International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 09, September-2018

Gender Recognition System for Manipuri Speech Signal by Using MFCC and Power Spectrum

ROHIT THOUNAOJAM(PG Student)[1]
HUIREM BHARAT MEITEI (M.Tech)[2]
Mrs. SHARMILA MEINAM (Assistant Professor)[3]

Electronics and Communication Engineering Department [1,2,3]
JNTUK,AP,(T.L Bechtel India Ltd,New Delhi)[2]
Manipur Institute of Technology,IMPHAL, Manipur,India[1,3]

Abstract
In this paper we are introducing a technique for gender classification as male or female. This paper proposed designing of feature extraction of Manipuri speech signal and gender Classification using MFCC and Power Spectrum .Previously there are so many technique for the classification, here we are comparing with all the classifiers and testing the results. The classifiers are used to differentiate the gender as male or female. The database for the speech recognition system is collection of male and female voices. The main features for the gender recognition are the base in the voice, softness, hardness in the voice. We are performing the results by using the Mel frequency cepstrum coefficient (MFCC), power spectrum and Euclidian distance classifier. By seeing the experimental results we can say that the proposed method is giving better results.

Keywords: MFCC, Power Spectrum, Euclidian distance, Mel-filter Bank.

I. Introduction

Talk hail does not contain talk information so to speak. Meanwhile, it contains information like age, sexual introduction, and energetic express that are related to the speaker [13]. At the vital period of all unmistakable confirmation system, features are settled. At this stage talk signal is changed over into estimation regards that have indisputable characteristics and less vacillation addressing talk properties. There are diverse systems for this change and they are designated parametric and non-parametric models.

In this paper we depict a speaker-assemble institutionalization estimation that we associated with both introduction institutionalization and speaker-institutionalization. To achieve parameter sharing the acoustic space is partitioned into classes. A biggest likelihood approach has been proposed under which the delta between the movement mean and its relating acoustic class is generally without speaker, however the techniques for the acoustic classes are generally speaker-subordinate. Right when associated with sexual introduction institutionalization, the slip-up rate diminish approaches that of a sex subordinate system anyway with a substantial part of the amount of parameters. For a speaker-institutionalized system, a 30% reducing in screw up rate was obtained in a bunch affirmation break down in a setting subordinate relentless thickness HMM structure.

II. Literature survey

[1] Alex Acero and Xuedong Huang: In parametric system a talk age exhibit is described in light of human talk creation segment. LPC examination of talk is overall used for this point. LPC models talk as the plan relationship of an excitation source E(Z) and an extraordinary framing channel H(Z) addressing vocal tract. Excitation source and vocal tract channel are used as a segment by removing them from talk movement by techniques for cepstral examination. In nonparametric system, Mel-Frequency cepstral coefficients (MFCC) which relies upon human voice perception part are used.

[2] C. Neti and Salim Roukos: Gender subordinate systems are by and large made by part the arrangement data into each sexual introduction and building two separate acoustic models for each sex. This method acknowledge that each state of a subphonetic show is reliably dependent on the sexual introduction. We use the begin that the acoustic recognize of various sub phonetic units are dependent on sexual introduction in moving degrees transversely over phones and simply more

particularly setting subordinate. We exhibit this is to make certain the case by using sexual introduction as a request despite phone setting request in the setting decision trees. Using these trees we collect phone specific sexual introduction subordinate acoustic models and display a novel procedure to pick between genders in the midst of translating in perspective of an extent of conviction of the decoded hypothesis. A difference in 6.3% in word screw up is proficient as for a sexual introduction free system.

[3] R. Vergin, A. Farhat and D. O'Shaughnessy: The makers present another modified male/female course of action procedure in light of the territory in the repeat space of the underlying two formants. This request relies upon another customized formant extraction which is snappier than an apex picking framework. Sexual introduction subordinate acoustic-phonetic models originating from this gathering are used in the INRS steady talk affirmation system with the ATIS corpora. A difference in 14% is gotten with these models conversely with the standard sans speaker system.

III. Proposed method

3.1. Mel-frequency cepstral coefficients

MFCC is a champion among the practically sometimes used features both in talk and speaker affirmation [13, 14]. Stevens and Volkman (1940) likely showed that human hearing structure sees the frequencies specifically up to 1 KHz and logarithmically above it. Association between observed repeat which is called Mel and certified repeat is given in as,

$$Mel(f) = 2595 * \log\left(1 + \frac{f}{700}\right) \qquad (1)$$

MFCC is a cepstral system which changes over talk into parameters as shown by mel-scale. The method whose square blueprint is given in Fig.1 has the going with propels.

