CONTEXT-AWARE MOBILE LEARNING WITH PEDIAPHON – A TEXT-TO-SPEECH INTERFACE TO THE FREE WIKIPEDIA ENCYCLOPEDIA FOR CELL PHONES AND MP3-PLAYERS

Andreas Bischoff

University of Hagen
Department of Electrical Engineering
Control Systems Engineering group (PRT)
Philipp-Reis-Building
Universitätsstraße 27
D-58097 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 and ubiquitous access. 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. A convenient cell phone interface to the Wikipedia content will be introduced.

KEYWORDS

Text-to-speech, Wikipedia, VoIP, mobile learning, Podcast, cell phone, Pediaphon

1. INTRODUCTION

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. Communication costs and bandwidth limitations are also limiting factors for possible users of mobile learning services. Bandwidth limitations are solved today (or will be solved in the near future) with UMTS enhancements like HSDPA and HSUPA (uplink). With the upcoming flat fees for UMTS and GPRS based Internet access, communication costs will be no longer an important issue. Nevertheless many users of mobile phones do not use mobile Internet access because they are afraid of the costs. 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. The production of audio learning material is expensive and time-consuming. As an alternative approach, automatically *generated* audio material can 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. 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 428 article of the 1,793,981 (2007-05-21, 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. Objectivity in the *content* of articles is often a topic of heavy discussion.

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 (Fig. 1). 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.

2.1 Text-to-speech

As a first step a web based Wikipedia article has to be preprocessed to split navigation links and the table of content from the pure article text. 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]. The just implemented French version of the Pediaphon service uses the DIC_PHON [14] text to phoneme translator. 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 Compression and distribution

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 or GPRS 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.

Todays 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 to 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. 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.

2.3 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 Podc7asting a renaissance of Audio based teaching occurs. Especially for mobile learning or microlearning purposes Podcasts are very successful. Like MP3 playback Pocasts are today's frequently used mobile media. The Pediaphon web service generates a valid feed for the Postcast XML description on the fly for every request.

3. Mobile Phone WAP interface

For mobile phone access we have established an WAP interface to the Pediaphon [4] service, but the communication costs are still a limiting factor for mobile users. The WAP interface to the Pediaphon service provides a WML 1.1 input mask as well as a WML server response. The response includes a link to the corresponding article at the Wapedia WAP service (a WAP representation of all Wikipedia articles) as well as a link to the generated MP3 audio file. Downloading the usually huge MP3 files can be very expensive for users without a GPRS or UMTS flat fee contract. To type in type URLs on mobile phones is usually painful.

Common approaches to assist the users for typing, like the T9 method, are not useful for URLs. Especially long URLs with special characters require much effort for potential users. To realize a convenient way to distribute URLs to a cell phone, 2D markers (barcodes) were used for the Pediaphon WAP URL (Fig.3). Software, which reads these 2D-barcodes is available for nearly all mobile phones with J2ME support. These kind of 2D barcodes are also used for the very successful Semapedia [12] project, which connects (links) physical existing objects to their representation in the Wikipedia encyclopedia.

4. Mobile Phone SMS/GSM interface

The main disadvantage of the WAP solution is the large amount of data which will be generated by each Pediaphon query. A MP3 file generated from a typical Wikipedia article will be about 12 MB of data size. For mobile internet user without a flat fee the download may be very costly. The downloaded file has to be stored locally on the device. Some mobile phones have very limited resources to store files. To avoid those communication costs and the limitation to users of the mobile internet we have created a pure SMS and GSM based service, too. No (probably expensive) GPRS or UMTS mobile Internet access is required. The service works with all mobile phones today without the need of additional configuration or expert knowledge. The costs are transparent for the users, only communications fees for an SMS request and a land line call have to be paid. Since the users are paying directly to the communications service provider, the Pediaphon service can be established without any costs for the university. The very successful Pediaphon GSM interface was realized by usage of voice over IP technology (VoIP) and the Asterisk [13] open source PBX (private branch exchange). The German language Pediaphon SMS interface could be accessed by sending a SMS with the text: 'pedia [search word]' to +49-171-5650513. After two minutes (to be sure that the SMS arrives and the text to speech (TTS) processing is done) the users have to call (with the same caller ID) the land line number +49231-5173852 (International, Germany). The announcement of the generated speech will start automatically. The communication costs are transparent to the users. This service is available for the English language version too. The SMS request for the English language version must include the token 'en': 'pedia en [search word]'.

5. USAGE AND USERS COMMENTS

Since its creation in February 2006 the web based Pediaphon service has answered to over 24000 request and has generated MP3 files (English, German and Frensh language Pediaphon) out of Wikipedia articles. With an average size of 12 MB of a MP3 file this is a total amount of 288 GB generated MP3 data. The equivalent playing time is about 25000 hours or 1042 days.

Only a small amount of requests (less than 5%) utilize the mobile phone interface. The main reason are the communication costs. User comments are rare, we received less than 100 comments since the creation of the service. Most of the comments cover audio rendering problems of single articles. Depending on the relative free formatting policies the structure of article vary heavily which causes failures in some articles. About 10 % of user feedback are feature requests. The English language version of the Pediaphon service is a direct outcome of such a feature request. The service itself is quite usable for short Wikipedia articles. It has been reported by some users that listening to a synthetic voice for a long time is inconvenient. The Pediaphon service is more usable with short articles just to refresh knowledge or to recover facts and issues. Especially the mobile phone version can be useful to act like a 'Hitchhiker's Guide to the Galaxy' [15] for mobile users.

6. FUTURE WORK

The control systems engineering group of the University of Hagen uses the same technique to generate audio teaching material (and a Podcast) out of text based material. Future work on 'Pediaphon' will cover the implementation of new target languages like Spanish and Russian and will improve the mobile phone based user interface.

REFERENCES

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

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

[3] N.N., 2007, *Spoken Wikipedia Project*
http://en.wikipedia.org/wiki/Wikipedia:Spoken_Wikipedia

[4] Andreas Bischoff, 2007, *Pediaphon*
http://prt-i61.fernuni-hagen.de/~bischoff/radiopedia/index_en.html

[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., *Mbrola*, 2007, http://tcts.fpms.ac.be/synthesis/mbrola.html

[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., 2007, *lame*, http://lame.sourceforge.net/index.php

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

[10] N.N., 2006, *Java released as open source*, http://www.sun.com/software/opensource/java/faq.jsp

[11] Marc Reichelt, 2007, *musicplayer*,
http://www.marcreichelt.de/spezial/musicplayer/

[12] NN, Semapedia, 2007, http://www.semapedia.org/

[13] N.N, 2007, *Asterisk PBX*, http://www.asterisk.org

[14] Bechet F., 2001, *LIA_PHON - Un systeme complet de phonetisation de textes*, revue Traitement Automatique des Langues (T.A.L.), volume 42, numero 1/2001, edition Hermes

[15] Douglas Adams, 1979, *The Hitchhiker's Guide to the Galaxy,*