DIGITAL RADIO AS AN ADAPTIVE SEARCH ENGINE
Verbal Communication with a Digital Audio Broadcasting Receiver

Günther Schatter, Andreas Eiselt and Benjamin Zeller
Faculty of Media, Bauhaus University Weimar, Bauhausstraße 11, D-99421 Weimar, Germany

Keywords: Digital Audio Broadcasting, Digital Radio, Text-to-Speech, Automatic Speech Recognition, Speech Dialog System, Information Retrieval System.

Abstract: This contribution presents the fundamentals for the design and realization of an improved digital broadcast receiver. The concept is characterized on the one hand by a content repository, which automatically stores spoken content elements. On the other hand a speech-based and a graphical interface are implemented, which enable users to directly search for audio and data content. For that purpose a DAB/DMB receiver is augmented to simultaneously monitor a variety of information sources. Other digital communication methods such as Satellite and HD Radios^{TM} or DVB are applicable as well. Besides broadcast receivers, the system can be integrated into next-generation mobile devices, such as cell phones, PDAs and sophisticated car radios. Priority is given to the idea of establishing a compact embedded solution for a new type of receiver, promoting the convergence of service offers familiar to users from the internet into the broadcast environment.

1 MOTIVATION

In recent years, the relevance of the classic medium radio is significantly decreasing. If information is required, it is more probable to use a web-based search engine than a radio. However, the radio is per se the first mobile and ubiquitous electronic medium with many proved benefits such as ubiquitous, mobile, topical, and free of charge reception.

If access to the Internet is not available, for instance in sparsely populated or less developed regions, a radio service usually exists. Radio encompasses everyone. The range of services offered on radio networks varies from entertainment to voice oriented information channels and to complex multimedia applications. The latter was possible by the development of digital radio concepts as Digital Audio resp. Multimedia Broadcasting (DAB/DMB), Digital Radio Mondiale (DRM), HD Radio^{TM}, satellite-based systems as 1worldspace^{TM}, Sirius XM Radio^{TM}, and Digital Video Broadcasting (DVB) as well. The level of standardization for several broadcasting one-way concepts is quite advanced. Even though digital broadcasting offers many potential advantages, its introduction has been hindered by a lack of international arrangements on common standards. However, the convergence of information technology and consumer electronics has also opened up a wide range of opportunities for a new generation of Digital Radio as a gateway to newly emerging services (Herrmann, 2007). Digital Radio is still in a growing state of penetration worldwide and has yet a huge development potential for numerous and still unknown applications.

Our contribution will present an automatic system that monitors several audio and data services of digital radio and allows selective retrieval of news and other information based on spoken and textual queries. The user may choose among the retrieved information and play back the stories of interest. The presented solution ist somewhat different from previous approaches in that the repository is expanded on data services and the user need not exclusively express a textual query. The results are applicable also for other digital broadcasting systems than DAB/DMB.

2 PREHISTORY

Information retrieval on spoken radio documents has been investigated since the 1990s because first Automatic Speech Recognition (ASR) systems were

Schatter G., Eiselt A. and Zeller B. (2009).
DIGITAL RADIO AS AN ADAPTIVE SEARCH ENGINE - Verbal Communication with a Digital Audio Broadcasting Receiver.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 157-164
DOI: 10.5220/0002234401570164

fed with recorded radio news, due to the fact that the news speakers were trained speakers, the reports were spoken clearly and were well-pronounced without background noise (Glavitsch, 1992). Preliminary investigations into the use of speech recognition for analysis of a news story was carried out by (Schäuble, 1995). Their approach used a phonetic engine that transformed the spoken content of the news stories into a phoneme string.

Later Whittaker et al. evaluated interfaces to support navigation within speech documents providing a mixed solution for global and local navigation by a graphical interface in combination with an audio output as well (Whittaker, 1999). Emnett and Schmandt developed a system that searches for story boundaries in news broadcasts, and enables a nonlinear navigation in the audible content with the help of different interface solutions (Emnett, 2000).

The problems of mobile devices in rough environments in relation with a speech-based interaction are reported on in (Sawhney, 2000). A notification model dynamically selects the relevant presentation level for incoming messages (such as email, voice mails, news broadcasts) based on priority and user activity. Users can browse these messages using speech recognition and tactile input on a wearable audio computing platform.

