*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 4, Ver. II (July – Aug. 2015), PP 01-08*
*www.iosrjournals.org*

# Government Policies Search Using Marathi Speech Recognition System– Based On Robust Feature Extraction Method

Mugdha Parande[1], Prof. Shanthi Therese[2] and Prof. Vinayak Shinde[3]
*[1]Mumbai University, Shri L.R. Tiwari College of Engineering, Kanakia Park, Mira Road (East). Thane,. Maharashtra, India*
*[2] ThadomalShahani Engineering CollegeP. G. KherMarg, (32nd Road), Linking Road. Bandra (West), Mumbai - 400 050.*
*[3] Shri L.R. Tiwari College of Engineering,KanakiaPark,Mira Road(East).Thane,India*

***Abstract:****Country like India, where 70% of the population lives in village and rural parts of the country, it becomes more necessary to utilize such tools for their social, economic, administrative and governance regeneration. The objective is to facilitate and improve Panchayat functioning on day-to-day basis, through two ways flow of information and content. The propose speech recognition application can facilitate most effective and reliable service to the user.*

***Index Terms:****Feature Extraction approach, Gammatone filter, SNR parameter, Marathi Speech recognition.*

## I. Introduction

Speech is the most natural means of communication between humans; it can be done without any tools or any explicit education. It is one of the first skills we learn to use. Speech is also the most important way of communicating. It has always been; before mankind invented writing, the spoken word was the only way of passing knowledge. Ancient poets like Homer and Ovidius originally wrote their even now famous epic poems for recitation not for reading.

Despite all our novel ways of communicating, like e-mail and chat, speech is still the number one means of communication, a fact once again proven by the immense popularity of cellular phones. So it is only logical that machine interface designers in their quest for a natural man-machine interface have turned to automatic speech recognition and speech production as one of the most promising interfaces. In the last 30 years researchers from areas like psychology, linguistics, and electrical engineering and computer science have worked on this subject. While the first systems could only differentiate between 'yes' and 'no', currently, speaker independent systems exist with a vocabulary of over 60000 words that can recognize continuous speech (that is complete sentences or paragraphs) with an accuracy of 90% .

## II. Problems With Speech Recognition

There are many issues which limit the performance of a speech recognition system. The main barrier is the variability of speech signal. A given word spoken by different persons can have different spectral and temporal properties due to variation in physiological characteristics, emotional status and cultural background. For example, females, in general have a shorter vocal tract than a male.

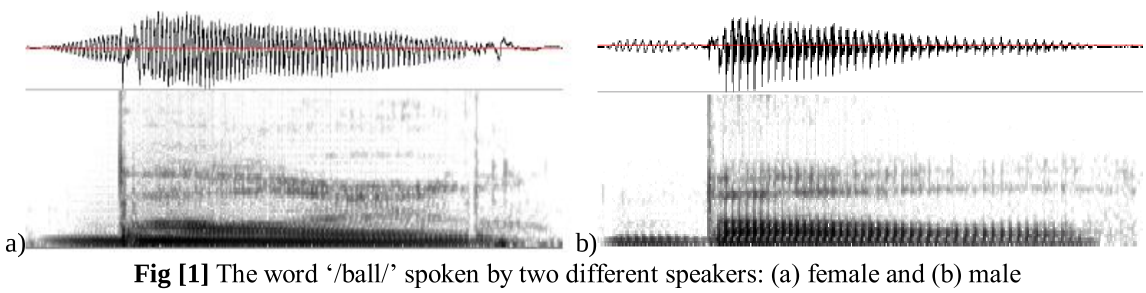

**Fig [1]** The word '/ball/' spoken by two different speakers: (a) female and (b) male

The figure [1] shows the word ball spoken by two different persons. The upper half of the figure shows the raw speech waveform, the lower half shows processed versions of the signal that highlight its formants, which are characteristic for a sound, So the formant frequencies of a vowel spoken by female speakers are, in general, higher than male.

Speech can be adequately described in terms of linguistic units called phonemes. For example, each character of Devanagari script represents one phoneme. However, there are no well-defined boundaries between phonemes in continuous speech. The spectral characteristics change continuously due to the inertia of the articulators which moves from the position of one phoneme to the position of the next phoneme. Consequently, the acoustic properties of a speech sound not only depend on the identity of the corresponding phoneme, but also on the neighboring sounds. This variability due to phonetic context, however, can be predicted unlike speaker dependent variations.

