ISSN: 2456 - 3935

International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 01 – 07

Modified Hidden Markov Model for Speaker Identification System

Dr. S. Kalaivani
Associate Professor, School of Electronics Engineering, VIT University, Vellore, India
Email: kalaivani.s@vit.ac.in

Rahul Singh Thakur
UG Scholar, School of Electronics Engineering, VIT University, Vellore, India
Email: rahulsinghthakur1992@gmail.com

Dharmendra
UG Scholar, School of Electronics Engineering, VIT University, Vellore, India
Email: dharmendra2011@vit.ac.in

Abstract: *Speaker recognition is an automated process of knowing who is speaking, on the basis of distinctive information included in speech signals. Every speaker recognition process involves enrollment of speakers during the training session and verification or identification during the testing session. Feature extraction and speaker modelling are the two very essential steps in any speaker verification or identification process. This paper discusses the development of speaker identification system using perceptual linear prediction (PLP) for feature extraction and hidden Markov model (HMM) for speaker modelling. The former technique uses concepts from the psychophysics of hearing to derive an estimate of the auditory spectrum: (1) the critical-band spectral analysis, (2) the equal loudness curve, and (3) the intensity-loudness power law, whereas the latter one is a doubly stochastic process especially known for its application in temporal pattern recognition such as speech, handwriting etc. Pitch is used as an extra feature to work on the limitation of HMM in order to enhance the speaker identification Rate (SIR) of the system. The Results of the system developed using only PLP are compared with the system developed using both PLP and Pitch.*

Keyword: *Hidden Markov model; perceptual linear prediction; spectrum; speaker identification rate; speaker recognition;*

1. INTRODUCTION

Speaker recognition is a biometric technique that uses an individual's voice for identification or verification purpose. There are many people who often confuse themselves thinking that voice (speaker) recognition and speech recognition are one and the same but there is a big difference between the two[1-4]. Speech recognition is recognizing what is being spoken whereas speaker recognition is identifying or verifying the speaker who has made an utterance.

There are two major applications of speaker recognition which are speaker verification and speaker identification. If the speaker claims to be of a certain identity and the voice is used to verify his/her claim, this is called verification. On the other hand, determining an unknown speaker's identity is called identification. Moreover if the uttered text is the same for enrolment and verification then it is called text-dependent recognition. If not, then it is called text-independent recognition [5]. In a text-dependent system, utterances can either be common for all speakers or unique for each whereas this in not the case in text independent system.

Speaker recognition is also categorized into closed-set recognition and open-set recognition [6-7]. The closed-set refers to the cases in which the unknown voice must come from a set of known speakers and the Open-set means unknown voice may come from unregistered speakers as well.

Speech is always regarded as the most powerful form of communication because of its rich dimensional character. Besides the speech text, the rich dimensions also refer to as the gender, emotion, attitude, health situation and identity of a speaker. This information is very important for establishing an effective communication.

The speaker recognition systems are developed in two phases [1, 8] training phase and recognition phase. In the training phase, each registered speaker provides samples of his/her speech so that the system can create a reference model corresponding to that speaker. In the testing phase, the input speech is compared with stored model(s) and a recognition decision is made.

One of the important decisions in any pattern recognition system is to choose how exactly to represent the signal that is to be classified. Through more than many years of research, many different feature extraction techniques of the speech signal have been suggested and tried like Mel Frequency Cepstrum Coefficient (MFCC) [9], LPC [10] etc. PLP [11] is one such technique. Moreover for speaker modelling, different modelling techniques for speaker recognition system have been identified such as Gaussian Mixture Model (GMM) [12] and Hidden Markov Model (HMM) [13-14], which are prevalent techniques in this field.

Speaker recognition is a difficult task. The fundamental source of variation is the speaker himself/herself. Speech signals in training and testing sessions can be largely different due to many factors such as age (people voice change with time), speaking rates, and health conditions and so on. There are also other factors that present a challenge to speaker recognition technology. Acoustical noise and variations in recording environments are two of them.

