Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.275 – 284

RESEARCH ARTICLE

Security System in Speech Recognition

[1]Sunita Dixit, [2]Dr. MD Yusuf Mulge
[1]Research Scholar, Pacific University, Udaipur
[2]Principal, PDM College of Engineering for Women, Bahadurgarh
bhardwajsunita23@gmail.com

Abstract: Speaker recognition is one of the effectively used biometric authentication system that actually identify the speaker on the basis of vocal characteristics. The speaker identification depends on different voice features such as the intensity analysis, voice pitch analysis, voice feature extraction etc. This recognition process is also affected from different factors such as the background noise, instrumentation noise etc. In this paper, noise effective approach is suggested to define an effective speaker recognition process. The robustness of the recognition system is improved with the definition of an integrated layered model.

Keywords: Speech Enhancement, Spectral Subtraction, LPC, HMM, ANN

I. INTRODUCTION

Speech is the most sophisticated signal naturally produced by humans. The speech signal carries linguistic information for sharing of information and ideas[1]. It allows people to express emotions and verbally share feelings. It is the most fundamental form of communication among humans. The aim of digital speech processing is to take advantage of digital computing techniques to process the speech signal for increased understanding, improved communication, and increased efficiency and productivity associated with speech activities. The field of speech processing includes speech analysis and representation, speech coding, speech synthesis, speech recognition and understanding, speaker verification, and speech enhancement. Speech is a complex signal that is characterized by varying distributions of energy in time as well as in frequency, depending on the specific sound that is being produced. The speech signal also possesses other characteristics that make it a very efficient means for carrying semantic (meaning) as well as pragmatic (task-dependent) information[2]. Speaker recognition is concerned with machine verification or identification of individual talkers, based on their speech, for authorization of access to information, networks, computing systems, services, or physical premises. At times, as an important biometric feature, a talker's speech may also be used for forensic or criminal investigations. Speaker Identification is becoming a high-relevant task in many fields, especially in the framework of security remote applications. These systems, usually developed under laboratory conditions, severely degrade their performance level when an acoustical mismatch appears among training and testing phases. The traditional framework for analyzing speech is the source-tract model first proposed by Homer Dudley at Bell Laboratories in the 1930s. In this model[3], as depicted in Figure 1, a speech excitation signal is produced by an excitation source and processed by a filter system that "modulates" the spectral characteristics of the excitation signal based on the shape of the vocal tract for the specific sound being generated.

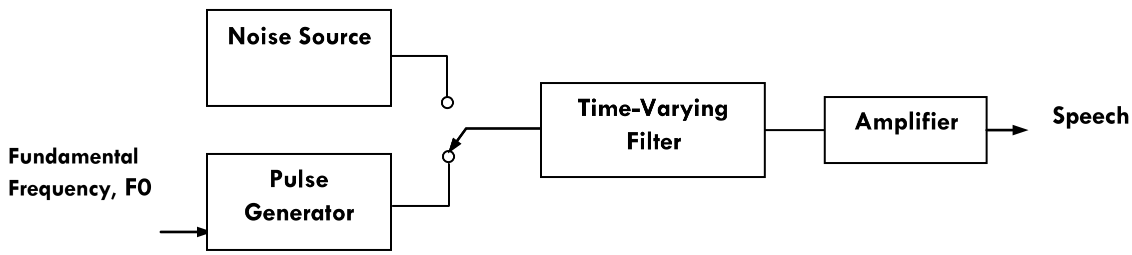

Figure 1 speech production model -- the basis for speech analysis

Rest of the paper is organized as follows. Section II defines related work done in this field. Section III explains proposed approach in detail. Section IV presents results obtained using proposed approach followed by section V which concludes the findings based on the obtained results.

II. RELATED WORK

In Year 2003, Chin-Teng Lin performed a work, "Single-Channel Speech Enhancement in Variable Noise-Level Environment". This paper discusses the problem of single-channel speech enhancement in variable noise-level environment. Commonly used, single channel subtractive-type speech enhancement algorithms always assume that the background noise level is fixed or slowly varying. In year 2006, Amarnag Subramanya performed a work," Speech Modeling with Magnitude-Normalized Complex Spectra and Its Application to Multisensory Speech Enhancement". A good speech model is essential for speech enhancement, but it is very difficult to build because of huge intra- and extra-speaker variation. In Year 2007, Esfandiar Zavarehei performed a work," Noisy Speech Enhancement Using Harmonic-Noise Model and Codebook-Based Post-Processing". This paper presents a post-processing speech restoration module for enhancing the performance of conventional speech enhancement methods. The restoration module aims to retrieve parts of speech spectrum that may be lost to noise or suppressed when using conventional speech enhancement methods. In Year 2008, Tim Fingscheidt performed a work," Environment-Optimized Speech Enhancement".