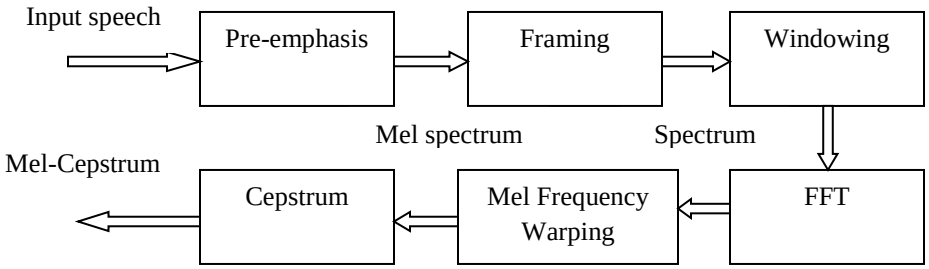

Fig.1 MFCC Block Diagram

1. Pre-emphasis:

In light of the structure of voice creation system, damping occurs in high-repeat territories. Consequently, the scopes of voiced areas are compensated by pre-emphasis which opens up high-repeat locale and performs filtering [15]. By and large used pre-complement channel is given as,

$$Y[n] = x[n] - a * x[n-1], a \approx (0.95 - 0.97) \qquad (2)$$

In this study we took a=0.97 2.

2. Framing and Windowing:

Like in all voice examination techniques, furthermore MFCC methodology is associated along the short bits where voice has stationary acoustic features [13]. These bits are generally picked as 20-30 milliseconds a move of 10-15 milliseconds along the banner. Thus every packaging contains a some piece of its past packaging. In voice applications Hamming window is generally supported. Hamming window is imparted as,

$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \qquad (3)$$

3. Frequency Spectrum:

Talk signals apportioned into examination windows are changed by FFT from time region to repeat space. This documentation addressing repeat allotment of talk signal is called plentifulness extend.

4. Mel-Frequency Warping:

To change over the got plentifulness extend into mel-scale, a channel set straightly concerning Mel-scale is used. This set contains triangle band pass channels that are covering half as showed up in Fig.2. All things considered, channel coefficient is picked some place in the scope of 20 and 30.