Since speaker recognition technique enables the system to use the speaker's voice for identity verification and controlling access to services such as voice dialling, banking & shopping by telephone, remote access to computers etc., therefore, a highly efficient and accurate speaker recognition system is a must. The speaker identification system proposed in this paper is a text dependent system which uses techniques like PLP for feature extraction and HMM for speaker modelling and pattern classification. HMM has a limitation. The higher the number of speaker models to compare the test voice signal with, lesser is the identification accuracy. Say for example if there are ‘N’ number of speaker models, the identification accuracy of the system will be more if the test voice sample is matched/compared with a number of speaker models, less than are ‘N’. Because of this issue, the efficiency of the system using HMM reduces on increasing the number of speaker models.

To overcome this drawback of HMM, pitch is used as an additional feature to categorize the speaker models into sets in order to reduce the number of comparisons for improved results. Moreover, the results of the system developed using PLP only are compared with that of the system developed using pitch as an additional feature along with PLP.

2. DESIGN APPROACH

The proposed system has been implemented using the sampling rate of 8000Hz with reference to human speech analysis standard. Voice samples are recorded in audio format and then converted into .wav format for further processing. Filtering of the Recorded voice samples is done using a Band-Pass Filter with the passband frequency range same as the audible frequency range of humans.

of the manuscript shall contain the sections such as introduction, literature review/related works, methods and materials, proposed method/algorithm, experimental setup, results and discussion, conclusion, acknowledgement, references, biography with optional photograph.

2.1 Silence Removal and End Point Detection

This algorithm is divided into two parts. Silence Portions of voice signal have been removed by initially labelling samples as voiced/silence using statistical properties of background noise. The mean and standard deviation of the N samples of the given utterance are calculated as:

$$\mu = \left(\frac{1}{N}\right) \sum_{k=1}^{N} \mathrm{x(i)} \qquad (1)$$

$$\sigma = \sqrt{\left(\frac{1}{N}\right) \sum_{k=1}^{N} (x(i) - \mu)^2} \qquad (2)$$

If x (i) represents the recorded voice signals, then μ is the mean and σ is the standard deviation. Background noise is characterized by Equation (1) ,(2). For each sample, if one dimensional Mahalanobis distance function is $\left(\frac{|x-\mu|}{\sigma}\right) \geq 3$ then the sample is treated as voiced sample, otherwise silence/unvoiced sample.

2.2 Modified Autocorrelation Method for Pitch Detection

The modified autocorrelation pitch detector is based on the centre-clipping method [15-16]. Generally Clipping Threshold (Cl) is about 30% of the maximum magnitude of signal. To get highCl, we can take the peak value of the first 1/3 and the last 1/3 of the signal and use the lesser one to be the maximum magnitude. Then the 60-80% of this maximum magnitude is set Cl.

$$clp[x(n)] = \begin{cases} x(n) - Cl, & If\ x(n) \geq Cl\ ; \\ 0, & If\,|x(n)| < Cl\,; \\ x(n) + Cl, & if\ x(n) \leq -Cl\ ; \end{cases} \qquad (4)$$

Centre clipping of the signal using Equation (4) is followed by the autocorrelation of the obtained centre clipped signal. Thereafter, the autocorrelation function is searched for its maximum value in every window of samples. The maximum value of the autocorrelation function is then compared with 0.55 times energy of the signal. If less, pitch is considered zero otherwise location of that maximum value is the pitch corresponding to those samples. Average of those values gives the pitch of the voice signal [16].

