Retrieval of Speech Information Efficiently in Smart Phones

Dr. G. Shanmugasundaram
Assistant Professor,
Dept. of Information Technology,
SMVEC

S. Soumya
Final Year,
Dept. of IT,
SMVEC

F. Monica Rexy
Final Year,
Dept. of IT,
SMVEC

D. Suruthi
Final Year,
Dept. of IT,
SMVEC

Abstract: **In this paper, an adaptive speech rate control technology is described to perform skimming that is similar to ultrafast listening. In today's world people prefer to listen audio books instead of reading. Thus the scope of audio books will increase when compared to reading printed materials. Here people feel difficulty in recalling the points and terms which they had read earlier. So the system must provide a method equivalent to skimming which will efficiently provide the needed information. Thus we have developed an adaptive speech rate conversion method to perform skimming at very fast play rate. This will also help visually impaired people to enjoy audio books.**

Index Terms – audio books, high-speed speech, speech rate conversion, visually impaired

1. INTRODUCTION

To gather information and data, people used to read printed books or relevant materials. That might be very difficult to concentrate in that specific progress for a long period of time. And so there is a chance of missing certain sentences in the middle of a paragraph. In order to avoid this we can go for audio information system. A fast listening algorithm must be developed and hence can be used by all people. This may also be difficult for people with normal eyesight and less reading practice. Hence an ultra-speed speech playback technology must be used for both blind and

Visually - impaired people. People find it difficult when a huge context is given in a text based information.

2. BACKGROUND

2.1. Related Work

Many methods have been proposed mainly for extracting the semantic clues from scenes. But they fail to catch the rate of speech to efficiently extract the contents. The Info media system helps us to play only the important segments and also uses fast play technique to retrieve information. Thus it will be effective in the case of skimming the speech contents. But this may reduce the capturing of information in case of listening quickly. The earlier research has shown that listening comprehension of audio playback that is increased uniformly will be limited to three times the normal speed. Till date there is no design to listen at high-speed playback. Thus it becomes difficult to listen when the speech rate exceeds the original speed.

2.2. Adaptive SRC method

This technology is used to capture the words and phrases efficiently. It allows even the ordinary users to listen the speech at faster rates. They may eliminate the less important information and pauses when they come across full-stop, comma, etc,. Fig.1 explains the concept of how the text is converted to speech.
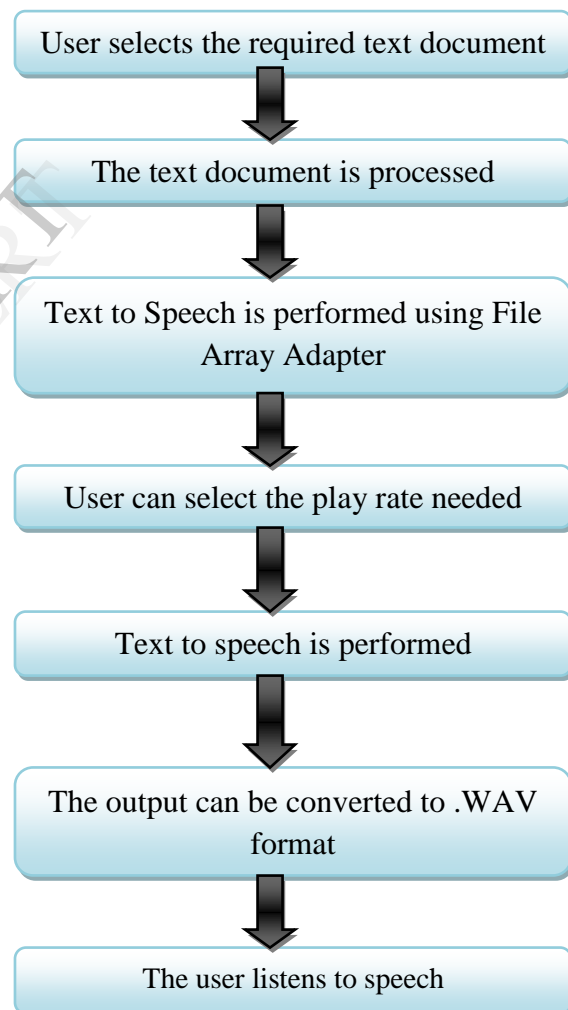

Fig.1. Flow chart of speech retrieval

2.3. New Listening Technology for all

This technology is mainly to focus on fulfilling the requirements of visually impaired people. The recent touch screen devices are difficult for the visually impaired people

and hence it must include some human-machine interfaces which make it easy for them. In future this method is expected to be effective for all people with respect to convenient listening to various multimedia applications.

3. ALGORITHM

People raise their pitch or power of their voice when they say something important. When the power and pitch of a voice are relatively low then it can be removed since it is difficult to hear in original recording.

The adaptive "SRC" method helps us to note the time fluctuations and speech signal. It also deletes the sections that fall below the threshold value. Fig.2 shows the principles of SRC method based on synthesis and waveform. The adaptive "SRC" was designed based on the following factors:

1.) To capture each utterance of the word correctly.
2.) The words with same power and pitch especially at the end of the sentences have less importance.

