direct warping on the FFT power spectrum by removing the filter bank processing step leads to the preservation of almost all the information in the short-term speech spectrum. We can now summarize the remainder of the proposed PMVDR algorithm as follows:

(a) Obtain the perceptually warped FFT power spectrum,
(b) Compute the "perceptual autocorrelations" by utilizing
(c) The IFFT on the warped power spectrum,
(d) Perform a Q^{th} order LP analysis via Levinson-Durbin recursion using perceptual autocorrelation lags,
(e) Calculate the Q^{th} order MVDR spectrum using Equation from the LP coefficients,
(f) Obtain the final cepstrum coefficients using the straight-forward FFT-based approach.

Flow diagram for the PMVDR algorithm is given in Fig. 4. [3]

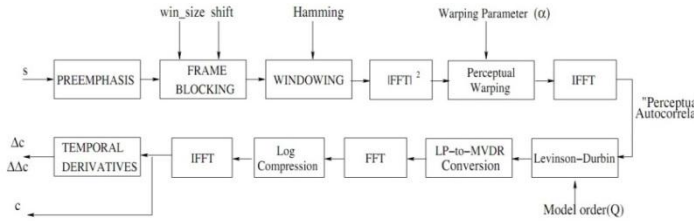

Fig -4: Schematic Diagram of PMVDR Front-End Computation

3.5 Speaker Recognition

Speaker recognition encompasses verification and identification. Automatic speaker verification (ASV) is the use of a machine to verify a person's claimed identity from his voice. The literature abounds with different terms for speaker verification, including voice verification, speaker authentication, voice authentication, talker authentication, and talker verification. In automatic speaker identification (ASI), there is no a priori identity claim, and the system decides who the person is, what group the person is a member of, or (in the open set case) that the person is unknown. [1]

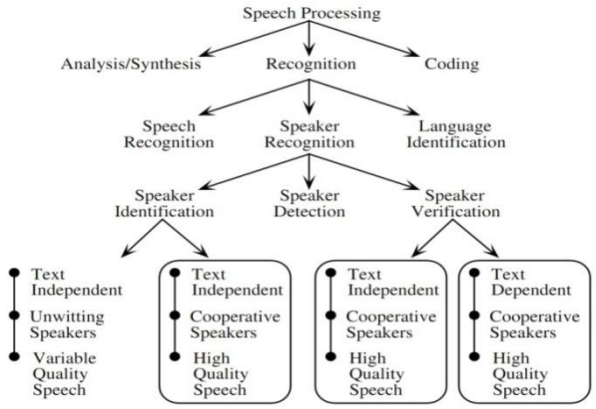

Fig -5: Speech Processing

The general approach to ASV consists of five steps: digital speech data acquisition, feature extraction, pattern matching, making an accept/reject decision, and enrollment to generate speaker reference models. A block diagram of this procedure is shown in Figure 6. Feature extraction maps each interval of speech to a multidimensional feature space. (A speech interval typically spans 10 to 30 ms of the speech waveform and is referred to as a frame of speech.) This sequence of feature vectors Xi is then compared to speaker models by pattern matching. This results in a match score Zifor each vector or sequence of vectors. The match score measures the similarity of the computed input feature vectors to models of the claimed speaker or feature vector patterns for the claimed speaker. Last, a decision is made to either accept or reject the claimant according to the match score or sequence of match scores, which is a Hypothesis-testing problem. [1]

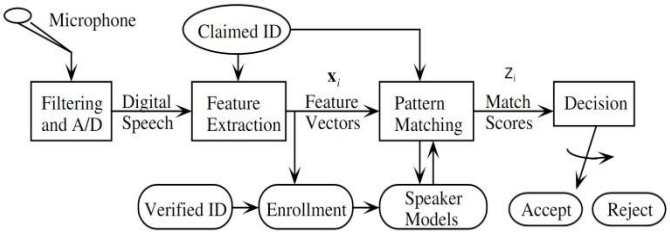

Fig -6: Generic Speaker Verification System

For speaker recognition features that exhibit high speaker discrimination power, high inter speaker variability, and low intra speaker variability is desired. Many forms of pattern matching and corresponding models are possible. Pattern matching methods include dynamic time warping (DTW), hidden Markov modeling (HMM), artificial neural networks, and vector quantization (VQ). Template models are used in DTW, statistical models are used in HMM, and codebook models are used in VQ. [1]

