

Podcast based m-Learning with Pediaphon - A Web based Text-to-Speech Interface for the free Wikipedia Encyclopedia

Andreas Bischoff
University of Hagen,
D-58084 Hagen, Germany
andreas.bischoff@fernuni-hagen.de

Abstract

This paper presents an approach to generate audio based learning material dynamically from web pages for m-Learning usage. It introduces Pediaphon (an audio interface to the free Wikipedia online encyclopedia) as an example application for microlearning. The effective generation and the deployment of the audio data to the user via podcast or progressive download (pseudo streaming) are covered.

1. Motivation

The ubiquitous availability of mobile communication devices which are connected to the Internet, makes it possible to use small amounts of spare time for mobile learning (m-learning). Travel and latency times can be used for the so called microlearning [1]. The term 'microlearning' describes a new e-learning paradigm with small or very small and short learning units. Learning material can be based on web logs (blogs) and social bookmarking.

The main reasons which limit the usage of m-Learning services for the end-user are:

- Usability problems, mainly the limited screen size and input facilities of highly mobile devices like smartphones and PDAs
- Bandwidth limitations
- Communication costs

Bandwidth limitations are solved today with UMTS enhancements like HSDPA and HSUPA (uplink). With the upcoming flatrates for UMTS and GPRS based Internet access, communication costs will be no longer an important issue.

As an alternative to display large text documents on very small displays, audio based learning material can be a solution for hand held devices. The usage of audio based learning material in distance education is state of the art since the seventies.

Nevertheless the production of audio learning material is expensive and time-consuming. As an alternative

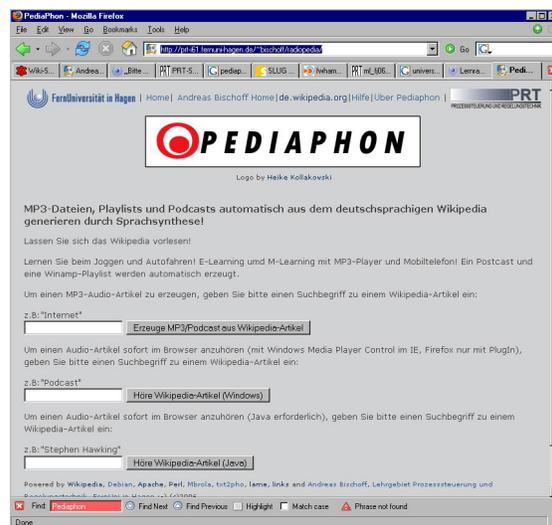


Fig.1. Pediaphon web based interface to the wikipedia
http://prt-i61.fernuni-hagen.de/~bischoff/radiopedia/index_en.html

approach, automatically *generated* audio material can



Fig.2. Pediaphon on a Window Mobile 2003 smartphone

replace time-consuming audio reproduction. Despite to the fact that the quality of text-to-speech generation is not perfect for m- and e-learning purposes, it is still usable for rapid prototyping of learning material. Especially to generate an audio representation of a text, *dynamically* text-to-speech conversion is the only solution.

The growing amount of high quality articles available via the online encyclopedia Wikipedia [2] is very suitable as *dynamic* content for microlearning purposes.



Fig.3. The Wikipedia logo

The established projects 'Spoken Wikipedia' [3] and 'gesprochene Wikipedia' (the German pendant) implement an audio representation of chosen Wikipedia articles by the help of various contributors. But these solutions are lacking the features of the text based Wikipedia articles in some points.

- *Actuality.*
The underlying principle of Wikipedia is user changeable content. The content of Wikipedia articles is changing often. Manual recorded audio representations of articles will reflect only a state of an article at a certain moment. For the users of 'Spoken Wikipedia' it is difficult/impossible to correct an article directly.
- *Completeness.*
Since the audio recording of articles is time consumptive the 'Spoken Wikipedia' project only covers 173 article of the 1.483.143 (2006-11-14, <http://en.wikipedia.org>) available articles.
- *Audio quality.*
Audio and speakers quality differ. Since 'Spoken Wikipedia' is supported by volunteers due to varying quality of their audio equipment (microphones, sound cards) general audio quality is not standardized. The speakers are non professionals and sometimes non native speakers so their pronunciation differ widely.
- *Objectivity.*
Since Wikipedia tries to establish an objective view of any topic, an emotional interpretation of an article may break this objectivity.

The only solution to the first two points, actuality and completeness, are automatically generated audio articles by text to speech techniques.

2. Pediaphon

The proposed solution the mentioned problems for 'Spoken Wikipedia' is the so called 'Pediaphon' [4]. 'Pediaphon' is a web based service which generates audio representations of Wikipedia articles dynamically. The tool is usable on- and off-line, as web based service to listen the articles directly in the web browser as well as to download MP3 files for later use in mobile devices like MP3-players and mobile phones. The realization of 'Pediaphon' combines different techniques:

- Text to speech audio generation
- on the fly MP3 compression
- Meta file generation for Winamp and Windows media player
- Podcast generation
- Pseudo streaming (progressive download)

2.1 Text-to-speech

Since text to speech audio generation is available since the eighties in differing quality for different languages it is still not a trivial task. A rule set for preprocessing of text files is necessary to cover special cases like spoken numbers, abbreviations and text formatting. Spoken language consists of a set of phonemes and the generation of these phonemes out of text files vary largely on this between different languages. Especially if the pronunciation of a word is depending on its meaning the phoneme generation will fail.