The speech and non-speech are separated based on the principles of SRC. The speech rate is changed uniformly using a linear rate function known as "linear SRC".

3.1. Adaptive SRC function

The algorithm that is used to predict the time of a breath group R(t) changed from r(s) to r(e) continuously.

$$R(t) = r(e)+(r(s)-r(e))1/2[\cos\{\pi(t-t0)/T\}+1.0] \quad (1)$$

Here r(s) should always be greater than 1.0

since it is the starting of the boosting rate. When the speech continues after T (T=2500ms) then the rate r(e) is continuously applied until the end of the speech. Due to the replaying at r(p) times the normal speed, the pitch periods are deleted or inserted in which the waveform l(n) from the starting to kth pitch period pl(k) should be as:

$$\text{Let } l(0) = 0, \quad (2)$$
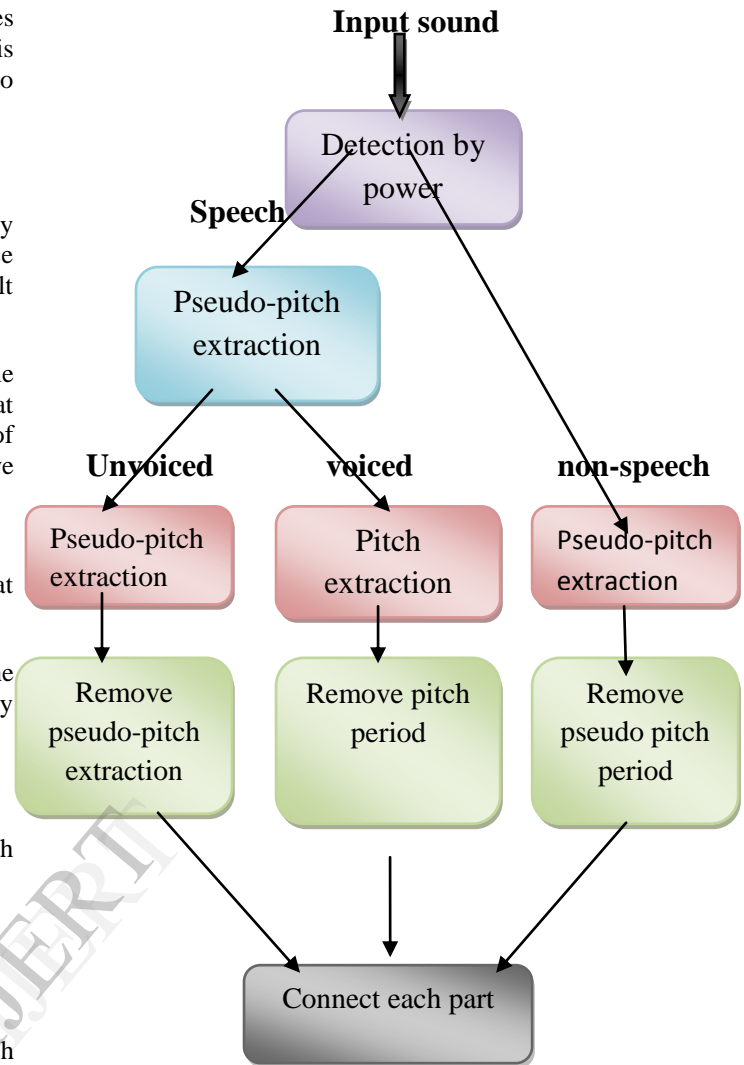
$$l(n) = 1/r(p) \sum pl(k).R(l(n-1))$$

Fig.2. SRC method based on waveform analysis and synthesis

4. EXISTING SYSTEM

The existing system does not provide a play rate which enables to retrieve the information efficiently. The fast play rate does not capture the clarity in words and becomes difficult for users to listen the speech clearly. And also there is no method that is equivalent to skimming which helps in fetching only the required part of information. This doesn't support to other languages except English. But people find this as a major drawback since they want the application to support more languages. The existing system is yet to be tested with the visually-impaired people.

5. PROPOSED SYSTEM

We have developed a new touch screen application to retrieve speech information efficiently. The proposed system has an increased speech rate to gain the information even at high play rate. We have implemented the concept of .WAV format conversion in which the selected files can be heard in easy method. This application is mainly concentrating on converting English text to speech. As a result we have

included some more languages that support in the process of text conversion.   We have developed a method of efficiently obtaining information from speech content that is equivalent to skimming printed books. The added advantage of this application is that it designed as a popular touch screen application for consumer use. The implemented speech-rate factor enables user to listen to recorded speech at even higher speeds.