There are prototypes for web-based Speech Dialog Systems (SDS) and audio search engines that are able to search spoken words of podcasts and radio signals for queries entered with a keyboard (ComVision, 2009) (TVEyes, 2009). In this context, the European Telecommunications Standards Institute has developed a standard for DAB/DMB-based voice applications. The VoiceXML profile will in the future explain the dialog constructs, the user input/output, the control flow, the environment and resources.

3 FOUNDATION

Several basics are necessary for the development of the system: An analysis of the different data transmission techniques, a survey of the available information sources in the Digital Radio environment, methods for the Music-Speech Discrimination (MSD) resp. speech extraction, and procedures for the Information Retrieval (IR) and the output of spoken information as well.

3.1 Information Sources

Two general types of information sources are distinguished in a digital broadcast receiver:

(1) The audible signals (internet audio and podcast files are applicable as well) are converted by an ASR system into plain text in order to perform a content-based analysis. The set of recorded information include such as breaking news, headlines, educational and cultural items, current affairs discussions etc.

(2) The data services are available as text (internet-based information is applicable too). Broadcast Websites (BWS) contain multifaceted news, press reviews etc. Other sources of information are Electronic Program Guides (EPG), Dynamic Label Plus, Intellitext, and Journaline (ETSI 2005).

Compared with (1) these information sources are more reliable with respect to structure and content, but less detailed and not always available. It was indispensable to establish a hierarchical sequence of sources depending on reliability, quality, and convenience.

3.2 Information Processing

For most digital receivers the Music/Speech Flag (M/S) is defined to distinguish music from speech, but broadcasters do not consistently broadcast this information. Another way to discriminate music from speech is to analyze the signal for typical patterns. Features which are abstracting this information were analyzed focusing on the requirements of MSD in digital broadcasting.

An ASR system is capable of transforming speech signals into machine-readable text. A Text to Speech System (TTS) synthetically generates a speech signal. There are nearly two dozen proprietary as well as open source systems for ASR and TTS systems on the market, differentiating in terms of speed, accuracy, speaker dependence, and robustness. A good overview of current research topics about the advanced development of SDS is given in (Minker, 2006).

To reduce the text generated by the ASR and make it searchable, a stemming algorithm is used. It reduces inflected or derived words to their stem. Although the derived stem of a word is not always the morphological root of the word, it is sufficient to be able to map different forms of a word to the same stem. The algorithm tries to reduce inflected or derived words to their stem. Another useful technique to reduce the amount of text is called stop word removal, which removes words with no semantic meaning. After words have been stemmed and the common ones removed, the system converts each text to a term vector if necessary using well-known log-entropy weighting schemes (Porter, 1980).

3.3 Metadata for Audio Content

The functionality of IR systems is highly dependent on the usage of appropriate metadata. The internal structure of all standards allows describing content elements in a hierarchical structure with any depth and level of granularity according to numerous descriptors. Even though music may be labeled with ID3-tags by internet radios, there is no wide-spread method to retrieve spoken content by tagging. As an alternative it is possible to apply subsets of the TV-Anytime standard (ETSI, 2004) offering XML-structures feasible in the field of search, recommender and archive applications (Schatter, 2006, 2007). The contained keywords resemble the radio-specific extensions of ID3v2 mainly concerning the music-related domain. An early proposal for a technique of machine-interpretable DAB content annotation and receiver hardware control, involving the utilization of Dynamic Label (DL) fields in the transmitted frames, was formulated by Nathan et al. in (Nathan, 2004). A similar approach was chosen in (Schatter, 2007). The separation of information is carried out in both cases by machine-readable control characters.

This specific metadata is commonly encoded in an XML-based data structure. There are a few standards enabling the description of AV-content such as EPG, TV-Anytime and MPEG-7. The defined catchwords and descriptors respectively derive from a controlled vocabulary that is capacious but of fixed extent. The TV-Anytime standard embraces a comprehensive collection of classification schemes. These schemes consist of elaborate arrangements containing pre-assigned terms that are used as catchwords to attach several categories to AV-contents. However, the descriptions appear not to be chosen systematically in every respect.

4 DESIGN OF THE SYSTEM

There are two fundamental designs proposed for the system:

(1) A user-based model: The navigation through the broadcast audio content is possible even without additional metadata from the broadcaster. In this case, all processes of IR have to be done on the receiver, see Figure 1, section B + C.