In Year 2009, Hairong Jia performed a work,"A Modified Speech Enhancement Algorithm based on the Subspace". A modified speech enhancement Algorithm based on the subspace is advanced. It reduces residual noise caused by wrongly estimating noise eigen value matrix and speech eigen value matrix, because in the traditional speech enhancement algorithm based on the subspace, the eigen value matrix of noise are attained by eigen decomposing to covariance matrix of noise, but covariance matrix of noise is estimated by using variance in the silence sequent, it cannot instead the whole noise, and lead to residual noise.

III. PROPOSED APPROACH

Speaker recognition is one of the effectively used biometric authentication system that actually identify the speaker on the basis of vocal characteristics. The speaker identification depends on different voice features such as the intensity analysis, voice pitch analysis, voice feature extraction etc. This recognition process is also affected from different factors such as the background noise, instrumentation noise etc. In this section, the noise effective approach is suggested to define an effective speaker recognition process. In this approach, the robustness of the recognition system will be improved with the definition of an integrated layered model. The robustness will be achieved for background noise and the instrumentation noise.

The approach is divided in two main stages. At the initial stage, the high level filtration over the noise is performed to remove the background noise. To perform this high level segmentation Spectral-subtraction method is used. In the later stage, to remove the instrumentation noise, linear probabilistic coding approach is used. This is the analytical approach that will perform the effective reduction of noise over the signal. At the final stage, the recognition process is performed using HMM improved neural network approach. The HMM[11] actually identifies the speech features and finally neural network performs the identification process. Later on these features are trained on neural to perform the effective recognition. The approach is robust again the noisy speech. It is a three stage model. The basic architecture of the proposed approach is given as under