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 01 – 07

are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

2.3 Perceptual Linear Prediction Model for Feature Extraction

In PLP, The speech spectrum obtained is modified by several transformations that are based on models of the human auditory system. The various steps involved are as follows:

Step 1: *Calculating the short term power spectrum of the voice signal*

Step 2: *Critical-Band Spectral Analysis*:

The spectrum$P(f)$ is warped along its frequency axis into the bark frequency (Ω) by the given formula:

$$\Omega = 6 * \log(\left(\frac{f}{600} + \sqrt{1 + \left(\frac{f}{600}\right)^2}\right) \quad (5)$$

The resulting warped power spectrum obtained using Equation (5) is then convolved with the power spectrum of the simulated critical-band masking curve i.e. ψ (Ω);

$$\psi(\Omega) = \begin{cases} 0\,; & \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)}\,; & -1.3 \le \Omega \le -0.5 \\ 1\,; & -0.5 \le \Omega \le 0.5 \\ 10^{-1(\Omega-0.5)}\,; & 0.5 \le \Omega \le 2.5 \\ 0\,; & \Omega > 2.5 \end{cases} \quad (6)$$

The discrete convolution of ψ (Ω) with power spectrum yields samples of the critical-band power spectrum.

Step 3: *Equal-Loudness Pre-emphasis:*

Hynek's magic equal-loudness-curve formula [17]is utilized, in which the function 'eql' is an approximation to the unequal sensitivity of human hearing at various frequencies and simulates the sensitivity of hearing at about the 40-dB level.

$$eql = \left(\frac{f_{sq}}{f_{temp}}\right)^2 * \left(\frac{f_{sq}+1.44e^6}{f_{sq}+9.61e^6}\right) \quad (7)$$

Where,

$f_{sq} = (f)^2$ and

$$f_{temp} = (f_{sq} + 1.6e^5) \quad (8)$$

Step 4: *Intensity-Loudness Power Law of Hearing:*

This operation simulates the nonlinear relation between the intensity of sound and its perceived loudness. It is similar to cube root amplitude compression.

Step 5:

Inverse Discrete Fourier Transform of the output obtained after cube root compression yields autocorrelation coefficients [18-19]. The first half of those values is used to determine the autoregressive coefficients which can be further transformed to cepstral coefficients for one's convenience.

The phases comprised in the perceptual linear prediction technique are given in Algorithm 1and the block diagram of modified autocorrelation pitch detection algorithm is depicted in Figure 1.

Algorithm.1 - perceptual linear prediction technique

Step 1	:	The short term power spectrum of the voice signal was calculated.
Step 2	:	The spectrum $P(f)$ was warped along its frequency axis into the bark frequency (Ω) using Eq. (5)
Step 3	:	Power spectrum of the critical-band masking curve $\psi(\Omega)$was determined using (6) and the bark frequencies (Ω) were obtained.
Step 4	:	The discrete convolution of $\psi(\Omega)$ with Power Spectrum yielded samples of the critical-band power spectrum.
Step 5	:	The critical bands were weighted with the Hynek's magic equal-loudness-curve values calculated using Eq. (7).
Step 6	:	Cube root amplitude compression was performed on the weighted critical bands.
Step 7	:	Inverse Discrete Fourier Transform was performed on the output obtained after cube root compression yielding autocorrelation coefficients
Step 8	:	The first half of those values was used to determine the autoregressive coefficients.
Step 9	:	The autoregressive coefficients were further transformed to cepstral coefficients for one's convenience.

2.4 Modified Hidden Markov Model (M-HMM) for speaker Modelling

Algorithm 2 Hidden Markov Model for speaker modeling

Step 1	**:**	Observation vector was computed by k means clustering algorithm using the obtained PLP features in both the training and identification phases
Step 2	**:**	The initial state probability distribution matrix was assumed as '[1 0 0 0 0]', as we considered 5 states
Step 3	**:**	Using the assumed initial state probability and the computed observation vector, HMM parameters like transition matrix (A) and emission matrix (B) were determined, in the training phase

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 01 – 07

Step 4 : The probabilities of match between test speaker and the speaker models were computed using the test voice signal's observation vector and the HMM parameters (A and B) of the registered speakers, in the identification phase.

HMM is a doubly stochastic process in which one of the processes producing the sequence of observation and the other one is describing the state evolution [20-21]. An HMM model is characterized by the number of hidden states in the model (N), the initial state distribution , State transition probability distribution (A) and observation symbol probability(B). In the proposed method, an estimation problem is solved using Forward-Backward algorithm to train the model parameters A and B to get maximum probability for the given observation sequence. The steps followed for HMM technique is given in Algorithm 2.

