Removal of Empty Bands from Speech Signals in Hindi

Dharamveer Sharma[1], Sukhdeep kaur[2]

[1]Department of Computer Science, Punjabi University, Patiala, Punjab, India

[2]School of Applied and Life Sciences, Uttaranchal University, Dehradun, Uttarakhand, India

E-mail: sukhdeepkaurk091@gmail.com

Abstract

In order to provide speech analysis, speech production and perception can be exploited with the help of various speech processing applications. Few properties or features can be extracted from speech signal S(n) with the help of speech analysis. With the objective of simplifying the speech signal and for removing the redundancy present within the speech signal, the transformation of S(n) is done into another signal. In this research work, the novel technique is proposed to remove empty bands from the speech signal. The proposed technique is based on the threshold value for the removal of empty bands. In the proposed technique, the frequency of each band is calculated and a band which has frequency below threshold value is removed from the signal. The performance of the proposed technique is tested in MATLAB and it is analyzed that proposed technique performs well in terms speech enhancement.

Key words: Empty bands, LDA.

1. Introduction

The easiest and most commonly used mode by human beings to perform communication is speech. With the help of speech, the information can be exchanged in the most natural and efficient manner. Thus, natural language speech recognition is considered to be an important area of research today. The mechanism through which the speech signal is converted into a sequence of words using an algorithm is known as speech recognition(Stolbov, M., et.al,2014).The conversion of text into speech in an automatic manner is known as Text-to-Speech synthesis (TTS). The computer is allowed to speak through this TTS technology. This process is very similar to the native speaker that reads the text in a particular language. There are two major phases involved within the TTS process(Prudnikov, A., et.al,2015). The input text is converted into a phonetic or any other linguistic representation through text analysis which is the initial phase. Further, from this phonetic and prosodic information, output is generated which is known as generation of speech waveforms and is the secondary phase(Deng, L., et.al,2004). Usually, high and low-level syntheses are the alternative names of these two phases. The information from high-level is used to generate the speech sound at the end using the low-level synthesizer(Korenevsky, M., et.al,2016). The text-to-phoneme or grapheme-to-phoneme conversion is known as the process through which the phonetic transcriptions are assigned to the words(Hirsch, et.al,2000). At the front end, the output is generated which is a symbolic linguistic representation that is made up by the phonetic transcriptions and prosody information. The symbolic linguistic representation is then converted into sound by the back-end which is commonly known as the synthesizer(Tomashenko, et.al,2014). Then it generates a prosody.

2. Research Methodology

The MFC application is developed by using concatenation method for implementing speech synthesis. Within visual Studio ultimate edition, the MFC application is generated. Generally, there are standard Windows applications, dialog boxes, forms-based applications, Explorer-style applications, and Web browser–style applications which are the five different types within which MFC executables fall basically. For its implementation, a dialog box is utilized. The Microsoft Foundation Class (MFC) Library is used as a base to generate an MFC application for Windows(Long Zhang, et.al,2017). The MFC application wizard is used for generating an MFC application in a very simple manner. Within the Visual Studio Express editions, the support of MFC projects is denied.

2.1. Pre-processing of text: The input text can be entered into a text box by the user. For entering the text, a dialog box was opened.

2.2. Segment entered text: Further, the entered text is segmented into words and then into graphemes within the next step. For any given language, the smallest unit of a writing system is the grapheme. Thus, the division of each word into its grapheme units is done here. For instance, the text "मेरानामनेहाहै" is entered by the user.

Further, following words are generated by segmenting the text:

मेरा, नाम ,नेहा, है

Further, the segmentation of each word is done into graphemes as follows: मेरा = मे,रा

नाम=ना,म

हा is left to be similar within the graphemes list.

2.3. Database Development: The pre-recorded sound units that are stored within the database are utilized by the concatenation synthesis of speech. Thus, the development of database is the most important requirement for implementing the concatenation method. Each grapheme is recorded here to generate phoneme since for any language phoneme is considered to be the smallest speech unit.

- Selection of graphemes: Unique grapheme units of Hindi language are identified initially for the development of database. Through the analysis it is seen that the database was developed by converting 830 graphemes to phoneme.
- Recording of phonemes: A native male or female Hindi speaker records each grapheme. Also, the recording can be done within the studio at specific pitch, bit rate or at other prosody properties. Any kind of noise present within the recorded speech is eliminated with the help of Audacity which is software that helps in improving the quality of recorded speech(Yash Vardhan Varshney, et.al,2017).

Labelling of phoneme: For instance, the labelling of sound "की" is to be labelled by the name itself, if it is recorded, such as "की.wav". This is done because within the wave file format that includes .wav extension, this each recorded sound is saved.