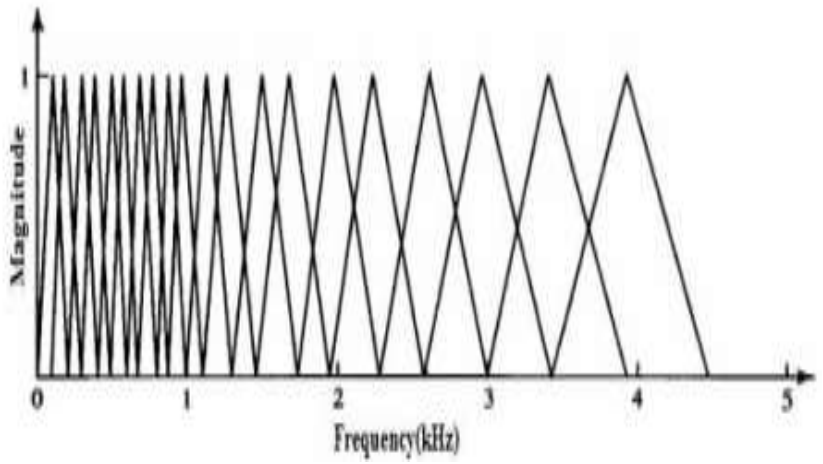

Fig 3.2 Mel Filter Bank

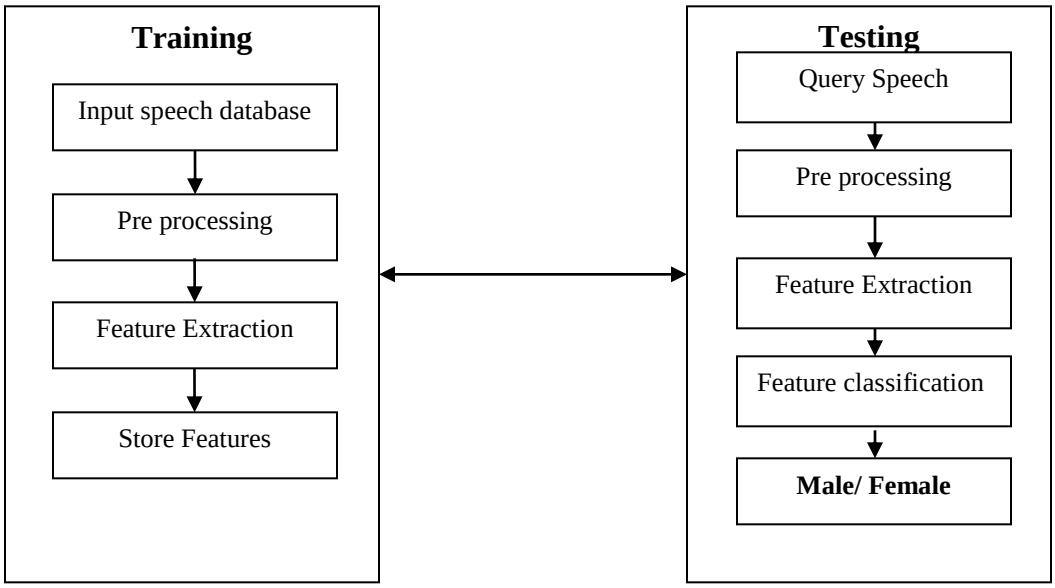

Fig 3.2 Proposed Method Block Diagram

5. Mel Spectrum and Cepstrum:

At this stage, mel go is gotten by expanding sufficiency scope of the banner with mel channel set and taking the logarithm of essentialness that is accessible in each channel. Since the logarithms of mel-extend coefficients are honest to goodness numbers, discrete cosine change given in condition (4) is used to come back to time region. Along these lines, the gained coefficients are called as mel-repeat cepstrum coefficients (MFCC).

$$\widetilde{C_n} = \sum_{K=1}^{K} \left(log\widetilde{S_K}\right) \cos\left(\mathrm{n}\left(K - \frac{1}{2}\right)\frac{\pi}{\mathrm{K}}\right), \mathrm{n} = 1,2,\ldots \mathrm{K} \qquad (4)$$

Here 243, k=1,2,...,K are mel run coefficients. Since the principle part got in view of progress addresses the ordinary logarithmic imperativeness, is removed from incorporate vector.

3.2 Power Spectrum Density

In the past zone we "calmly" evaluated the power supernatural thickness of a limit c(t) by taking the modulus-squared of the discrete Fourier difference in some constrained, inspected stretch of it. In this section we'll do by and large a comparable thing, anyway with amazingly more critical respect for purposes of intrigue. Our thought will uncover a couple of wonders.

The important detail is control go (in like manner called a power ghost thickness or PSD) institutionalization. Overall there is some association of proportionality between an extent of the squared plentifulness of the limit and an extent of the sufficiency of the PSD. Grievously there are a couple of one of a kind conventions for depicting the institutionalization in each region, and various open entryways for getting mistakenly the association between the two spaces. Accept that our

ability c(t) is tried at N centers to make regards c0 ...cN−1, and that these centers length an extent of time T, that is T = (N − 1)Δ, where Δ is the examining break. By then here are a couple of particular depictions of the total power:

$$\sum_{j=0}^{N-1} |C_j|^2 \equiv \text{sum squared amplitude} \qquad (5)$$

$$\frac{1}{T}\int_0^T |c(t)|^2\, dt \approx \frac{1}{N}\sum_{j=0}^{N-1} |C_j|^2 \equiv \text{"mean squared amplitude"} \qquad (6)$$

$$\frac{1}{T}\int_0^T |c(t)|^2\, dt \approx \Delta \sum_{j=0}^{N-1} |C_j|^2 \equiv \text{"time-integral squared amplitude"} \qquad (7)$$

PSD estimators, as we will see, have a significantly more prominent assortment. In this area, we consider a class of them that give gauges at discrete estimations of recurrence fi, where I will extend over whole number qualities. In the following segment, we will find out about an alternate class of estimators that deliver gauges that are constant elements of recurrence f. Regardless of whether it is concurred dependably to relate the PSD standardization to a specific portrayal of the capacity standardization

3.3 Euclidean Distance:

In the speaker acknowledgment stage, an obscure speaker's voice is spoken to by a grouping of highlight vector {x1, x2 … .xi), and afterward it is contrasted and the codebooks from the database. Keeping in mind the end goal to distinguish the obscure speaker, this should be possible by estimating the bending separation of two vector sets in light of limiting the Euclidean distance[6].The equation used to figure the Euclidean separation can be characterized as following:

The Euclidean distance between two points P = (p1, p2…pn) and Q = (q1, q2...qn),

$$= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \qquad (8)$$

The speaker with the lowest distortion distance is chosen to be identified as the unknown person.