First of all note that the speaker has mispronounced the word. In other word, pronunciation of same word may differ by same user or different user. Another scenario is that two different words have same pronunciation for example “see “and “sea”. Human beings have no problem in translating this to the correct word. However, such things pose a problem for a machine.

Figure [2] shows the phoneme /e/ spoken by the same person in a number of words. The exact shape of the speech signals also depends on the speed with which is spoken and the mood and the temper of the speaker.

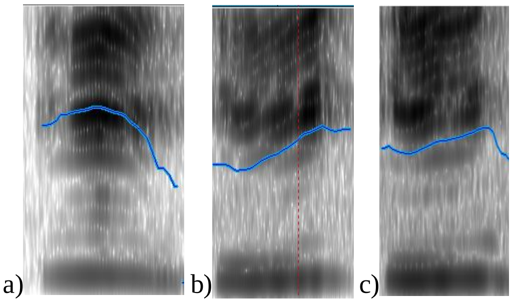


**Fig [2]**The phoneme /e/ spoken by the same person

Superposition of background noise and extraneous signals in the transmission channel decrease the signal to noise ratio of speech signal. Acoustic characteristics of room, microphone and transmission channel get convolved with the speech signal. All such effects add to the variations of speech signal.

## III. Proposed Application Of Speech Recognition

There are various ways to communicate with each other such as writing, speaking etc. Speech is the most desirable medium of communication. As we know Government of India facilitates tones of policies and scheme for development and growth but every person cannot take benefits from it because lack of policies reaches. This is the main hurdle which resists our India to reach at the top of the world. The propose recognition application help to interact with system and to get Government of India’s policies and schemes.

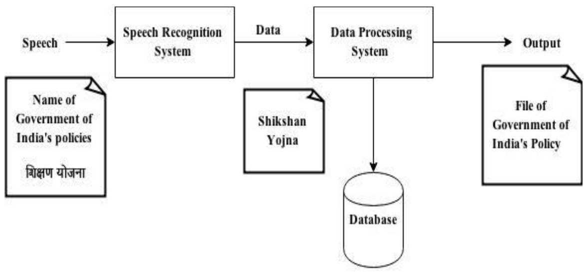


**Fig [3]** Proposed application for speech recognition

In the Figure [3] shows flow of the proposed speech recognition application. There are two modules:-

1) Speech Recognition system to recognize sound data and convert it into text data which uses proposed speech recognition approach as mentioned above.
2) Processing System to perform Retrieval function which retrieves Government of India’s policies and schemes from database.

Finally displaying the output as per the user requires. In the propose application some highlight features which makes this application more user friendly and real time applicable. The propose recognition application is user independent. In other word, system is train to work efficiently in multiple user environment. This application now has a limited scope because it trains for Marathi language dataset. It recognize name of the government of India’s policy and scheme spoken by user in Marathi language.

## IV. Sequence Diagram Of Proposed Speech Recognition System

In figure [4] shows proposed speech recognition system, voice input is given by user to Voice Type class. Voice Type Class converts input signal into short array by using MicroInput class. Array of data is given to Feature Extraction class as an input. Feature Extraction class is return MFCC array as an output. It applies to Point class which is mapping data based on indices. For mapping function codebook is use which is generated into training phase. When return word from dictionary to VoiceType class. Same word is use as input to retrieve file from local hard disk and display it.