(2) A provider-based model: All IR processes will be done by the broadcaster who submits the gathered information to the DAB receivers using an appropriate meta description, see Figure 1, section A + B.

To match the fundamental premises of an independent and ubiquitous system, two additional features were proposed with the focus on improved usability in a mobile environment:

(1) The entire controllability of the radio device by verbalized user queries,

(2) A memory function allowing not only to search in the current broadcast content but in stored content as well.

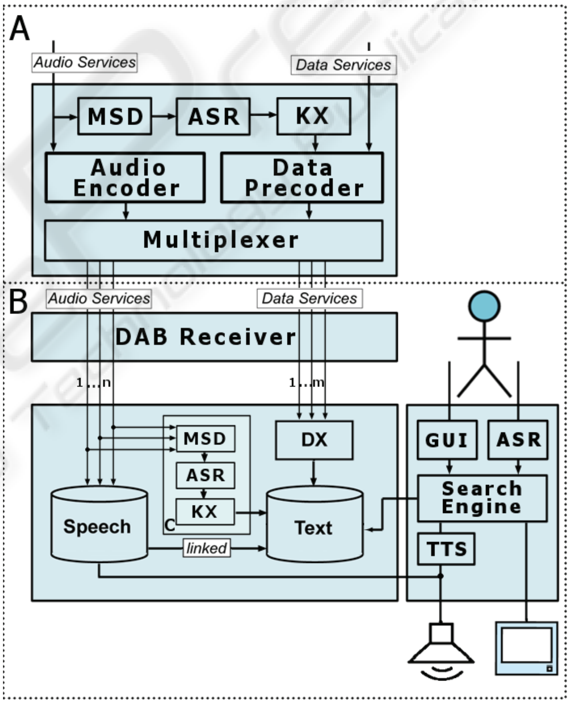


Figure 1: System overview of provider- and user-based model.

4.1 Comprehensive Structure

4.1.1 User-based Model

The concept of the system is realized on the mobile device itself, which consists of three main parts: the monitoring of the audio services by MSD, the ASR, and the keyword and metadata extraction. The MSD and the keyword extraction can be done simultaneously and in real time for more than one monitored service, whereas the ASR is much more time-consuming. There are three principal system architectures for ASR on mobile devices proposed in (Zaykovskiy, 2006), which take account of this aspect. Because two of these concepts are involving additional mobile devices whose availability could not be guaranteed in our case, the third, an embedded approach, is recommended.

4.1.2 Provider-based Model

In this design, the duties are arranged to avoid the restrictions brought about by limited resources on mobile devices. The process of IR was centralized and sourced out to the broadcaster, who adds the gathered information to the data services. Due to these changes and because the broadcaster is able to use high-capacity hardware the quality of the retrieved data may increase as well. The disadvantage of this model compared to the first, is the dependency on the broadcast information. Our experience is that radio stations for reasons of costs do not consistently broadcast this information.

4.2 Information Retrieval and Management

4.2.1 Central processes

The process of IR is divided into three sub-processes which are successively executed: the MSD, the ASR and the Keyword Extraction, see Figure 2.

The MSD was complicated by common arrangements in radio such as background music and cross-fadings. To solve the problem, a new audio feature was proposed which is more insensitive to such audible structures. The feature based on the fact that speech is recorded by one microphone, whereas music is recorded by at least two microphones. Due to that, a speech signal has no significant phase differences between the audio channels.

In order to retrieve information from the broadcast audio, the speech-based content is processed by an ASR system converting the audio data into plain text. To make the data searchable, the extraction of semantic information is processed by a stop word removal and stemming algorithm. In order to enrich the searchable data, information about the current program is extracted from EPG, BWS, DL and if available from appropriate sources in the internet. Those data from different sources and services are complementing the information pool on the hard disc storage.

4.2.2 Strategy to manage File Space in Case of Low Capacities

Of course the capacity of mobile devices storing content locally is limited. Therefore a strategy was implemented that is deleting those content elements that are probably not relevant for the user. For that purpose the personal preferences of users are stored separately based on the work in (ETSI, 2005). Here the preferences of users are centrally defined in a standardized format. Each element in the user's preferences defines a preference value ranging between -100 and 100 for a certain content type or element. Depending on the value, the content element is persisted for a longer or a shorter time.

