

Efficient Intralingual Text To Speech Web Podcasting And Recording

Neetu Bhojar¹, Priyanka Sen², Nitu Kothe³, Shubhangee Chavan⁴, Mitali Ingle⁵

¹Dept.of Computer Science and Engineering, Dr. Babasaheb Ambedkar College of Engineering Nagpur, Maharashtra-441110, India

²Dept.of Computer Science and Engineering, Dr. Babasaheb Ambedkar College of Engineering Nagpur, Maharashtra-441110, India

³Dept.of Computer Science and Engineering, Dr. Babasaheb Ambedkar College of Engineering Nagpur, Maharashtra-441110, India

⁴Dept.of Computer Science and Engineering, Dr. Babasaheb Ambedkar College of Engineering Nagpur, Maharashtra-441110, India

⁵Lecturer, Dept.of Computer Science and Engineering, Dr. Babasaheb Ambedkar College of Engineering Nagpur, Maharashtra-441110, India

Abstract : The objective of this project is to create a web browser which will provide information in the form of audio. It is a desktop application. This web browser connects with different sites and converts text into speech. In market numbers of convertors are available which converts different file formats such as word, doc, txt, .dat into speech. This is an innovative idea to create a web browser that converts text from different file formats (word, doc, txt, .dat, PDF, rtf, html etc) to speech. This will reduce the reading effort. Different file formats can be opened in the browser which will read the text available in it. In this browser, History will be available in the form of images which will help the user to recall previously performed activities. It will also shows date and time along with the images. This will help the user to keep update with previous work done by him.

Keywords –speech, text, record, browser, audio, image history.

I. INTRODUCTION

Technology today is rapidly changing and most of these changes are value adding whether at the personal or commercial level. On a personal level technological developments have helped increase our standards of living and on the commercial level technology has helped in raising productivity and efficiency. In fact, the use of technology today has become even more significant. For example, text reader software is a good example of how modern technology can be of great assistance to us. Software that is able to convert text into speech gives us the opportunity to do tasks concurrently. The ability to multi-task can help the user of the program be more effective and efficient with his or her time which in turn will result in him/her being more productive. In effect, the software can be very effective in saving you a lot of time. In market numbers of convertors are available which converts different file formats such as word, doc, txt, .dat into speech. This is an innovative idea to create a web browser that converts text from different file formats (word, doc, txt, .dat, PDF, rtf, html etc) to speech. This will reduce the reading effort. This project is a web browser which will provide information in the form of audio.

The main modules of the project are:

1. Browser
2. Text to speech converter
3. Recorder
4. Image history

II. LITERATURE REVIEW

In 1968 the first working text-to-speech system for English was developed in the Electrotechnical Laboratory, Japan by Noriko Umeda and others. It was based on an articulatory model and included a unit for syntactic analysis with complicated heuristics. The speech was quite accurate but dull and far away from the current systems quality. Text to speech synthesis still required much work. In 1976 Kurzweil invented the first reading aid with optical scanner. The Kurzweil Reading Machines for the Blind were able to read very well even the text with multifont. However, the system was really costly for normal customers, so it was operated in libraries and service centers for visually impaired people. People were beginning to see the various applications of text to speech synthesis.

In late 1970's and early 1980's, significantly amount of commercial text-to-speech and speech synthesis products were commercially available. The first IC for speech synthesis is likely to be the Votrax chip which comprised cascade formant synthesizer and simple low-pass smoothing circuits. An inexpensive Votrax-based Type-n-Talk system was presented by Richard Gagnon in 1978. In 1980, Texas Instruments presented linear prediction coding (LPC) based Speak-n-Spell synthesizer comprising a low-cost linear prediction synthesis chip (TMS-5100). It was operated in an electronic reading aid for children and it succeeded to a significant limit. Echo low-cost diphone synthesizer was introduced by Street Electronics in 1982; it was built on a newer version of the same chip as in Speak-n-Spell (TMS-5220). During the same period Speech Plus Inc. presented the Prose-2000 text-to-speech system. First commercial models of the common DECtalk and Infovox SA-101 synthesizers were introduced a year later.