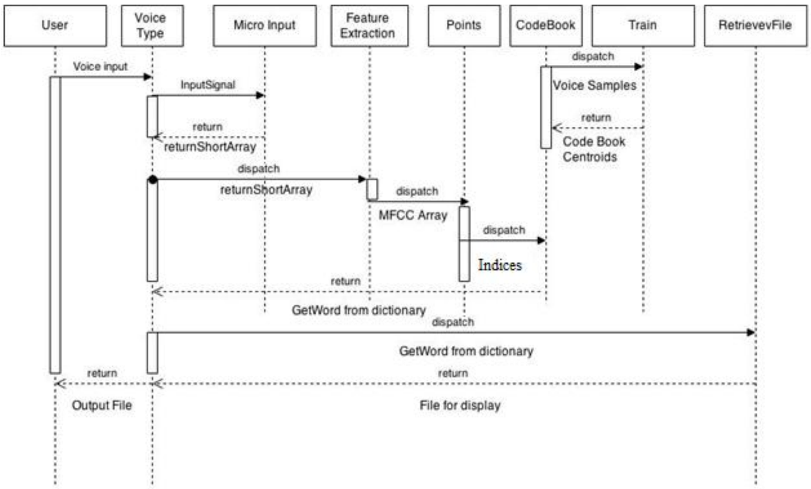


**Figure [4]** Sequence diagram of proposed speech recognition system

## V. Proposed Speech Recognition Application

Proposed Speech Recognition Application composes the following steps

1. Need to Train dataset.
2. Creates dictionary for mapping codebook.
3. After training of dataset, Vector Quantization generate text file which contain features.
4. Start the application with Login Username and Password (To administer for security purpose).
5. Search Indian Government policies and schemes with one's voice input.
6. Retrieve Indian Government policies and schemes.

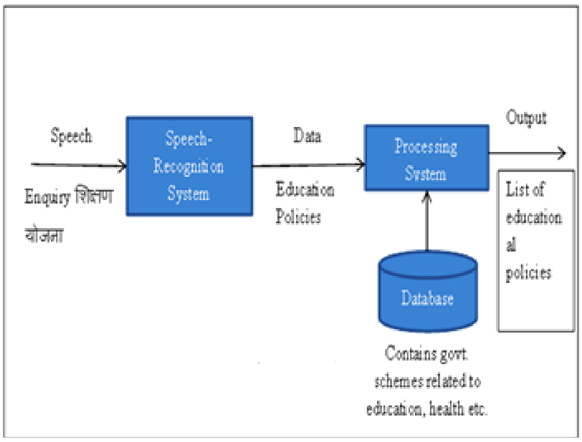


**Fig [5]** Data flow of proposed speech recognition application

In the figure [5] shown the data flow of proposed speech recognition application. Speech is input to speech recognition system process which recognized speech and convert into text data. Recognized text data is input to processing system which can retrieve government of India's policies or schemes. Finally retrieve file is display on the screen. In the figure [5] shown that user ask for 'शिक्षणयोजना' it is recognize by speech recognition system such as 'Shikshanyojna' and passed it as input to processing system which retrieve file of education policy and display it on the screen.

## VI. System Implementation

Working of proposed speech recognition Application with snapshots is given below:-
Login page is for administration purpose. Proposed application is served in public sector so it is supported to open access feature.


**Fig [6]** Login Page of proposed speech recognition application

Administrator need to train data with 20 samples for each word. To demonstrate, 15 words trained by 5 different users where each word with 20 samples (4 samples of each user) in Marathi language and 20 words trained by single user similarly each word with 20 samples in Marathi language.To generate VQ codebook for new word click on FILE -> NEW. It opens new dialogue box which shows in figure [7].

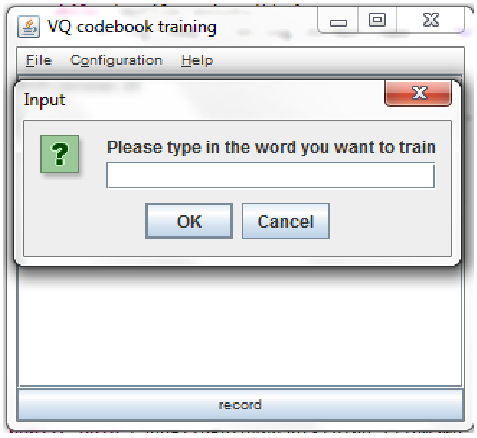


**Fig [7]** Dialogue box to generate VQ code book for new word

Enter word that you want to train and click 'Ok' button. In figure [8] shows that 'Sukanya' word type for train dataset and generate VQ codebook for it.