3. EXPERIMENTAL RESULTS AND DISCUSSION

In the proposed method various Realistic Design Constraints have been considered and analyzed.

3.1 Voice Modulation of a speaker

Natural voice changes may affect speaker recognition accuracy like a temporarily hoarse voice caused by a cold or other sickness, different emotional states that affect voice (i.e. a cheerful voice versus a tired voice) and ddifferent pronunciation speeds during enrolment and identification. The aforementioned voice and user behavior changes can be managed in two ways: Separate enrolments for the altered voice, storing the records in the same person's template and a controlled, neutral voice during enrolment and identification.

3.2 Sensitivity of the Recording Device

The speaker recognition accuracy depends on the audio quality during enrolment and identification. There are no particular constraints on models or manufacturers when using regular PC microphones, headsets or the built-in microphones in laptops, smartphones and tablets. However the same microphone model is recommended (if possible) for use during both enrolment and recognition, as different models may produce different sound quality.
Some models may also introduce specific noise or distortion into the audio, or may include certain hardware sound processing, which will not be present when using a different model. And also the same microphone position and distance is recommended during enrolment and recognition.

Headsets provide optimal distance between user and microphone; this distance is recommended when non-headset microphones are used. Web cam built-in microphones should be used with care, as they are usually positioned at a rather long distance from the user and may provide lower sound quality. The sound quality may be affected if users subsequently change their position relative to the web cam.

3.3 Noise in the Environment

Speaker recognition algorithm is sensitive to noise in the background; they may interfere with the user's voice and affect the recognition results. These solutions may be considered to reduce or eliminate these problems: (i) A quiet environment for enrolment and recognition, (ii) Several samples of the same phrase recorded in different environments can be stored in a biometric template. Later the user will be matched against these samples with much higher recognition quality (iii) Close-range microphones which are not affected by distant sources of sound to be used and (iv) Third-party or custom solutions for background noise reduction, such as using two separate microphones for recording user voice and background sound, and later subtracting the background noise from the recording. In the proposed method, two identification systems were developed. One that uses PLP only, the other one uses PLP and Pitch both.

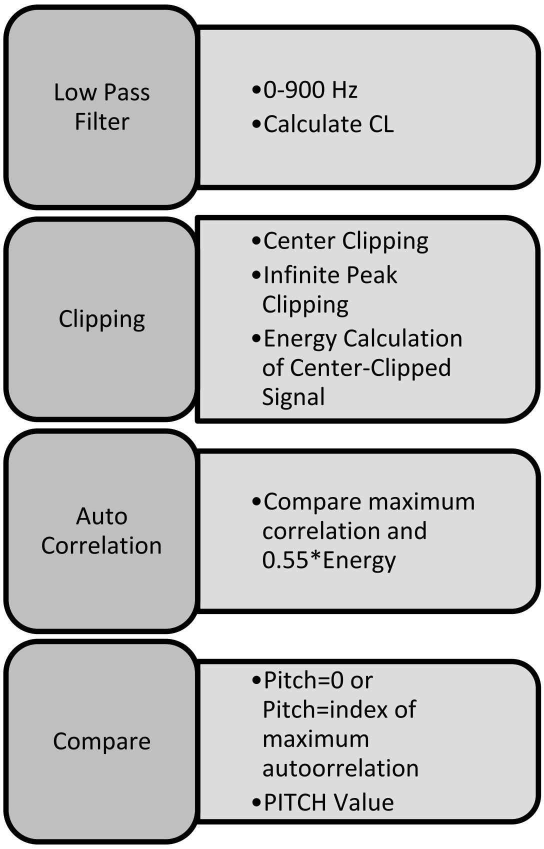

Figure.1 Block diagram of modified autocorrelation pitch detection algorithm

In both the systems, HMM is used for speaker modelling and comparison purpose. In the 'PLP Only System', the features of test voice signal are matched with all the 10 speaker models. This is not the case in the 'PLP-Pitch System' as in here since speaker models are divided into sets, the number of comparisons have reduced. This way higher identification accuracy of the system has been achieved. Since there are ‘10’ speaker models in total created in the training phase, it was observed that in PLP Only System, the no of comparisons taking place was more than that in PLP-Pitch System. Moreover, pitch of all the 10 speakers were calculated and accordingly based on the threshold values, categorized into sets.