IV. Results

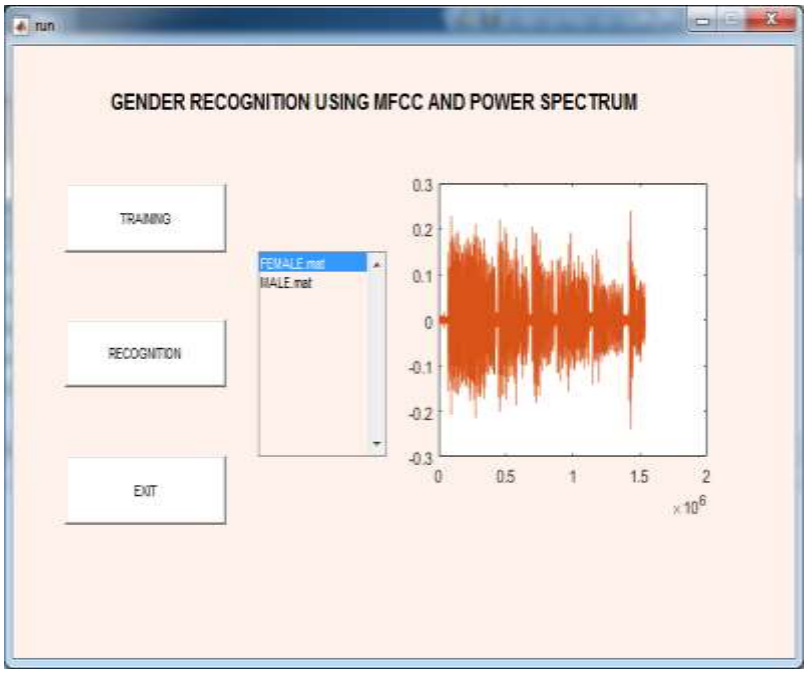

Fig.1 Overall Structure of GUI for Proposed Work

Above there is discussion of overall structure of our development using GUI (graphical user interface). The main steps in this development are training the samples with their features and testing samples for unknown samples for gender identification.

While training and testing we are using main feature as MFCC and power spectrum. Finally for features comparison we used Euclidian distance classifier.

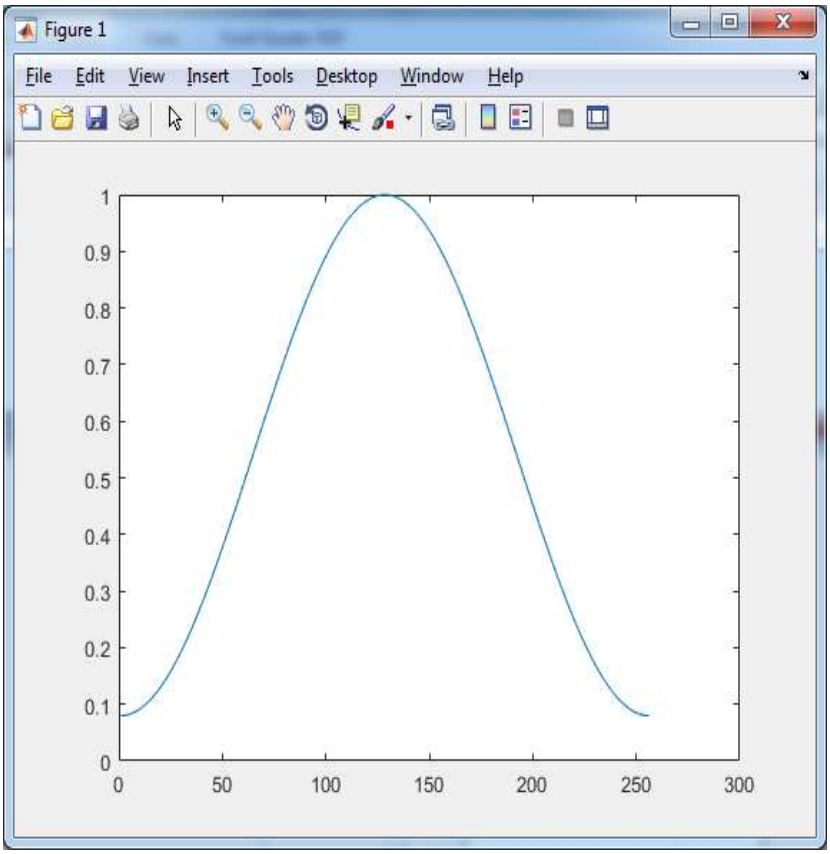

Fig.2 Hamming Window

Here is the Hamming window which is used for analysis and reduction of side lobes in the speech signal.