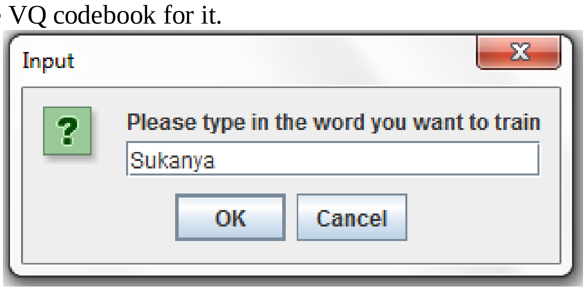


**Fig [8]** 'Sukanya' word type for training process

Select file Path to create VQ codebook after finished training process for new word. In figure [6.4] shows the details such as num_samples (20), word to train (Sukanya) and file path to generate VQ codebook (C:\Test\Sukanya)

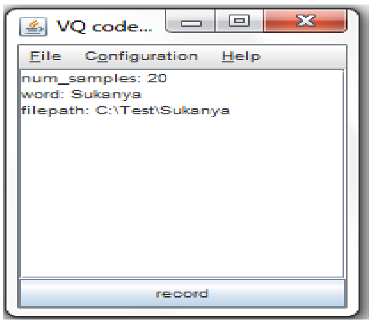


**Fig [9]** Select path to create VQ file for 'Sukanya' word

Click on 'Record' button and record word such as 'Sukanya' and then click on 'stop' button. This process is repeated 20 times. It means your 20 samples is recorded for each new word. After training process is finished, VQ codebook of 'Sukanya' is generated at given file path.

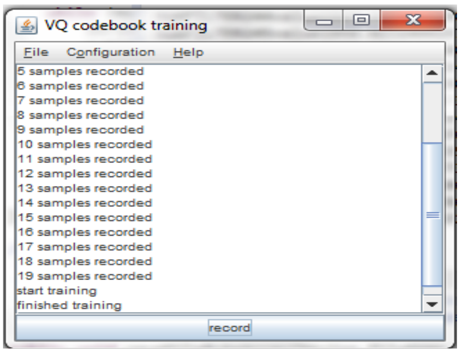


**Fig [10]** Training process for new word

Open VQ codebook file it shows like figure [11]. These are features extracted from recorded samples based on 12 dimension and 256 frames.

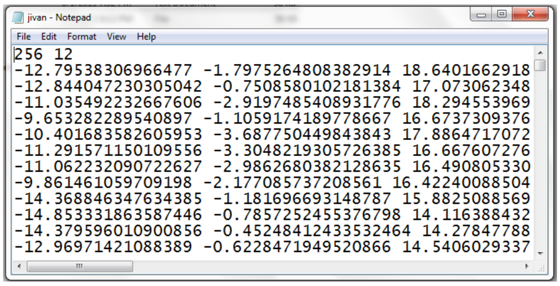


**Fig. [11]** Generated VQ code book contains

Dictionary contains list of words which are trained. To demonstrate 20 words are trained for speaker dependent model and 15 words are trained for speaker independent model in Marathi language.

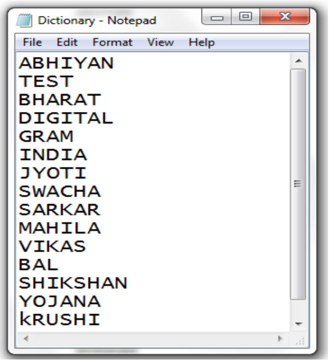


**Fig [12]** Dictionary for proposed application

On user side, proposed speech recognition application displays dialogue box which shows in figure [6.8]. Dialogue box contains four buttons and one text area. 'Clear text' button is to clean text from the text area. 'Record' button is record voice and process it for recognition.

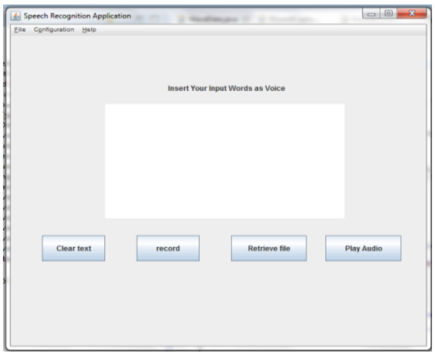


**Fig [13]** Screen of recognition in application

After word is recorded word by application. It displayed in text area such that in figure [6.9] shown the 'Sukanya' word into text area.