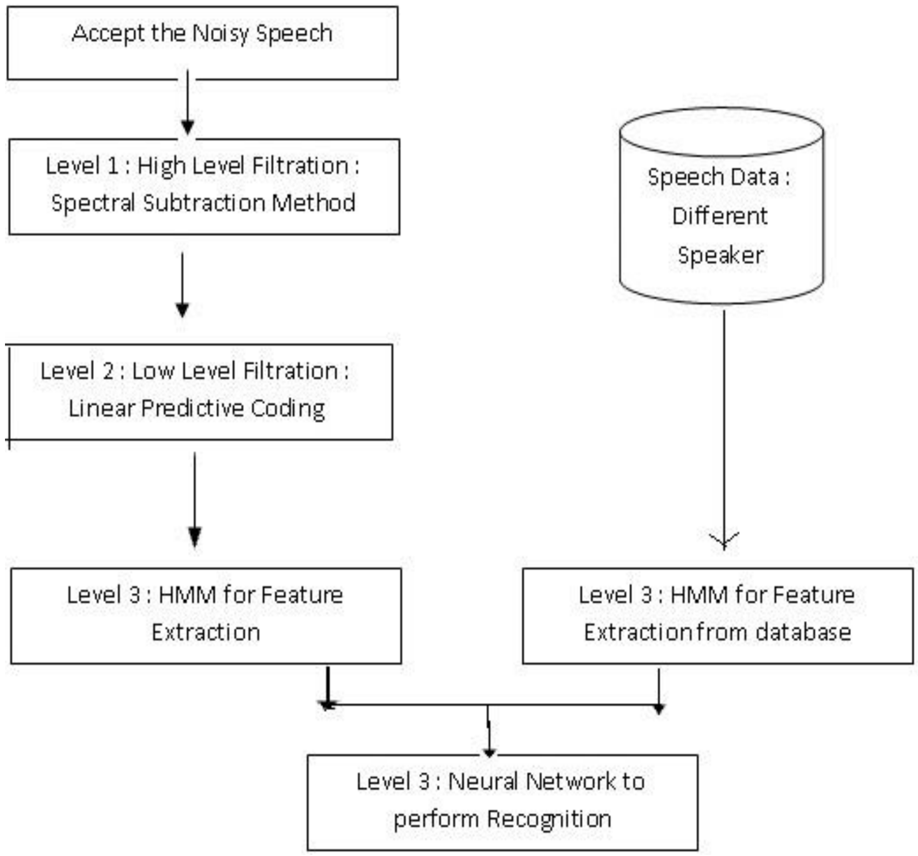

Fig. 2 Proposed model

As shown in the figure 2, the proposed model is having three main stages. First two stages are defined to perform de-noising over the signal to improve the robustness of recognition process. In the third stage, the HMM improved neural approach is defined to perform the recognition.

The de-nosing process will be effective against the background noise as well as the instrumentation noise. At the earlier stage, the spectral subtraction method is used to reduce the background noise over the speech signal and later on Linear predictive coding is used for low level filtration. This filtration process will reduce the instrumentation noise.

The recognition process defined in this model is the hybridization of HMM and the neural network. The HMM will actually used for feature extraction and the neural will use the classification approach to perform the recognition. The recognition will be performed on the featured speech dataset. The aim of the work is to improve the recognition ratio.

Filtration techniques: As shown above, first two stages are defined to perform de-noising over the signal to improve the robustness of recognition process. Following two techniques are used for high level and low level filtration:

- Spectral subtraction method
- Linear predictive coding

A. Spectral subtraction method:

The goal of spectral subtraction is to suppress the noise from the degraded signal. It is based on the principle that one can estimate and update the noise spectrum when speech signal is not present and subtract it from the noisy speech signal to obtain clean speech signal spectrum. It is assumed that the noise is additive and its spectrum doesn't change with time. It means noise is stationary or slowly varying with time. The Block Diagram of Spectral Subtraction Method is shown in figure below:

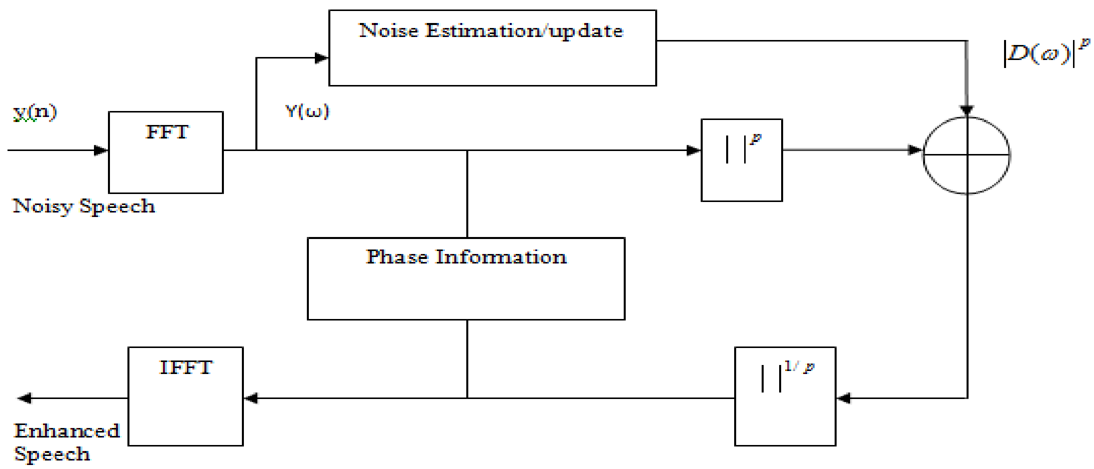

Figure 3: Block Diagram of Spectral Subtraction

The enhanced speech is obtained by subtracting the estimated spectral components of the noise from the spectrum of the input noisy signal. The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present.