2.4. Concatenation of Phonemes: For synthesizing the speech, the concatenation of phoneme units is done with the help of concatenation algorithm. C++ programming language s used to write the code. For instance, the phonemes of "मेरानामनेहाहै" are found from the database.

2.5. Remove Empty Band and Re-Generate Signal: The signal which is generated by the text to speech convert has noise or contain empty band which affect quality of the signal. The threshold-based technique is applied which can remove empty band from the signal(Mehmet AlperOktar, et.al,2016). When the empty bands get removed from the signal it leads to increase quality of the speech signal.

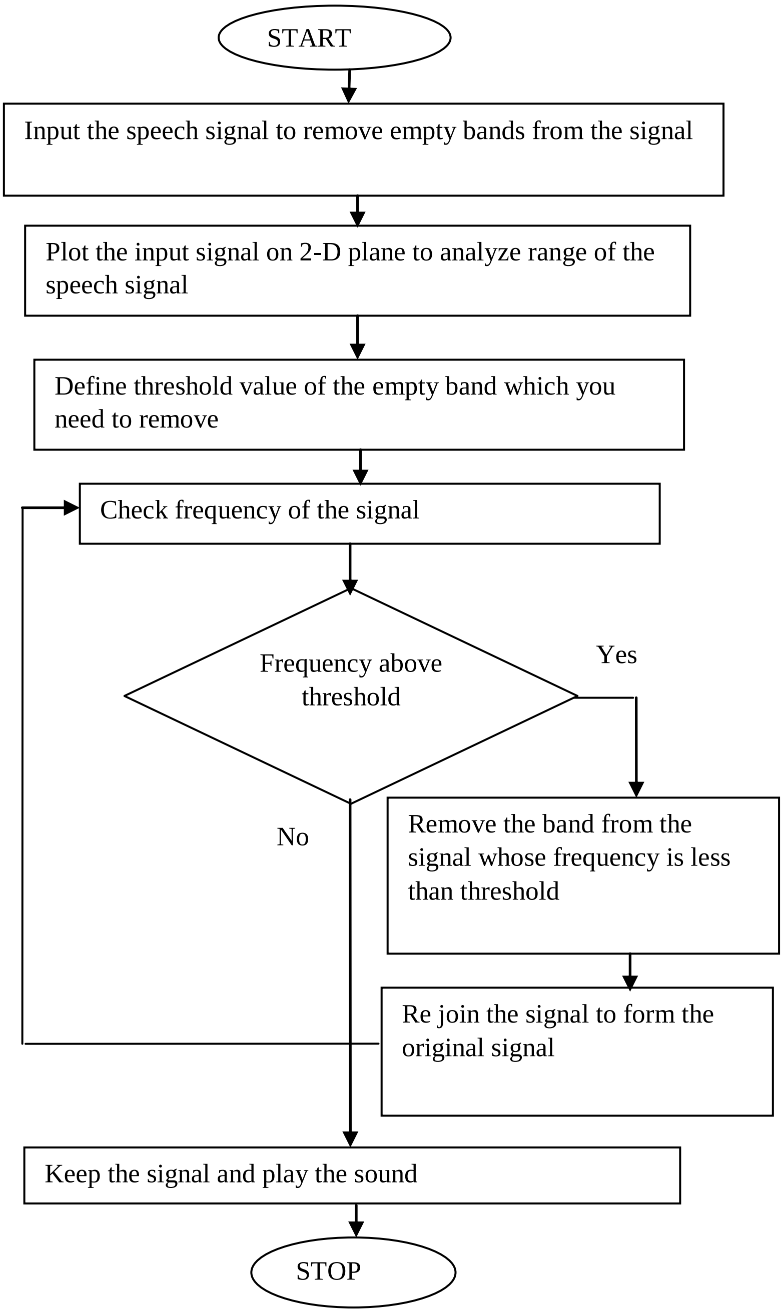

START
Input the speech signal to remove empty bands from the signal
Plot the input signal on 2-D plane to analyze range of the speech signal
Define threshold value of the empty band which you need to remove
Check frequency of the signal
Frequency above threshold
Yes
No
Remove the band from the signal whose frequency is less than threshold
Re join the signal to form the original signal
Keep the signal and play the sound
STOP

Figure 1: Proposed Flowchart

3. Results and Discussion

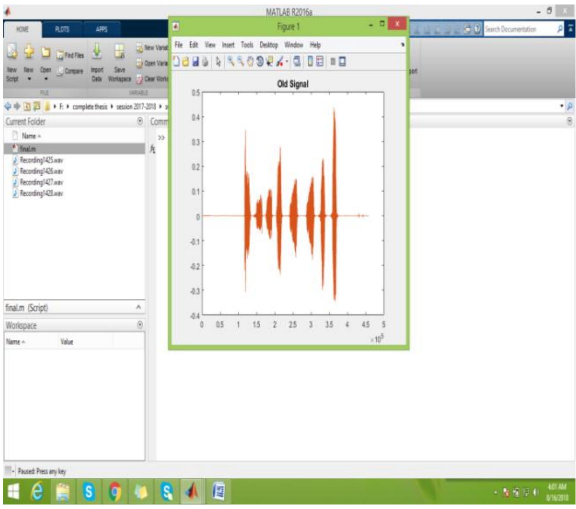
The proposed work is implemented in MATLAB and the results are evaluated as shown below.