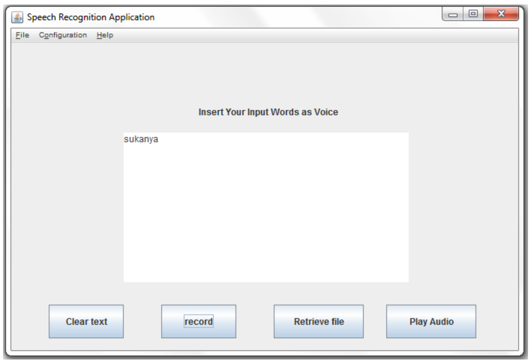


**Fig [14]** Screen of recognition in application

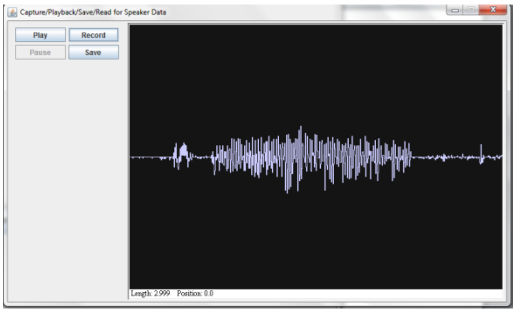


**Fig [15]** Playing audio which is recorded by user

Proposed Application is trained Marathi language to generate VQ codebook file. Click 'Record Button' and speak name of the policy or the scheme. Automatically, spoken voice is converted into text and displays into the text area. Click 'Retrieve Button' to get the policy or the scheme file from database on the application screen. Government of India's policies and schemes files are store into local hard disk.

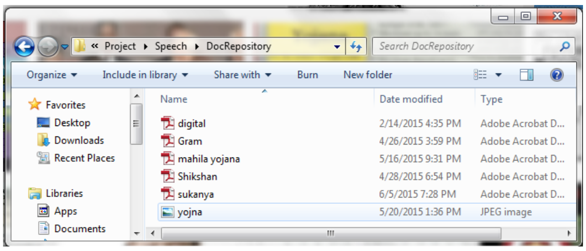


**Fig [16]** Files stored into local hard disk

## VII. Average Percentage Of Speech Recognition Accuracy Rate

To compare between speaker dependent and speaker independent without and with background noise respectively. In this section, determined the result of accuracy in four different scenarios and compared their results.

Table [I] Average of speech recognition accuracy rate

$$\mathbf{Avg} = \frac{\text{Totalofaccuracyrecognitionrate}}{20}$$

| Parameter | Speaker-dependent (In %) | Speaker-Independent (In %) |
|---|---|---|
| Without Background noise | 94.45 | 95.46666667 |
| White Noise | 94.2 | 94.8 |
| Train Noise | 85.65 | 88.33 |
| Car Noise | 86 | 87.33 |
| Thunder Storms Noise | 83.5 | 85.33 |
| Bell Noise | 86 | 84.67 |

Speech recognition accuracy of speaker independent is more than speaker dependent because in this experiment result of speaker independent with 5 different user consists of 3 male and 2 female voice samples which generated strong training set which is more accurately recognized even in the noisy environment than speaker dependent.

## VIII. Conclusion

The Speech recognition application can effectively be used in public sector. This application is very useful to illiterate people in respect to being familiar with Indian Government policies and schemes. Implementation of application based on MFCC with Gammatone Filter as pre-process which helps to improved accuracy of recognition in noisy environment. Proposed Speech recognition system needs well trained dataset because training dataset does most impact on accuracy of recognition. Speech recognition is difficult because the system performance is depended on real time input therefore minor mistake in input leads to wrong perceptions. Proposed system performed much better than traditional MFCC. It is supported both speaker dependent and speaker independent model effectively and generated efficient result. Marathi Language is used as training dataset which is one of the important features of proposed system because it expanded the application's usage area. There are some drawbacks of proposed system like needs of wide search area for policies and improvement in probability of getting search estimation more accurately.

## References


[1]. Jun qi, dong wang, yijiang, runshengliu, "auditory features based on gammatone filters for robust speech recognition," ieee transactions on audio, speech, and language processing, vol. 10, no. 10, pp. 1109-11023, march,2012.

[2]. Vimala.c, radha.v, "suitable feature extraction and speech recognition technique for isolated tamil spoken words," international journal of computer science and information technologies, vol. 5 (1) , 2014, 378-383.

[3]. Hynkeyhermanskey , "design and development of automatic speech recognition of isolated marathi words for agricilturalpurpose,"iosr journal of computer engineering, vol.16,issue no. 3, pp. 79-85, may-jun 2014.

[4]. R. K. A. A. M. Dave, "using gaussian mixtures for hindi speech recognition system," international journal of signal processing, image processing and pattern recognition, vol. 4, no. 4, pp. 50-64, december 2011.