TABLE 1 PITCH OF SPEAKERS’ VOICES (ONE TIME RECORDING)

Speakers	Pitch (Hertz)
Speaker 1	138.34
Speaker 2	173.82
Speaker 3	118.46
Speaker 4	160.21
Speaker 5	105.24
Speaker 6	187.41
Speaker 7	213.09
Speaker 8	208.70
Speaker 9	198.46
Speaker 10	232.39

In this work, voice signals were recorded from 10 speakers using a recording device in a noise free environment. All the speakers were asked to utter a word ‘zero’. The time duration of every recording was 4 seconds. Threshold values for the case of One Time Recording are: Pitch1 = 120 Hz, Pitch2 = 180 Hz, Pitch3 = 200 Hz and Threshold values for the case of Two Times Recording are: Pitch1 = 150 Hz, Pitch2 = 170 Hz, Pitch3 = 200 Hz. Pitch of speakers’ voices for one time recording and two time recording were listed in Table 1 and Table 2 respectively.

TABLE 2 AVERAGE PITCH OF SPEAKERS’ VOICES (TWO TIMES RECORDING)

Speakers	Pitch (Hertz)
Speaker 1	150.14
Speaker 2	172.69
Speaker 3	146.35
Speaker 4	155.37
Speaker 5	137.75
Speaker 6	187.84
Speaker 7	213.09
Speaker 8	208.70
Speaker 9	198.46
Speaker 10	232.39

Both the systems were tested for a range of K values, where K is the order of the Autoregressive Coefficients computed in Perceptual Linear Prediction Method. For both the systems, we observed the number of speakers which were unidentifiable by the system as well as the number of speakers which were nearly identifiable.

TABLE 3 OBSERVATIONS FOR ONE TIME RECORDING

K Value	Method	No. of Speakers (Un-identifiable)	No. of Speakers (Nearly Identifiable)
4	PLP	9	1
	PLP-P	3	2
5	PLP	6	1
	PLP-P	2	1
6	PLP	2	1
	PLP-P	0	1
7	PLP	4	0
	PLP-P	2	0
8	PLP	7	0
	PLP-P	5	0

TABLE.5 OBSERVATIONS FOR TWO TIME RECORDING (SERIES 2)

K Value	Method	No. of Speakers (Un-identifiable)	No. of Speakers (Nearly Identifiable)
4	PLP	10	0
	PLP-P	4	0
5	PLP	5	0
	PLP-P	4	0
6	PLP	4	0
	PLP-P	1	0
7	PLP	6	0
	PLP-P	1	0
8	PLP	8	0
	PLP-P	7	0

The efficiency of the system is determined by the parameter called speaker identification rate which is the percentage of the number of speakers which got identified by the system out of the total no of speakers and were listed in Table 3, 4 and 5. Performances of both the systems are compared for different values of the autoregressive model order for both the cases of one time recording and two times recording and was given in Figure 2, 3 and 4.

For each speaker in the database, there is a corresponding Hidden Markov Model. The basic structure of speaker identification (identification phase), as-

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 01 – 07

suming there are 'M' speakers in the database is that speaker identification has to perform 'M' pattern matching between the unknown speaker and 'M' known speakers. With a large number of speakers in the database, the performance of speaker recognition will decrease. Moreover since HMM is a doubly stochastic process, this models are too flexible and hard to train. As a result the high recognition accuracy is hard to be achieved with a large number of speaker models to compare with. This issue has been avoided in this project by incorporating pitch as an additional factor for the categorization of speaker models and to achieve improvement in HMM algorithm.