6.   APPLICATION

In this application, there are three operational parameters used namely "play rate", "start rate" and "voice modulation" shown in Fig.3
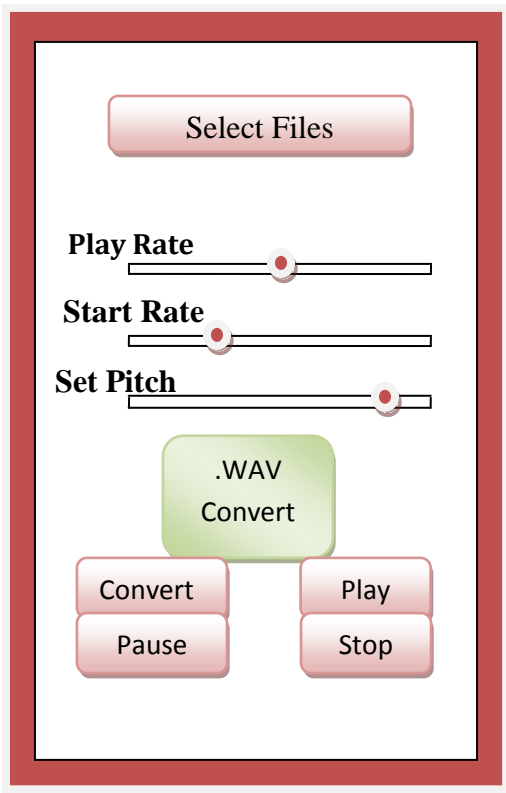

Fig. 3. Operational parameters to control the   speech rate efficiently

"Play rate" allows to set the reproduction speed and "start rate" will help to adjust the intelligibility to listen the fast speech according to user rate of interest. Once the needed text document is selected it can be listened with required play rate which allows to easy understanding of the speech. This will determine the total replay time. The "start rate" points the difference between faster and slower portions. This is developed as a touch screen application.

The"voice modulation" helps user to listen to the speech at required voice pitch rate. According to that the words are read either slowly or at a high speed. The user can also change the needed text document to a .wav file format which enables a new way of listening. Since all the features all designed as touch screen applications it will reach the audience easily and can be soon used for many multimedia applications. Thus we believe that a convenient touch screen interface will be developed for visually impaired people as early as possible.

7.   CONCLUSION

We have developed a new touch screen application to retrieve speech information efficiently using improved play rate. This report shows that our proposed system provides users to listen to speech at effective speech-rate factors and provide the technique of skimming. Also the conversion plays the speech with considering break in statements, comma and quotes. The options are provided in such a manner that it is recognized by the voice of the user. This may be an efficient method for visually impaired people. This also allows the text to speech conversion in languages other than English. The selected files to convert to speech can also be heard as .WAV format which gives an additional feature.

          Although the proposed method converts text to supported languages in the application, there is no method to convert them into Tamil. For the benefit of visually impaired people, the operational parameters can be made through the recognition of user's speech input. Much wide range of applications can also be included for the ease of use for all the users regardless of their visual capacity.

8.   REFERENCE

1.   N.Tazawa, S.Torihara, Y.Iwahana, A.Imai, N.Seiyama, and T.Takagi, "Rapid Listening of DAISY Digital Talking Books by Speech Rate Conversion Technology for People with Visual Impairments", in Proc.ICCHP(1), 2010, pp.62-68.

2.   N.Tazawa, S.Torihara, Y.Iwahana, A.Imai, N.Seiyama, and T.Takagi, "Adaptive High-speed Playback Technology and Multilingual Support for DAISY Book Listening", in Proc. CSUN, 2011, BLV 2004.

3.   M. A.Smith and T.Kanade, "Video Skimming and characterization through the combination of image and language understanding techniques" in Proceedings International Conference Computer Vision Pattern Recognition, pp.775-781, 1997.

4.   A. Imai, R.Ikezawa, N. Seiyama, A.Nakamura, T. Takagi and E.Miyasaka, "An Adaptive Speech-Rate Conversion Method for News Programs without Accumulating Time Delay", IEICE Transactions A, Vol.J83-A, No.8, pp. 935-945, Aug.2000.

5.   A.Imai, N.Seiyama, T. Takagi and E. Miyasaka, "Application of Speech Rate Conversion to Video Editing", Audio Engineering Society 20th International Conference, pp.96-101 Oct.2001.

6.   T. Takagi, N.Seiyama and E.Miyasaka, "A Method for Pitch Extraction of Speech signals using autocorrelation function through multiple window-lengths", IEICE vol. J80 No.9 pp.1341-1350 Sept.1997.

7.   T.Watanabe, "A Study on Voice settings Readers for Visually-Impaired PC Users", IEICE Transactions on Information and Systems, D-I, Vol. J88-D-I, No.8, pp.1257-1260, Aug.2005.

8.   S.Torihara, "Oblique Listening System-Speed-reading System for the Visually Impaired using Syntactic Information", Technical Report of IEICE, 5th Meeting of the Technical Committee on Well-being Information Technology, Nov.2000.

9.   T.Nishimoto, S.Sako, S.Sagayama, K.Oda, T.Watanabe, "Evaluation of text-to-speech synthesizers at fast speaking rates", Technical Report of IEICE, WIT2005-5, pp.23-28, May 2005.

10.   M.Furini, "Fast Play: A Novel Feature for Digital Consumer Video Devices", IEEE Transaction on Consumer Electronics, Vol.54, No.2,pp513-520, May2008.