3.6 Voice Recognition to Interact With A Virtual Environment

The VoiceDirect 364 speech recognition kit For the purpose of integrating voice recognition in a virtual environment, we used a virtual environments browser called DIVE (Distributed Interactive Virtual Environments), developed by the Swedish Institute of Computer Science because this software is free for academic use, it can be used in various computer platforms, and it is easy to program, among other features. DIVE can run TCL programs and also has its own graphics language. A way to obtain the voice commands from the Voicedirect speech recognition kit is to connect it to a computer parallel port and read its output values (the n^{th} recognized word from the previously trained words) from the parallel port via a program. However, DIVE and TCL languages do not control parallel port directly. Thus, a program made in Visual Basic was developed to read the parallel port and saved the data in a temporary text file. A DIVE/TCL program easily read this file, and the data from the text file could serve to manipulate virtual objects in DIVE. Figure 1depicts a schematic view of this process. [9]

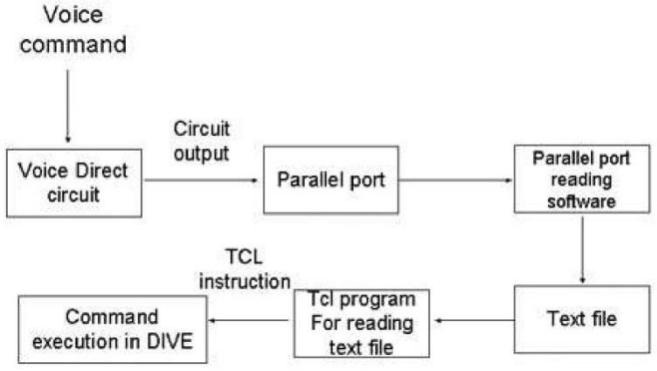

Fig -7: Schematic Diagram of The Voice Direct 364 Circuit Connections To The PC.

The VR Stamp evaluation board has been used to develop customized firmware in the C++ language. In addition to the VR Stamp, the circuitry also includes a speaker amplifier, relays for switch outputs, and status LED indicators. The system is powered by a 9V battery. A 3.3V voltage regulator is used to supply power to the VR Stamp unit. [10]

The VR Stamp can be operated in either the speaker independent mode or the speaker dependent mode. For the speaker independent mode the voice commands to be recognized are from a standard library and preprogrammed into the processor. For the speaker dependent model the voice commands are programmed for the specific speaker and can be re-recorded any time.[10]