3.1 Signal with empty bands

As shown in figure 2, the speech signal which is generated with the text to speech converted is displayed in the figure. The generated speech signal has the empty spaces and we have analyze the threshold frequency of the empty bands

3.2 Signal without white spaces

As shown in figure 3, the speech signal which is generated with the text to speech converted is displayed in the figure. The generated speech signal has the empty spaces and we have analyzed the threshold frequency of the empty bands. The frequency which is above the threshold value is kept and all other frequency bands will be removed from the speech signal.



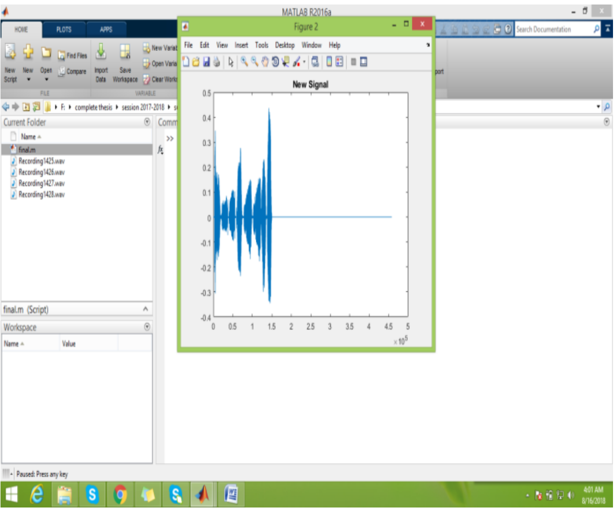


Fig.2 **Fig.3**

4.Conclusion

Natural language speech recognition is considered to be an important area of research today. The mechanism through which the speech signal is converted into a sequence of words using an algorithm is known as speech recognition. In this research work, the text to speech converter is designed which can convert the Hindi text into Hindi signal. It is analysed that in some text to speech converters the empty bands are present which affect its efficiency. In this work, the module is added which can remove empty bands from the signal. To remove empty bands from the signal, the threshold-based technique is proposed(Wang, et.al,2016). In that technique, the bands which have high frequency than the threshold value is kept and all other willbe removed from the signal. The Proposed algorithm is implemented in C sharp and MATLAB. The MATLAB is used to remove empty bands from the input signal.

Conflict of interest:

Authors declare that they have no conflict of interest.

References:

1. Stolbov, M., et.al., “Speech enhancement with microphone array using frequency-domain alignment technique”, 2014, Proceedings of 54-th International Conference on AES Audio Forensics, pp. 101–107
2. Wang, et.al “MMSE-LSA based Wavelet Threshold Denoising Algorithm for Low SNR Speech”, 2016, JieWei, Ming IEEE
3. Prudnikov, A., et.al, “Adaptive beamforming and adaptive training of dnn acoustic models for enhanced multichannel noisy speech recognition”, 2015, Proceedings of 2015 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2015), pp. 401–408
4. Deng, L., et.al, “Enhancement of log mel power spectra of speech using a phase-sensitive model of the acoustic environment and sequential estimation of the corrupting noise”, 2004, IEEE Trans. Speech Audio Process. 12(2), 133–143

5. Korenevsky, M., et.al, “Feature space VTS with phase term modeling”, 2016, Speech Comput. Lect. Notes Comput. Sci. 9811, 312–320
6. Hirsch,et.al, “The aurora experimental framework for the performance evaluations of speech recognition systems under noisy conditions”, 2000, Proceedings of ISCA ITRWASR2000 on Automatic Speech Recognition: Challenges for the Next Millennium
7. Tomashenko, et.al, “Speaker adaptation of context dependent deep neural networks based on MAP-adaptation and GMM-derived feature processing”,2014, Proceedings of Interspeech, pp. 2997–3001
8. Long Zhang, et.al “Supervised Single-Channel Speech Dereverberation and Denoising Using a Two Stage Processing”, 2017, IEEE
9. Yash Vardhan Varshney, et.al, SNMF Based Speech Denoising With Wavelet Decomposed Signal Recognition”, 2017, IEEE
10. Mehmet Alper Oktar, et.al, “Denoising Speech by Notch Filter and Wavelet Thresholding in Real Time”, 2016, IEEE