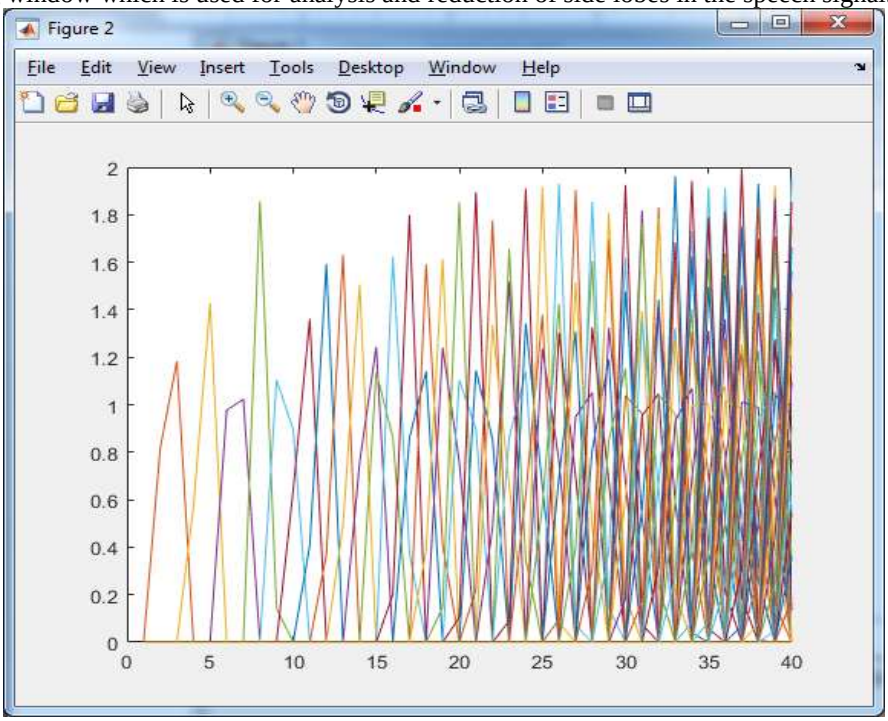

Fig.3 MFCC Features

MFCC features determine the overall shape and structure of speech spectrum. This is very important feature for detection and identification of speech signal.

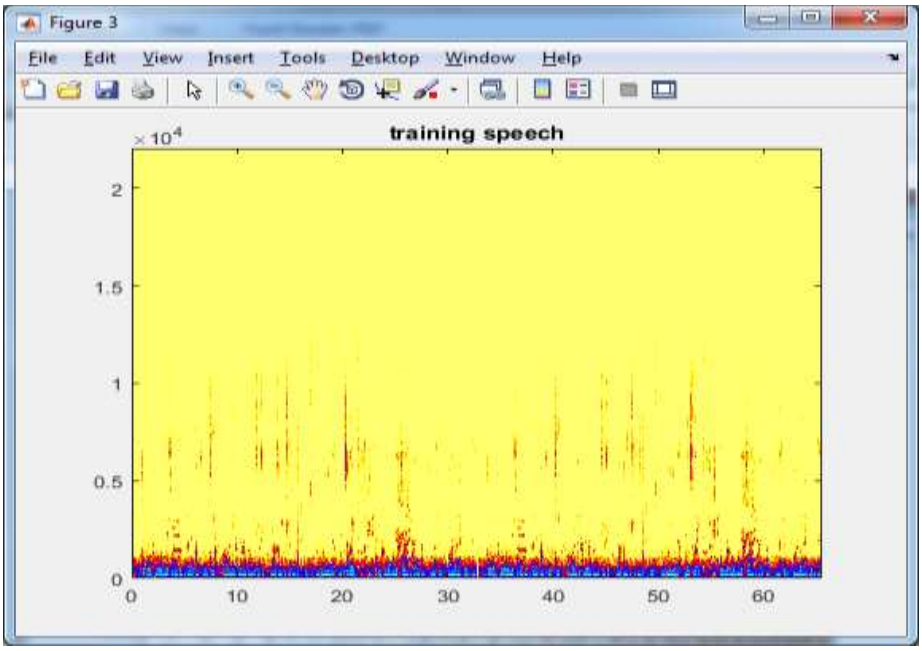

Fig.4 Power Spectrum

This power spectrum represents the visual representation of speech frequency spectrum with respect to time.