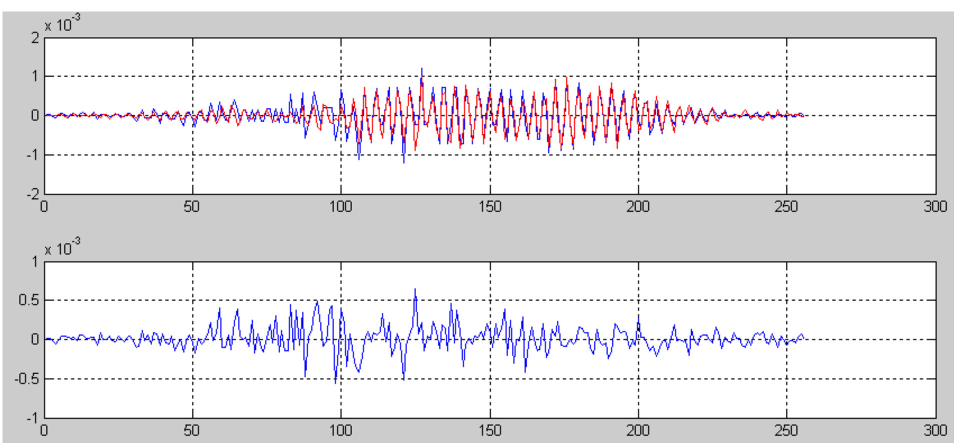

Figure 4: Input Speech Processing

As the input signal is taken, the spectral subtraction method is applied for high level filtration. Here figure 4 is showing the high level filtration process applied over the speech signal. The signal is here filtered to remove the instrumentation noise over the speech. The figure is also showing the improved speech signal.

B. Linear predictive coding

The LPC is one of the strongest tools in speech signal processing. The idea of this analysis is that each sample of the speech sign can be expressed as a linear equation of previous inputs and outputs. The transform function of the system can be achieved by applying the Z transform. An all pole model is very good estimation for the transform function. The important point in computing the *LPC* is that these coefficients can be directly driven from the speech signal for this reason and because of the dependence of the speech signal on times first, windowing is done the signal then the *LPC* coefficients are calculated in short frames.

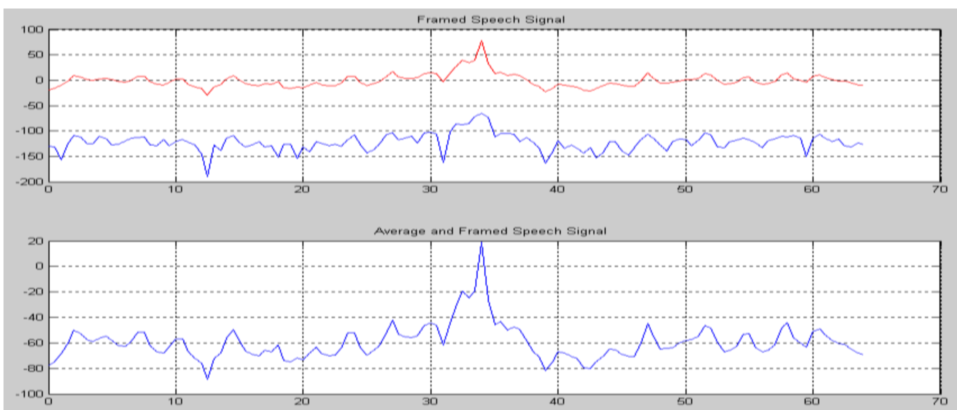

Figure 5: Frame Averaging Analysis

Here figure 5 is showing the frame averaging analysis applied over the speech signal. The window processing is used to perform the block by block analysis over the speech so that the partial analysis will be performed under vector quantization. Here the LPC method is been implemented to analyze and filter the speech signal.

