MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Creating MP3 Audio Files from Scientific Articles

Chase L. Andreason[a], Scott L. DuVall[b]

[a] *Undergraduate Research Opportunity Program, University of Utah, United States of America*
[b] *Department of Biomedical Informatics, University of Utah, United States of America*

Abstract

Reviewing scientific literature is an important part of research and education. In the University of Utah Department of Biomedical Informatics published articles are assigned as required reading. As an alternative to reading these articles, students were given the opportunity to listen to audio reproductions of the articles. To create audio files from Portable Document Format (PDF) files took several steps and an average of two hours per article. Articles were converted from PDF files to plain text, reviewed for layout and text conversion errors, reviewed for pronunciation and word usage mistakes, adjusted for pace, and saved as Moving Picture Experts Group Audio Layer-3 (MP3) audio files. This process was used for converting weekly required reading articles for a fifteen week informatics course.

Keywords:

speech synthesizer, computer-assisted instruction, education

Introduction

In the course of training in the University of Utah Department of Biomedical Informatics scientific articles are referenced and given as required reading. In a fifteen week informatics course, an option was given to students to listen to the required reading articles as an alternative to reading them. Speech synthesis was used to convert Portable Document Format (PDF) files to Moving Picture Experts Group Audio Layer-3 (MP3) files.

Speech synthesis has been studied mostly in the realm of communication aids for the disabled.[1-3] Research has shown that non-technical, fictional literature is best understood by listeners.[3] Creating audio files from published literature will allow the Department to investigate the comprehension of technical concepts in a graduate school environment.

Methods

Scientific articles were acquired from journals as PDF files using the University of Utah institutional license. A commercial text-to-speech engine was used for converting PDF files into plain text. Special formatting, figures and tables, and peripheral page text caused the conversion to include artifacts or misinterpret page flow. Each article had to be reviewed for layout and text conversion errors. As expected, many medical and technical words were not in the text-to-speech application's dictionary and were pronounced incorrectly and needed to be spelled out phonetically. In addition, the pronunciation of heteronyms like "read" and "record" were determined automatically (sometimes unsuccessfully) from surrounding words. Human review was needed to ensure correct word usage. The text-to-speech engine had difficulty reading the complex, protracted sentences found in scientific literature. To correct issues with pace, sentences were divided up in to smaller parts and pauses were inserted. Article sections were individually converted into audio files as tracks for easier navigation. The commercial text-to-speech engine included the ability to save audio in the MP3 format. Tracks for each article were bundled in a compressed folder for distribution.

Results

Fifteen scientific papers were successfully converted from PDF files to MP3 audio files using the workflow described above. Each conversion took approximately two hours and resulted in a compressed folder of audio files.

Discussion

Scientific articles given in audio format was seen as a convenience and a time-saver for graduate students. The process of creating the audio files proved time consuming. While commercial products claim to enable conversion, layout and text conversion errors, pronunciation and word usage mistakes, and pace issues occurred.

Acknowledgments

This research was supported by a training grant from the National Library of Medicine and Robert Wood Johnson Foundation and funding support from the Jenifer C. DuVall Research Foundation.

References

[1] Drager KD, Reichle JE. Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: applications for AAC. J Speech Lang Hear Res. 2001 Oct;44(5):1052-7.

[2] Levitt H. Processing of speech signals for physical and sensory disabilities. Proc Natl Acad Sci USA. 1995 Oct 24;92(22):9999-10006.
[3] Hensil J, Whittaker SG. Visual Reading Versus Auditory Reading by Sighted Persons and Persons with Low Vision. J Vis Impair Blind 2000 Dec;94(12):762-70.