Fig.5 Result Obtained as Female Gender

After using the Euclidian distance classifier the index with which features are matching we can retrieve and the gender related to that particular index.

V.Conclusion

In this paper we are doing the classification for gender detection as male or female by using the Mel frequency cepstrum coefficient (MFCC) and power spectrum. The recognition of speech is very difficult process because whenever we are speaking there are some chances for introduction of noise. For the classification we are using Euclidian distance classifier. By including more number of subjects from various ethnic groups, a user independent and general system for gender classification is considered from any type of speech. By using the proposed method we can easily classify the gender as male or female and by seeing the experimental results we can say that this classification is effective for the gender classification.

References

[1] Alex Acero and Xuedong Huang, Speaker and Gender Normalization for Continuous-Density Hidden Markov Models, in Proc. of the Int. Conf. on Acoustics, Speech, and Signal , IEEE, May 1996
[2] C. Neti and SalimRoukos.Phone-specific genderdependent models for relentless talk affirmation, Automatic Speech Recognition and Understanding Workshop (ASRU97), Santa Barbara, CA, 1997.
[3] R. Vergin, A. Farhat and D. O'Shaughnessy, "Intense sex subordinate acoustic-phonetic showing in perpetual talk affirmation in perspective of another customized male/female request", Proc. of IEEE Int. Conf. on Spoken Language (ICSLP), pp. 1081-1084, Oct. 1996.
[4] S. Slomka and S. Sridharan, "Modified sexual introduction ID improved for vernacular opportunity", Proc. of IEEE TENCON'97, pp. 145-148, Dec. 1997.
[5] E.S. Parris and M.J. Carey, Language Independent Gender Identification, ICASSP, pp 685-688, 1996.

[6] Ting, H, Yingchun, Zhaohui, W., Combinning MFCC and Pitch to Enhance the Performance of the Gender Recognition, IEEE, 2006.
[7] D.A. Reynolds and R.C. Rose, "Enthusiastic substance self-sufficient speaker ID using Gaussian mix speaker models," IEEE Trans. Talk and Audio Process., 3 (1), 72– 83,Jan. 1995.
[8] M. H. Sedaaghi, "A Comparative Study of Gender and Age Classification in Speech Signals", Iranian Journal of Electrical and Electronic Engineering, Vol. 5, No. 1, pp. 1-12, March 2009.
[9] Djemili, Rafik, HocineBourouba, and Mohamed Cherif Amara Korba. "A talk hail based sexual introduction recognizing confirmation structure using four classifiers." Multimedia Computing and Systems (ICMCS), 2012 International Conference on.IEEE, 2012.
[10] Abdulla, W. likewise, Kasabov, N. 2001 .Improving talk affirmation execution through sexual introduction division. In Proc. Int. Conf. Fake Neural Networks and Expert Systems (ANNES), pages 218– 222, Dunedin, New Zealand.
[11] Pronobis, Marianna, and Mathew Magimai Doss. "Examination of F0 and Cepstral Features for Robust Automatic Gender Recognition."
[12] Deiv, D. Shakina, and Mahua Bhattacharya. "Customized Gender Identification for Hindi Speech Recognition.". Overall Journal of Computer Applications (0975 – 8887). Volume 31– No.5, October 2011
[13] L. Rabiner and B.- H.Juang, Fundamentals of Speech Recognition, Englewood Cliffs (N.J.), Prentice Hall Signal Processing Series, 1993.
[14] J. R. Deller, J. H. L. Hansen, J. G. Proakis, Discrete-Time Processing of Speech Signals, IEEE Press, Piscataway (N.J.), 2000.
[15] Picone, J., Signal exhibiting frameworks in talk affirmation, Proceedings of the IEEE 81, pp. 1215– 1247, 1993.
[16] Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
[17] G.McLachlan, Mixture Modeles. New York: Marcel Dekker, 1988
[18] Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Most outrageous likelihood from lacking data by methods for the EM algorithm."Journal of the Royal Statistical Society. Plan B (Methodological) (1977): 1-38.