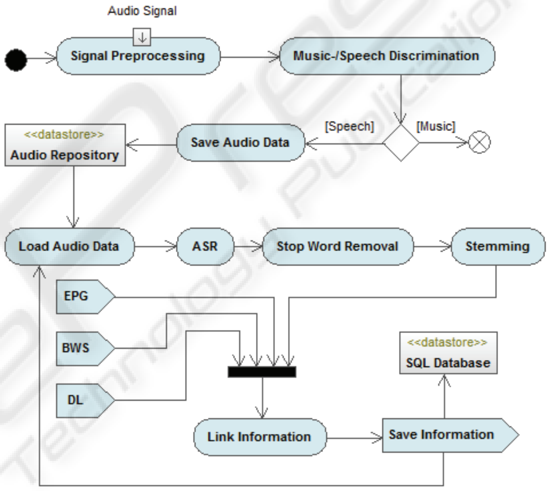


Figure 2: Activity diagram of the IR process.

4.2.3 Automatic Generation of Metadata

The data which were extracted during the preceding IR process are subsequently subject to the automatic generation of metadata. They are transmitted in parallel to the audio services. There are three entities comprising this process:

(1) The extracted data conforming to a proprietary structure,

(2) The standardized metadata to transmit (EPG, TV-Anytime, MPEG-7), and

(3) The converter that is automatically generating the metadata on the basis of the extracted data and the metadata structure to transmit.

4.3 Interfaces

The system incorporates a multimodal user-interface, which enables the user to interact in two ways:

- by a Speech-based User Interface (SUI)
- by a Graphical User Interface (GUI)

The structure of the system is based on the client-server model, see Figure 3.

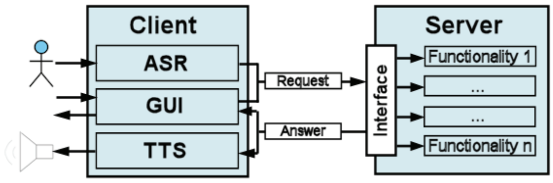


Figure 3: Structure of the multimodal interface.

At the one hand the client incorporates all functionalities related to ASR, speech synthesis and The GUI and at the other hand the server enables the client to access the entire functional scope of the radio.

The choice of voice commands for the SUI was based on the following requirements:

- Memorability: Commands had to be easily memorable in order to enable the user to reliably utilize all commands.
- Conciseness: Users should be able to easily associate commands to functions.
- Briefness: The length of a command was kept at a target size of 1-2 syllables.
- Unambiguousness: The use of homonyms was strongly avoided.
- Tolerance: The use of synonyms was desirable to a high degree.

5 IMPLEMENTATION

Our solution was realized on the basis of a DAB receiver (DR Box 1, Terratec) connected by USB. The System has been implemented in Java JDK 6/MySQL 5.1 on a standard laptop (Core2Duo; 1,8 GHz; 4GB DDR3-RAM; 500 GB HDD) and operates quite mobile. The system was exemplarily implemented based on the user-based model pursuing the two design ideas:

The identification of speech-based content and the possibility for users to directly search for specific audio content was implemented with a graphical user interface.

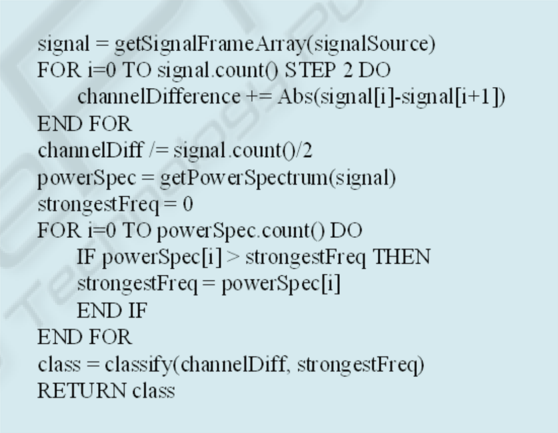
The accumulation of text-based data services in combination with the capability to search for desired information was realized with a speech-based user interface.The system consists of two main parts. The first is the monitor itself comprising the three main processes of the information retrieval subsystem: MSD, ASR and keyword. The second part is a graphical user interface allowing the user to directly search for content and access the audio files related to the results found.