Zamzar is a web application to convert files. The name is derived from Franz Kafka's The Metamorphosis. Its main character is called Gregor Samsa and it is from his surname that Zamzar is derived. It was created by brothers Mike and Chris Whyley in England. It allows user to convert files without downloading a software tool. Users can type in a URL or upload one or more files (if they are all of the same format) from their computer, Zamzar then converts the files to another format. For example, from a Flash streaming media file to MP3. Users receive an email with a URL from where they can download the converted file. Zamzar is currently in the public beta stage of development.

Current speech synthesis technologies have very sophisticated techniques and algorithms. Hidden Markov models (HMM) are among the methods applied recently in speech synthesis. HMMs have been used with speech recognition since late 1970's; they served speech synthesis systems for more than 3 decades. The various advances over the years have made today's text to speech synthesis very common place.

All the above software is good at the features they are providing. But In this existing web based podcasters we have to give an input text explicitly. If user wants the online contents in audio format, then for searching web browser is used and for audio conversion podcaster is used. Then the idea about this project emerged and the decision to develop such software which will have these two features together is taken and that is implementing in this project.

III. APPLICATIONS

- a. With the aid of this project, user can easily change text, whether they are in word, PDF, HTML or any other file format to speech.
- b. This software can be utilized for particular tasks like pronunciation teaching for several languages. It can also be utilized with interactive educational applications.
- c. This Software is able to convert text into speech gives the opportunity to do tasks concurrently.
- d. From time to time reading too much text can be tiresome and create eye-strain. In these instances, our browser is a relief to those that regularly read texts from computers on a day-in-day-out basis.

IV. Block Diagram

4.1 Text to Speech converter:

This module will convert different file formats text into speech.

4.2 Recorder

Recording can be performed in two ways:

- 1. Direct Recording: In this method, by selecting the specified text and performing recording function. After completion of recording function the audio file will be saved and can used in future.
- 2. Indirect Recording: In this method, first the speech function will be operated and then recording will be performed.

4.3 Image History:

This module will show the history in the form of images along with date and time.

