SubTTS: Light-weight automatic reading of subtitles

Sandra Derbring
Peter Ljunglöf
Maria Olsson
DART: Centre for AAC and AT
Gothenburg, Sweden

Abstract

We present a simple tool that enables the computer to read subtitles of movies and TV shows aloud. The tool extracts information from subtitle files, which can be freely downloaded or extracted from a DVD, and reads the text aloud through a speech synthesizer.

The target audience is people who have trouble reading subtitles while watching a movie, for example people with visual impairments and people with reading difficulties such as dyslexia. The application will be evaluated together with user from these groups to see if this could be an accepted solution to their needs.

1 Background

1.1 Why read subtitles aloud?

Spoken subtitles could be a solution if, due to sight disorder or poor reading skills, a person is unable to read subtitles and the language spoken in the movie is unknown, or not known good enough.

A speech synthesizer able to read the text file of the DVD could make the text audible. This would make the vast sea of foreign language movies and TV shows on DVDs accessible to people with reading disabilities and visual handicaps.

Swedish Association of the Visually Impaired (Synskadades Riksförbund)[1] has around 12,000 members but there are most likely more people with poor eyesight. The number of people with reading disabilities is unknown, but according to the Swedish Dyslexia Association[2] between 5 and 8 percent of the population have significant difficulties to read and write. A survey by OECD (Organisation for Economic Co-operation and Development) in 1996 showed that "8 per cent of the adult population [in Sweden] encounters a severe literacy deficit in everyday life and at work" (OECD, 2000, p. xiii). For other countries, the problems were even bigger: "In 14 out of 20 countries, at least 15 per cent of all adults have literacy skills at only the most rudimentary level" (OECD, 2000, p. xiii).

To hear the subtitles along with the original audio track of the movie may not suit everyone, but making these movies and shows accessible could bring a huge value for people who would use it.

To reduce the risk of a large amount of auditive information disrupting the experience of watching the movie, we plan to investigate what kind of speech synthesis is best suited and what modifications can be made to the sound in the synthesis as well as in the movie.

1.2 Previous work

There have been some previous work done on automatic reading of subtitles.

A project by the Swedish Association of the Visually Impaired, in cooperation with Svenska Enter Rehabilitering AB,[3] developed a prototype that used OCR to interpret subtitles, which then were spoken aloud using TTS. The project estimated that a batch product would cost around 2500€, which they concluded would be too much for ordinary users (Eliasson, 2006, pp. 63–64).

Swedish Public Service Television (SVT) uses speaking subtitles since 2005.[4] The speech is transmitted through a second channel, which means that the user needs two digital boxes. This solution only works on SVT's own programs.

[1] http://www.srfriks.org

[2] http://dyslexiforeningen.se/om_dyslexi.html

[3] "Framtagande av TV-textläsare för syntetisk uppläsning av TV:s textremsa" (Development of a TV text reader for synthetic reading of TV subtitles)

[4] http://svt.se/svt/jsp/Crosslink.jsp?d=22138&a=274311

Kristiina Jokinen and Eckhard Bick (Eds.)
NODALIDA 2009 Conference Proceedings, pp. 272–274

Very similar to our project is Hanzlíček et al. (2008), who describe a system for reading Czech subtitles aloud. Their motivation is similar to ours, but they focus their attention on technical details on how to synchronize speech and subtitles.

1.3 Problems with existing solutions

The main problems with the existing solutions (apart from those in the Czech project described above) are that:

- they are overly complicated, by for example using OCR to scan the subtitles.
- they are closed, which means that they do not work for all kinds of movies and formats.

2 Implementation

Our prototype implementation is very simple. It consists of a script that reads a subtitle file and calls a speech synthesizer at the correct times.

To provide the speech synthesizer with input, the texting from the program needs to be extracted. There are different techniques for producing subtitles onto a TV show or a movie. Soft or closed subtitles are plain text files that are run separately from the video file, which makes them easy to edit and to extract information from. This format is often used in the subtitles available from the Internet. The files consist of the spoken lines together with time stamps that signals when the text should be displayed during playback:

```
00:00:41,549 --> 00:00:42,419
You have to go.
```

The above example means that the subtitle should be displayed 41.549 seconds into the movie and disappear at 42.419 seconds.

Subtitles are available from several sites on the Internet,[5] both in the original language and in translations into different other languages. For our purpose, the Swedish translations are of interest. In addition, subtitles are also available in purchased DVDs, often in multiple languages. Those are called prerendered subtitles and are separate video frames laid over the original streams during playback. They are usually made as an image, which makes them hard to edit. However, there is special software that can be used to extract and convert the information into soft subtitles with the help of OCR.

The present implementation is a script that extracts information from the subtitles file and uses it to provide input when communicating with the speech synthesizer. The script is currently run in parallel with the media player, but a future extension includes having it automatically synchronized.

3 Discussion

3.1 Social and pedagogical advantages

People with visually impairments and/or reading difficulties often use text-to-speech to cope with school work, and to keep up with society. Spoken subtitles further increase the accessibility of foreign movies and TV shows for these people.

Hypothetically, people with reading difficulties may also learn better how to read by using spoken subtitles. The theory is that looking at the text as it is spoken by the speech synthesis, may benefit the reading process but this is yet to be tested.

3.2 Future work

To further ease the user friendliness and the availability, the current implementation is planned to be built into a module for the open-source and cross-platform VLC Media Player.[6]

According to Hanzlíček et al. (2008), 44 percent of the Czech subtitles had overlaps when spoken with TTS. Even though we have no figures for Swedish, some overlap is to be expected also here, which is an issue that should be addressed. One possible simple solution is to modify the speech rate.

An important factor for the experience of the speech synthesizer together with a video playback would be the settings of the audio channels. Hypothetically, a listener would want to keep both the original background cues, like music, and the original voices. However, these sounds must not interfere with the speech synthesizer that is the source of information for the listener. Balancing these two criteria to get the optimized result is of great interest.

We also have plans to evaluate the application together with different users in the target groups. The aim is to discover if this approach is appreciated and if it could be an accepted solution to

[5] Two examples are `http://www.undertexter.se` and `http://www.opensubtitles.org`

[6] `http://www.videolan.org/vlc/`

the need of text interpretation during movie playback. Factors that could be evaluated and used to improve the implementation could be, for example, type of voice, type of speech synthesizer, and filter settings on the audio channels.

If the program would be used for language learning, or to help slow readers to comprehend, the feature of highlighting the word that is spoken could be a very useful additional feature.

Acknowledgements

We are greatful to three anonymous referees for their valuable comments.

References

Folke Eliasson. 2006. *IT i praktiken – slutrapport*. Hjälpmedelsinstitutet, Sweden.

Zdeněk Hanzlíček, Jindřich Matoušek, and Daniel Tihelka. 2008. Towards automatic audio track generation for Czech TV broadcasting: Initial experiments with subtitles-to-speech synthesis. In *ICSP '08, 9th International Conference on Signal Processing*, Beijing, China.

OECD. 2000. *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*. OECD Publications, Paris.

ISSN 1736-6305 Vol. 4
http://hdl.handle.net/10062/9206