It is important to note that both interfaces could be utilized for either case.

5.1 Information Retrieval and Management

The first main process records and processes the incoming audio data by MSD. The MSD is accomplished by at first decompressing the incoming MPEG signal. The raw PCM format is the processed by two audio feature extractors calculating the channel difference and the strongest frequency, to classify the current content, see Table 1.

Table 1: Music-Speech Discrimination in pseudo code.



```
signal = getSignalFrameArray(signalSource)
FOR i=0 TO signal.count() STEP 2 DO
    channelDifference += Abs(signal[i]-signal[i+1])
END FOR
channelDiff /= signal.count()/2
powerSpec = getPowerSpectrum(signal)
strongestFreq = 0
FOR i=0 TO powerSpec.count() DO
    IF powerSpec[i] > strongestFreq THEN
    strongestFreq = powerSpec[i]
    END IF
END FOR
class = classify(channelDiff, strongestFreq)
RETURN class
```

All audio data is recorded to a repository with separate folders for each digital broadcast service and child folders for music and speech. In order to persist content elements in the original sequence the according audio files are labeled with the time stamp of their starting time.

The second main process monitors the audio repository in parallel, see Table 2. This process permanently retrieves new files in the repository and converts the audio data into plain text (ASR). The prototype uses a commercially available, large-vocabulary ASR engine that was designed for speech-to-text dictation. Those dictation recognizers do not perform well at information retrieval tasks, but they can operate as speaker-independent systems.

After the extraction the text is processed by a stop word removal and the stemming algorithm. The resulting set of words is written to a database including obligatory information about the associated audio file and contextual metadata. Additional metadata are included by parsing the BWSs, EPGs, DLs, and appropriate internet services.

Table 2: Automatic Speech Recognition in pseudo code.

```
WHILE monitorIsActive
   files = getFilesNotProcessedInAudioRepository()
   FOREACH files AS filename DO
        text = asr(filename)
        procText = keywordExtraction(text)
        service = getServicenameFromFilePath(filename)
        timestamp = getTimestamp(filename)
        meta = getRelatedMeta(service,timestamp)
        saveToDB(sender,timestamp,procText,meta)
   END FOREACH
END WHILE
```

In case of the provider-based model it is necessary to map the proprietary data to standardized XML-based metadata. This requires a converter for each pair of proprietary data structures and possible metadata standards to use for broadcasting. This aspect was realized exemplarily by a converter mapping the data structure to the EPG metadata standard, see Figure 4. The converter automatically maps the information from the extracted data to the standardized descriptors of the metadata structure.

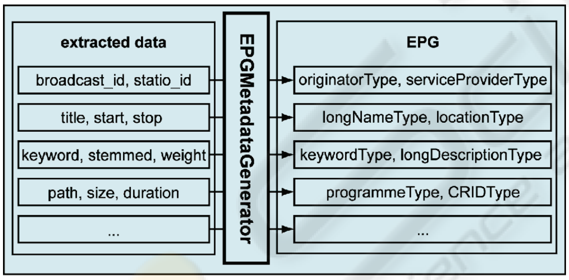
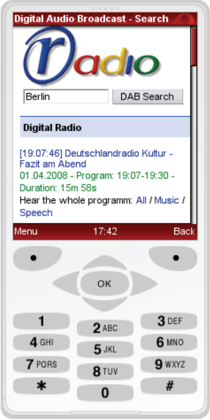


Figure 4: Schematic mapping of proprietary data structure to standardized metadata.

5.2 User Interface

The second part of the system is the User Interface (SUI/GUI). The user is able to specify a query by voice or over via a web interface. Subsequently the system parses/interprets the user's input and searches for corresponding data in the database. The results are listed in the GUI as shown in Figure 6. The user is able to select a content element similar to a web browser and to listen to the associated audio file. Although our prototype utilizes underlying textual representation and employs text-based information retrieval techniques, this mechanism is hidden to a great extent from the user.



[19:07:46] Deutschlandradio Kultur - Fazit am Abend
01.04.2008 - Program: 19:07-19:30 - Duration: 15m 58s
Hear the whole programm: All / Music / Speech

Figure 5: Design of a GUI on mobile devices for searching audio content.