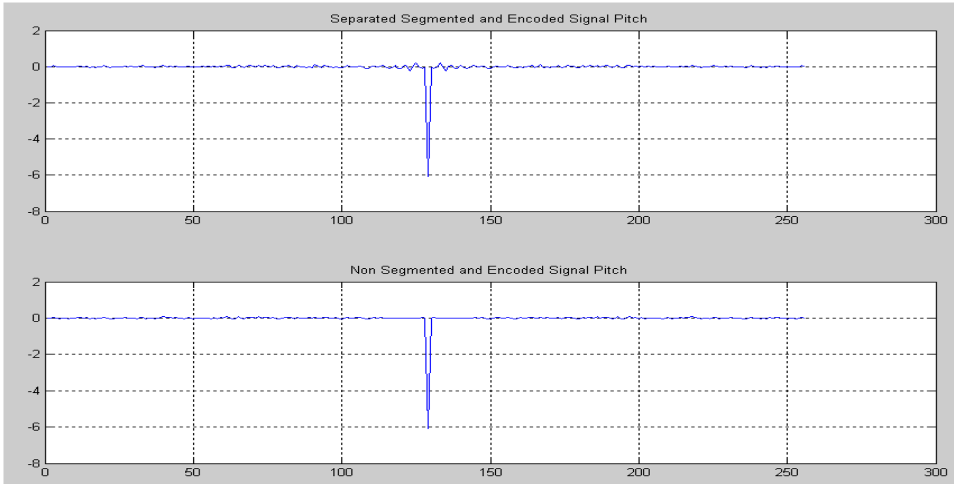

Figure 6: Segmented Speech Signal

Here figure 6 is showing the segmented speech signal obtained from initial LPC processing. The segmentation process is applied at reduce the signal noise under magnitude level analysis. The figure is showing the segmentation stage applied over the speech.

Recognition technique: The recognition process defined in our proposed model is the hybridization of HMM and the neural network.

A. HMM model:

HMMs are most simple networks that can produce speech by using a number of states for each model and modeling the short-term spectra associated with each state with, usually, mixtures of multivariate Gaussian distributions (the state output distributions). The parameters of the model are the state transition probabilities and the means, variances and mixture weights that characterize the state output distributions. Each word, or each phoneme, will have a different output distribution; a HMM for a sequence of words or phonemes is made by concatenating the individual trained HMM for the separate words and phonemes.

B. Neural Network

Neural networks have been used in many aspects of speech recognition such as phoneme classification, isolated word recognition, and speaker adaptation. Neural networks make no assumptions about feature statistical properties and have several qualities making them attractive recognition models for speech recognition. When used to estimate the probabilities of a speech feature segment, neural networks allow discriminative training in a natural and efficient manner. Few assumptions on the statistics of input features are made with neural networks. However, in spite of their effectiveness in classifying short-time units such as individual phones and isolated words, neural networks are rarely successful for continuous recognition tasks, largely because of their lack of ability to model temporal dependencies. Thus, one alternative approach is to use neural networks as a pre-processing e.g. features transformation, dimensionality reduction, etc.

IV. RESULTS

To recognize the speaker under noise vector a two stage model is defined in previous section. The initial stage is for the speech signal filtration and second is to perform the speaker recognition. The signal improvement or enhancement is again defined under two main approaches called spectral signal analysis and LPC approach. The recognition is performed under hybrid model using HMM and neural network. The implementation of work is done on real time voices for different users using MATLAB.

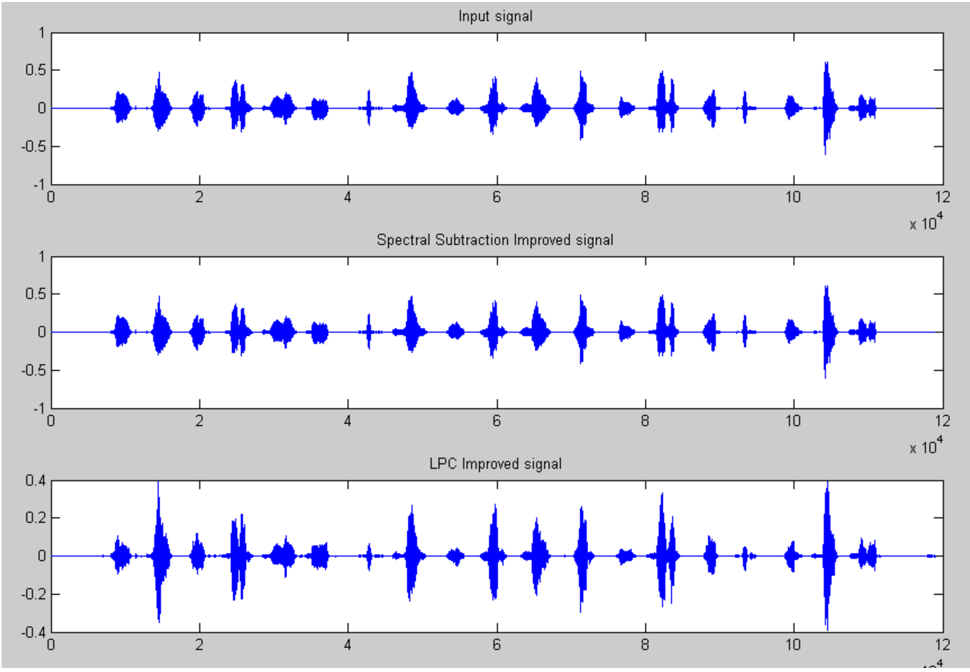

Figure 7: Improved Speech Signal

Here figure 7 is showing the improved speech signal after the spectral subtraction and LPC approaches. The figure is showing the input signal is enhanced so that effective speaker recognition can be performed over it. Therefore, now speaker recognition can be obtained using HMM and neural network based recognition model over noise filtered database.