[5]. Octavian cheng, waleedabdulla, zoransalcic, report on" performance evaluation of front-end processing for speech recognition systems," department of electrical and computer engineering department, the university of auckland. March 2012.

[6]. Bharti w. Gawali, santoshgaikwad, pravinyannawar, sureshc.mehrotra"marathi isolated word recognition system using,", vol. 01, no. 01, pp. 21-26, 2011.

[7]. Lawrence rabiner,biing-hwangjuang and b.yegnanarayana. "fundamentals of speech recognition", dorlingkindersley (india) pvt. Ltd. 2009. Pp 135-136.

[8]. Shanthitherese s., c.l. "review of feature extraction techniques in automatic speech recognition," international journal of scientific engineering and technology, vol. 02, no. 06, pp. 479-484, 2013.

[9]. R. K. A. A. M. Dave, "using gaussian mixtures for hindi speech recognition system," international journal of signal processing, image processing and pattern recognition, vol. 4, no. 4, pp. 50-64, december 2011.

[10]. A thakur and r. Kumar, "automatic speech recognition system for hindi utterances with regionalindian accents: a review 1."No. 10, pp. 1109-1119, march 2012.

[11]. Agarwal, k. Samudravijaya, and k. Arora, "recent advances of speech database development activities for indian languages," in international symposium on chinese spoken language processing (iscslp 2006), 2006.

[12]. G. Anumanchipalli, r. Chitturi, s. Joshi, r. Kumar, s. P.singh, r. Sitaram, and s. Kishore, "development of indian language speech databases for large vocabulary speech recognition systems," in proc. Specom, 2005.

[13]. S. S. D. R. Devamita, aggarwal, "on the performance of front-ends for hindi speech recognition with degraded and normal speech," symposium on translation support systems, strans 2001, iitkanpur, 15.

[14]. M. I. Vishal chourasia, samudravijaya k and m. Chandwani,"hindi speech recognition under noisy conditions,"j. Acoust. Soc. India, vol. 54.

[15]. M. Kumar, r. K. Aggarwal, g. Leekha, and y. Kumar,"ensemblefeature extraction modules for improved hindi speech recognition system," International Journal of Computer Science, vol. 9.

[16]. P. Saini and P. Kaur, "Automatic speech recognition-areview," International journal of Engineering Trends & Technology, pp. 132–136.
[17]. S. S. D. R. DevAmita, Aggarwal, "On the performance of front-ends for hindi speech recognition with degraded and normal speech," Symposium on Translation Support Systems, STRANS 2001, IIT Kanpur, 15.
[18]. M. I. Vishal Chourasia, Samudravijaya K and M. Chandwani,"Hindi speech recognition under noisy conditions,"J. Acoust. Soc. India, vol. 54.
[19]. M. Kumar, R. K. Aggarwal, G. Leekha, and Y. Kumar,"Ensemble feature extraction modules for improved hindi speech recognition system," International Journal of Computer Science, vol. 9.
[20]. R. M. S. Chanwoo Kim, "PNCC in Matlab," 2012.[Online]. Available: http://www.cs.cmu.edu/ robust/archive/algorithms/PNCC_IEEETran.
[21]. B. Milner, "A comparison of front-end configurations for robust speech recognition.," in ICASSP, pp. 797–800,IEEE, 2002.
[22]. D. P. W. Ellis, "PLP and RASTA (and MFCC,and inversion) in Matlab," 2005. [Online]. Available:http://www.ee.columbia.edu/ dpwe/resources/matlab/rastamat/.
[23]. D. Pearce and H.-G. Hirsch, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions.," in INTERSPEECH,pp. 29–32, ISCA, 2000.
[24]. S. Young, "The htk hidden markov model toolkit: Designand philosophy," Entropic Cambridge Research Laboratory,Ltd, vol. 2, pp. 2–44, 1994.
[25]. J. Picone, "Signal modeling techniques in speech recognition," Proceedings of the IEEE, vol. 81, no. 9,pp. 1215–1247, 1993.
[26]. Audacity Manual "http://www.audacity/tutorialmanual.pdf"
[27]. PraatManual"htttp://www.praat/tutorialmanual.pdf"
[28]. Speech Processing, Transmission and Quality Aspects (STQ);Distributed Speech Recognition; Advanced Front-end FeatureExtraction Algorithm; Compression Algorithms, European Telecommunications Standards Institute ES 202 050, Rev. 1.1.5,Jan. 2007.