Speaker's voice signal varies due to various factors and the fundamental source of variation is the speaker himself/herself. Moreover, speech signals in training and testing sessions can be largely different due to many factors. Even estimating the exact pitch value of the voice signal is not easy. We don't have highly accurate pitch detection methods. Therefore in order to get almost accurate pitch value, we went for multiple recording of the voice signal, calculating pitch of each of the recordings and finding the average of all the pitch values.

TABLE 4: OBSERVATIONS FOR TWO TIME RECORDING (SERIES 1)

K Value	Method	No. of Speakers (Un-identifiable)	No. of Speakers (Nearly Identifiable)
4	PLP	10	0
	PLP-P	4	0
5	PLP	6	0
	PLP-P	3	0
6	PLP	3	0
	PLP-P	2	0
7	PLP	6	0
	PLP-P	2	0
8	PLP	8	0
	PLP-P	7	0

Hence, with reference to the data in tables and figures, for every value of the auto-regressive model order it can be observed than the number of unidentified speakers in 'PLP Only System' is more than that in 'PLP-Pitch System'. Moreover, it can be seen that the Speaker Identification Rate (the number of speakers out of the total which are identified by the system) of the 'PLP-Pitch system' is more than that of the 'PLP Only System', for each and every value of K.

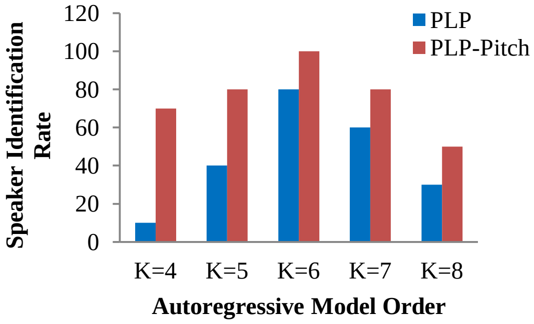

Figure 2 K-value versus SIR (for one time recording)

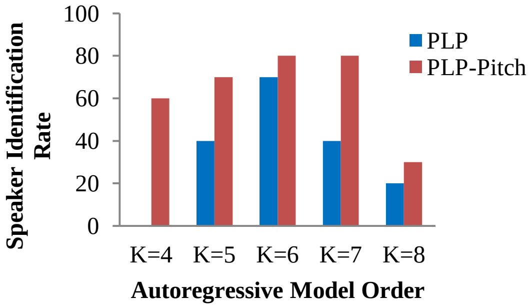

Figure 3 K-value versus SIR (for two times recording – Series 1)

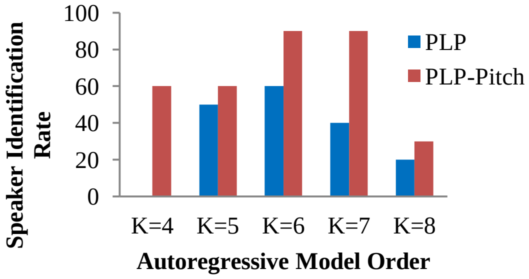

Figure 4 K-value versus SIR (for two times recording – Series 2)

4. CONCLUSION

In this paper, a speaker recognition systems was developed in two phases: training phase and recognition phase. In the training phase, each registered speaker provides samples of his/her speech so that the system can create a reference model corresponding to that speaker. In the testing phase, the input speech is compared with stored model(s) and a recognition decision is made. In HMM the efficiency of the system reduces on increasing the number of speaker models. Hence, to overcome this drawback of HMM, pitch was used as an additional feature to categorize the speaker models into sets in order to reduce the number of comparisons for improved results.