V. CONCLUSION

In this paper, we have defined an approach for noisy speech signal. The noise can be some instrumentation noise or the background noise. The approach is divided in two main stages. In first stage, the signal filtration is performed using two layer model. In first layer, the spectral subtraction approach is defined to perform high level filteration and later on linear predictive model is applied for low level filteration. After this filteration stage, the recognition is applied using HMM improved neural network approach. The approach is tested on real time dataset. The results show the effective recognition of speaker over the database.

REFERENCES

[1] Sharon Gannot,” Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms”, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING 1063–6676/98@1998 IEEE

[2] Huaiyu Dai,” A Joint Speech Coding-Enhancement Algorithm for MBE Vocoder”, Intemational Conference on Communication Technology ICCT’98

[3] Nathalie Virag,” Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System”, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING 1063–6676/99@1999 IEEE

[4] Benito Carnero,” Perceptual Speech Coding and Enhancement Using Frame-Synchronized Fast Wavelet Packet Transform Algorithms”, IEEE TRANSACTIONS ON SIGNAL PROCESSING 1053–587X/99@1999 IEEE

[5] Chin-Teng Lin,” Single-Channel Speech Enhancement in Variable Noise-Level Environment”, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS 1083-4427/03© 2003 IEEE

[6] Te-Won Lee, ”Speech Enhancement By Perceptual Filter With Sequential Noise Parameter Estimation”.

[7] Amarnag Subramanya, ”Speech Modeling with Magnitude-Normalized Complex Spectra And Its Application To Multisensory Speech Enhancement”, ICME 2006 1424403677/06©2006 IEEE

[8] Esfandiar Zavarehei, ”Noisy Speech Enhancement Using Harmonic-Noise Model and Codebook-Based Post-Processing”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING 1558-7916© 2007 IEEE

[9] Anuradha R. Fukane, ”Enhancement of Noisy Speech Signals for Hearing Aids”, 2011 International Conference on Communication Systems and Network Technologies 978-0-7695-4437-3/11© 2011 IEEE

[10] Biing-Hwang Juang, “Minimum classification error rate methods for speech recognition”, 1997, IEEE Transactions on Speech and Audio Processing (Volume: 5, Issue: 3)

[11] Rabiner, L.,”A tutorial on hidden Markov models and selected applications in speech recognition” 1989 Proceedings of the IEEE (Volume: 77, Issue: 2)

[12] Gillick, L, "Some statistical issues in the comparison of speech recognition algorithms" International Conference on Acoustics, Speech, and Signal Processing, 1989

[13] Chibelushi, C.C. "A review of speech-based bimodal recognition", 2002 IEEE Transactions on Multimedia, (Volume: 4, Issue: 1).

[14] Reddy, D.R, "Speech recognition by machine: A review", Proceedings of IEEE (Volume: 64, Issue: 4) ISSN: 0018-9219.

[15] Potamianos, G., "Recent advances in the automatic recognition of audiovisual speech", Proceedings of the IEEE (Volume: 91, Issue: 9) ISSN: 0018-9219

[16] Bahl, Lalit R, "A Maximum Likelihood Approach to Continuous Speech Recognition" IEEE Transactions on Pattern Analysis and Machine Intelligence, (Volume: PAMI-5, Issue: 2).

[17] Sakoe, H, "Dynamic programming algorithm optimization for spoken word recognition", IEEE Transactions on Acoustics, Speech and Signal Processing (Volume: 26, Issue: 1)

[18] Bahl, Lalit R., "A tree-based statistical language model for natural language speech recognition", IEEE Transactions on Acoustics, Speech and Signal Processing (Volume: 37, Issue: 7).