A free digitally available pronunciation encyclopedia for the target language will simplify this task. The open source tool 'txt2pho' [5] is used to get a text based representation of the phonemes for a given German language text file. The English language variant of 'Pediaphon' takes advantage of a similar tool for British English (freephone) [6].

After identifying all phonemes each single phoneme must be synthesized as digital audio output. The free 'mbrola' [7] speech synthesizer is used for this task. 'Mbrola' is a universal solution for voice files from different languages.

2.2 MP3 compression

Since the uncompressed audio representation of a typical Wikipedia article is about 80 MB of data, a suitable compression is needed. To avoid the transfer of a huge amount of data efficient MP3 audio compression is used. Especially for later progressive download it is useful to match a target data rate of 32 kBit/s which fits modem

users requirements for real time playback, too. A typical average sized Wikipedia article can be reduced to an amount of 5 MB. The open source MP3 encoder 'lame' [8] performs this task for the 'Pediaphon' application. Suitable MP3 ID3-tags (MP3 meta data, author, name of Wikipedia article) will be generated on the fly by Pediaphon too.

2.3 Metafile generation

Today's web browsers are unable to play-back MP3 audio directly. Suitable web browser plugins are needed embedding an audio player into the web page. Three different methods of embedded plugins are implemented into the 'Pediaphon' web service.

- Window media player plugin. An ActiveX control for the windows media player is available.
- Java based MP3 player. The open source Javazoom Tinyplayer [9] provides a MP3 playback feature. Java itself was released as open software too, since 2006-11-13 [10].
- A open source Macromedia Flash-based MP3 player (emff) [11].

The 'Pediaphon' application dynamically generates a html page with suitable plugin parameters and metafiles for these players. A link to the origin Wikipedia article will be included dynamically.

2.4 Progressive download

To avoid a delay for the download of a huge file before the playback starts, streaming techniques will be used often. Streaming is defined as UDP based transmission of IP packets. Delays due to retransmissions of IP packets will be avoided by the usage of streaming. To realize a 'pseudo streaming' functionality, the so called progressive download method was chosen. The web browser downloads just a metafile with a description of the audio media (URL of the MP3 file, transfer method). Then the web browser audio plugin or an external player will fetch the MP3 file and will start the playback during download. One of the features of the MP3 file format is that playback is possible even with a partially downloaded



file. If the download rate is larger than the data rate of the MP3 (here 32kBit/s) a real time playback is possible. To playback an uncompressed audio stream, a up to ten times larger bandwidth is necessary.

Fig.4. The unofficial Podcast logo

2.5 Podcast

The word 'Podcast' was derived from iPod and broadcast. Since the word 'Podcast' is the word of the year 2005 (elected by the New Oxford American Dictionary) Podcasts and Podcasting are buzzwords today.

In a technical sense a podcast is nothing more than a MP3 file and a XML based RSS-Newsfeed (Really Simple Syndication) meta description. But for users, with no or little knowledge about file formats and meta data, a Podcast is a very convenient way to synchronize their audio data to their mobile devices or to create an own Podcast. Also for producers of a Podcast very convenient tools like 'GarageBand' and 'PodcastMaker' are available. The very successful Apple iTunes software, (available for MAC OS and Windows) and Apples iPod MP3 player series have created a high demand for audio based teaching material. Due to the success of Podcasting a renaissance of Audio based teaching occurs. Especially for mobile learning or microlearning purposes Podcasts are very successful.

The Pediaphon web service generates a valid feed for the Podcast XML description on the fly for every request.

3. Conclusion and future work

Since its creation in February 2006 the Pediaphon service has answered to 11400 request and has generated MP3 files (English and German language Pediaphon) out of Wikipedia articles. With an average size of 12 MB of a MP3 file this is a total amount of 145 GB generated MP3 data. The equivalent playing time is about 12500 hours or 523 days.

The control systems engineering group of the University of Hagen uses the same technique to generate audio teaching material (a Podcast) out of text based material.

Future work on 'Pediaphon' will cover the implementation of new target languages like French and Spanish and distribution a of xhtml and WAP based solution for mobile phones.

An evaluation of the acceptance of text to speech based teaching material will take place during a online seminar in spring 2007.

4. References

[1] Roger Fischer, Microlearning with Mobile Weblogs, Microlearning Conference 2005, http://microlearning.org/micropapers/MLproc_2005_fischer.pdf

[2] N.N., http://en.wikipedia.org/wiki/Main_Page , 2006

[3] N.N., http://en.wikipedia.org/wiki/Wikipedia:Spoken_Wikipedia , 2006

[4] Andreas Bischoff, http://prt-i61.fernuni-hagen.de/~bischoff/radiopedia/index_en.html , 2006

[5] Portele, T., Steffan, B., Preuss, R., Sendlmeier, W.F., Hess, W. (1992): "HADIFIX - a speech synthesis system for German", Proceedings of the International Conference on Spoken Language Processing (Banff, Alberta), vol. 2, pp. 1227-1230.

[6] N.N. <http://tcts.fpms.ac.be/synthesis/mbrola.html> , 2006

[7] Dutoit, T., Pagel, V., Pierret, N., Bataille, F. and van der Vreken, O. (1996) The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes, *Proc. ICSLP'96, Philadelphia, vol. 3, pp. 1393-1396*

[8] N.N., <http://lame.sourceforge.net/index.php>, 2006

[9] N.N., <http://www.javazoom.net/applets/tinyplayer/tinyplayer.html> , 2006

Java itself was released as open software too since 13.11.2006

[10] N.N., <http://www.sun.com/software/opensource/java/faq.jsp> , 2006

[11] Marc Reichelt, <http://www.marcreichelt.de/spezial/musicplayer/> , 2006