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 01 – 07

REFERENCES

[1] C. H. Lee, F. K. Soong and K. K. Paliwal, Automatic Speech and Speaker Recognition:Advanced Topics, Boston: Kluwer Academic Publishers , 1996.
[2] H. Beigi, Fundamentals of Speaker Recognition, New York: Springer Science and Business Media, Inc, 2011.
[3] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, New Jersey: Prentice - Hall, 1978.
[4] R. L. Klevans and R. D. Rodman, Voice Recognition, Boston: Artech House, 1997
[5] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, A study of inter-speaker variability in speaker verification, in IEEE Transactions on Audio, Speech and Language Processing , 2008
[6] Pillay, S., Ariyaeeinia, A., Sivakumaran, P., Pa wlewski, M.: Open-Set Speaker Identification under Mismatch Conditions. Proc. 10 the Annual Conference of the International Speech Communication Association (Inter speech'09), 2347-2350, 2009.
[7] Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin -Chagnolleau, I., Meignier S., Merlin, T., Ortega - Garcia, J., Petrovska, D., Reynolds, D. A. "A tutorial on text - independent speaker verification", EURASIP Journal on Applied Signal Processing, pp. 430 - 451, 2004.
[8] Sturim, D.E., Reynolds D.A., Dunn R.B., Quatieri T.F. "Speaker Verification using Text - Constrained Gaussian Mixture Models", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp. 677 - 680, May 2002.
[9] G. Fant, Acoustic Theory of Speech Production ,Mouton,The Hague, The Netherlands, 1970.
[10] L.R.Rabiner and B. Juang, Fundamentals of Speech Recognition, New Jersey: Prentice - Hall, Inc, 1993.
[11] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," Journal Acoust. Soc. Amer. , vol. 87, no. 4, pp. 1738–1752, 1990.
[12] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," Speech Communication , vol. 17, no. 1-2, pp. 91–108, 1995
[13] G. A. Fink, Markov Models for Pattern Recognition: From Theory to Applications. Berlin: Springer - Verlag, 2008.
[14] W. Ching and M. K. Ng, Markov Chains: Models, Algorithms and Applications, New York: Springer Science+ Business Media, Inc, 2006
[15] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal. "A comparative performance study of several pitch detection algorithms". IEEE Transactions on Audio, Signal, and Speech Processing 24 , 399-417 1976.
[16] M. M. Sondhi, "New methods of pitch extraction," IEEE Trans.Audio Electroacoust., vol. AU- 6, pp. 262-266, June,1968
[17] Hermansky, H., and Junqua, J. C. . "Optimization of perceptually based ASR front-end," in Proceedings of the IEEE International Conference on Acoustics; Speech, and Signal Processing Paper S5.10, pp. 219- 222., 1988
[18] R.N.Bracewell, The Fourier Transform and Its Applications , McGraw-Hill, New York, NY, USA, 1965.
[19] .V.OppenheimandR.W.Schafer, Discrete-Time Signal Processing , Prentice-Hall, Englewood Cliff s, NJ, USA, 1989.
[20] Shun-Zheng Yu and Hisashi Kobayashi , "Practical Implementation of an Efficient Forward – Backward Algorithm for an Explicit- Duration Hidden Markov Model", IEEE Transactions on Signal Processing, Vol. 54, no. 5, pp - 1947-1955, May 2006.
[21] Mark Gales and Steve Young, "The Application of Hidden Markov Models in Speech Recognition", Journal of Foundations and Trends in Signal Processing, Vol. 1, No. 3, pp. 195 – 304, 200

Authors Biography

Dr. S.Kalaivani is currently working as Associate Professor in VIT University, Vellore. She received her B.E. degree in Electronics & Communication Engineering from Government College of Engineering, Salem, Tamilnadu under Madras University, India in 2000 and M.E. degree in Communication Systems from Anna University, Chennai, India in 2006. She completed her Ph.D in 2012 under Anna University, Chennai with the specialization of Image Processing. She is having 15 years of teaching experience in under graduate and post graduate level of Engineering courses in Anna University affiliated colleges. She has presented her research papers in 10 international, National Journals and 6 International conferences, more than 15 National conferences. Her research interest includes speckle reduction in SAR and ultrasound images, Multi-resolution processing, Laplace transforms and diffusion.

Rahul Singh Thakur completed his B.Tech degree in Communication Engineering from VIT University in 2015. His research interest includes signal analysis, audio signal processing and speech signal synthesis.

Dharmendra completed his B.Tech degree in Communication Engineering from VIT University in 2015. His research interest includes signal analysis, audio signal processing and speech signal synthesis