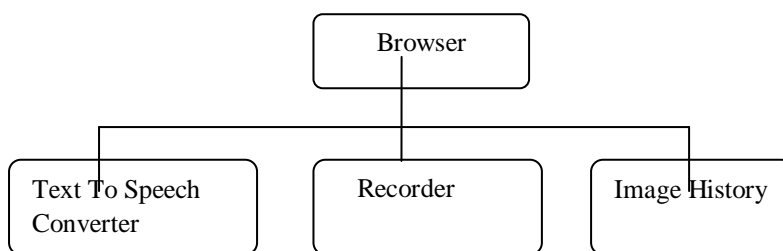


Fig 1.1 Block Diagram

V. IMPLEMENTATION

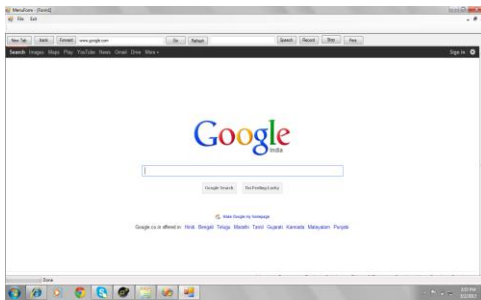


Fig 1.2 Browser

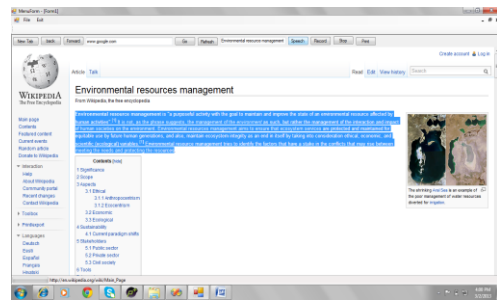


Fig. 1.3 Speech

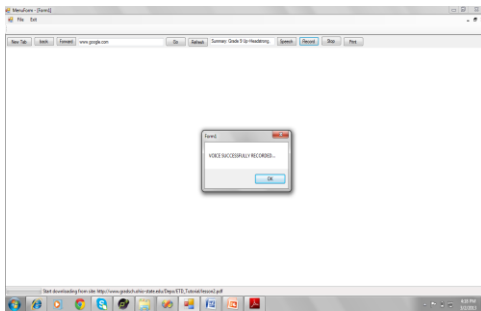


Fig. 1.4 Recording

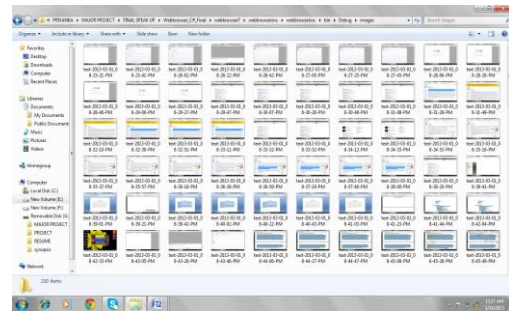


Fig. 1.5 Image History

REFERENCES

- [1] Hussein Mohsen, Karim Jahed, Marwan Fawaz “ Multi-Purpose Speech Recognition and Speech Synthesis System” IEEE Multidisciplinary Engineering Education Magazine, Vol. 6, No. 4, December 2011
- [2] O'Malley M.H, Berkeley Speech Technol., CA, USA “Text-To-Speech Conversion Technology” Vol. 3, Issue 8, August 1990
- [3] El-Imam, Y.A, IBM Kuwait Sci. Center, Safat, Kuwait Banat, K. “Text-To-Speech Conversion On A Personal Computer” Micro, IEEE, Vol. 10, Issue 4, August 1990
- [4] L. R. Rabiner, (1989). “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286.
- [5] L.M. Friesen, R.V. Shannon, D. Başkent, and X. Wang, (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *Journal of the Acoustical Society of America*, vol. 110, pp. 1150-1163.
- [6] T. Scarlatos and F. Gallarotti, “iSign: Making the Benefits of Reading Aloud Accessible to Families with Deaf Children ,” in *Proc. of the Sixth IASTED International Conference on Computer Graphics and Imaging*, Hawaii, USA, 2003.
- [7] (2005). EduSpeak [Online]. Available: <http://www.eduspeak.com/products/eduspeak.shtml>.
- [8] P. Duhamel and M. Vetterli (1990). "Fast Fourier transforms: a tutorial review and a state of the art," *Signal Processing* 19, pp. 259–299.
- [9] G. Strang, (1999). “The discrete cosine transform,” *SIAM Review*, vol. 41, no.1, pp. 135–147.
- [10] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). New York: Cambridge University Press, 2007, ch.16.
- [11] (2008). Sphinx-4, a speech recognizer written entirely in Java Programming Language [Online]. Available: <http://cmusphinx.sourceforge.net/sphinx4>.
- [12] M.S. Allen, S. Hunnicutt, and D. Klatt, *Text to Speech: The MITalk system*. Cambridge University Press, 1987.
- [13] P. Rubin, T. Baer, and P. Mermelstein, (1981). "An articulatory synthesizer perceptual research". *Journal of the Acoustical Society of America*, vol. 70, pp. 321–328.
- [14] P. H. Jan, R.W. Sproat, J. P. Olive, and J. Hirschberg, *Progress in Speech Synthesis*. New York: Springer, 1997. B.H. Juang, and L.R. Rabiner, *Encyclopedia of Language and Linguistics Elsevier*. Citeseer, 2005, pp. 1-24.
- [15] A.J. Hunt, A.W. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," in *Proc. of the IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, Georgia, 1996, vol.1, pp.373-376.
- [16] I.G. Mattingly, *Current Trends in Linguistics, Vol. 12*. Paris: Mouton, 1974, pp. 2451-2487.
- [17] P. H. S. Jan, (1994). “Assignment of segmental duration in text-to-speech synthesis,” *Computer Speech and Language*, vol. 8 pp. 95-128.