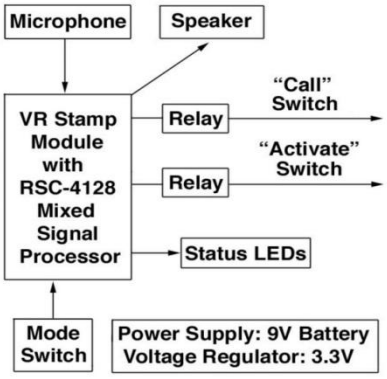

Fig -8: Block Diagram of the Voice Recognition System

4. SELECTION OF RECOGNITION TECHNIQUE

Here we have discussed many types of voice or speech detection and recognition techniques. Among all the technique best suited technique is selected as per application we have. If multi users are their then call – type recognition is best suited. If numbers of serves are connected in voice recognition system at that time distributed architecture is used. In Automatic Voice Signal Detection (AVSD) technique Wavelet or Fourier Transformation is necessary for processing the data. Perceptual Minimum Variance Distortion less Response (PMVDR) technique straight-forward FFT and Front – End Computation is carried out. Automatic speaker verification (ASV) is the use to verify a person's identity from his voice. Also it can carry out work like speaker verification, including voice verification, speaker authentication, voice authentication, talker authentication, and talker verification. In automatic speaker identification (ASI) is used to identity person, and the system decides who the person is, what group the person is a member of, or that the person is unknown. Many forms of pattern matching and corresponding models are possible. Pattern matching methods include dynamic time warping (DTW), hidden Markov modeling (HMM), artificial neural networks, and vector quantization (VQ). Template models are used in DTW, statistical models are used in HMM, and codebook models are used in VQ. VoiceDirect 364 speech recognition kit is used for conversion of voice command of speech in to the text file with the help of Dragon's Naturally Speaking and IBM's Via Voice or Microsoft Speech Recognizer or any Other Computer Based Voice Detecting software. VR Stamp Module with RSC – 4128 Mixed Signal Processor is used to interface with hardware like fan, lamp, heater, blower, etc. with help of relay as intermediate device for isolation as well as amplification purpose.

5. CONCLUSIONS

There are different types of techniques used for the conversion of voice data for the storage purpose as well as for generation of documents and to develop records. Also some kits are developed to deal with hardware to interface with outside environment. All these techniques are very help full while using computer or other equipment.

REFERENCES

[1] Joseph P. Campbell, Jr., "SPEAKER RECOGNITION",Department of Defense, Fort Meade, MD.

[2] KadriHacioglu& Bryan Pellom, "A DISTRIBUTED ARCHITECTURE FOR ROBUST AUTOMATIC SPEECH RECOGNITION", Center for Spoken Language Research, University of Colorado at Boulder.

[3] Umit H. Yapanel& John H. L. Hansen, "A New Perspective on Feature Extraction for Robust In – Vehicle Speech Recognition", Robust Speech Processing Group, Center for Spoken Language Research Univ. of Colorado at Boulder, CO, 80309, USA, EUROSPEECH 2003 – GENEVA.

[4] Shiv Kumar, Member, IACSIT, IAENG AdityShastri and R. K. Singh"An Approach for Automatic Voice Signal Detection (AVSD) Using Matlab", International Journal of Computer Theory and Engineering, Vol. 3, No. 2, April 2011, ISSN: 1793 – 8201.

[5] Min Tang, Bryan Pellom, KadriHacioglu, "CALL – TYPE CLASSIFICATION AND UNSUPERVISED TRAINING FOR THE CALL CENTER DOMAIN",Center for Spoken Language Research, University of Colorado at Boulder, Boulder, Colorado 80309 – 0594, USA.

[6] Ö. Salör, B. Pellom, T. Ciloglu, K. Hacioglu, M. Demirekler, "On Developing New Text and Audio Corpora and Speech Recognition Tools for the Turkish Language" ICSLP-2002:Inter. Conf. on Spoken

Language Processing, vol. 1, pp. 349-352, Denver, CO USA, Sept. 2002.

[7] John Kirriemuir,"Speech Recognition Technologies", March 30th 2003, TSW 03 – 03, March © JISC 2003.

[8] HsinEu and Alan Hedge," Survey of Continuous Speech Recognition Software Usability", Cornell University, Department of Design & Environmental Analysis, MVR Hall, Ithaca, NY 14853, © HsinEu and Alan Hedge, Cornell University, September 1999.

[9] Miguel A. Garcia-Ruiz and Cesar R. Bustos-Mendoza," Using Hardware-based Voice Recognition toInteract with a Virtual Environment", Virtual Reality Laboratory, University of Colima, CEUPROMED, Colima, 28040,Mexico.

[10] AfeezOlalekan, Alex Page, Ying Sun, PhD, "Optimizing the Functionality of a Voice Recognition System for Assistive Technology", Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881-0805 USA.