Over the GUI the user is able to decide for each result if he wants to hear only the piece where the keyword occurred, the whole program in which this piece occurred or only the speech/music of this program. A possible result-set is exemplified in Table 3. How this result-set could be presented to the user by the GUI is shown in Figure 5.

6 USE CASE AND EXPERIENCES

The user is able to search the indexed content through verbal queries or accepts textual queries via a web-based interface with a query window, see Figure 5. In case of a spoken request, the system utilizes a speech recognition engine to retrieve a searchable string, which in the case of the web-based interface is entered by the user directly. The results could be presented either as spoken response utilizing TTS technologies, see Figure 6 or through a text-interface, see Table 3.

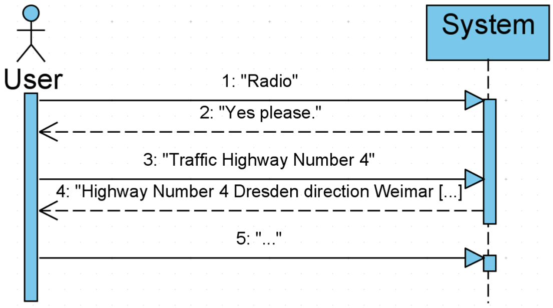
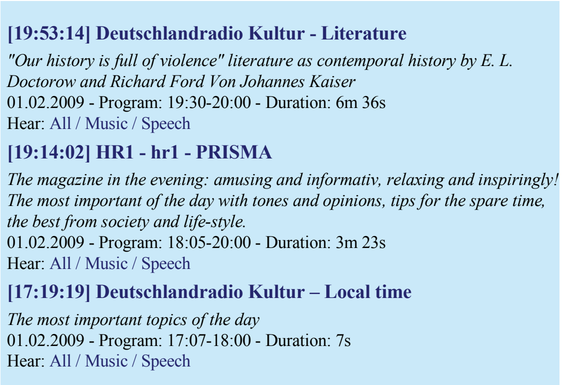

User
System
1: "Radio"
2: "Yes please."
3: "Traffic Highway Number 4"
4: "Highway Number 4 Dresden direction Weimar [...]
5: "..."

Figure 6: Example spoken dialog about traffic.

The results are furthermore ordered by their relevance represented by the count of occurrences of the keywords and the time when the related audio content was broadcast. For each result the user can decide to hear only the piece where the keyword occurred, the whole program in which this piece occurred or only the speech/music of this program.

Table 3: Excerpt of results for search query "*Literature*".



[19:53:14] Deutschlandradio Kultur - Literature
"Our history is full of violence" literature as contemporal history by E. L. Doctorow and Richard Ford Von Johannes Kaiser
01.02.2009 - Program: 19:30-20:00 - Duration: 6m 36s
Hear: All / Music / Speech

[19:14:02] HR1 - hr1 - PRISMA
The magazine in the evening: amusing and informativ, relaxing and inspiringly! The most important of the day with tones and opinions, tips for the spare time, the best from society and life-style.
01.02.2009 - Program: 18:05-20:00 - Duration: 3m 23s
Hear: All / Music / Speech

[17:19:19] Deutschlandradio Kultur – Local time
The most important topics of the day
01.02.2009 - Program: 17:07-18:00 - Duration: 7s
Hear: All / Music / Speech

Through relatively recent improvements in large-vocabulary ASR systems, recognition of broadcast news has become possible in real-time. Though, problems such as the use of abbreviations, elements of foreign languages, and acoustic interferences are complicating the recognition process. The combination of informal speech (including dialect and slang, non-speech utterances, music, noise, and environmental sounds), frequent speaker changes, and the impossibility to train the ASR system on individual speakers results in poor transcription performance on broadcast news. The result is a stream of words with fragmented units of meaning.

We confirmed with our experiments an older study of ASR performance on broadcast news of Whittaker to this day (Whittaker, 1999), who observed wide variations from a maximum of 88% words correctly recognized to a minimum of 35%, with a mean of 67% (our results: 92%, 41%, 72%). Unfortunately most ASR programs do not show additional information; they do not offer any measure of confidence, nor do they give any indication if it fails to recognize anything of the audio signal. When the speech recognizer makes errors, they are gaps and deletions, insertions and substitutions of the inherent word pool, rather than the kinds of non-word errors that are generated by optical character recognition. Recent proper nouns, especially names, contribute significant error because they can not be transcribed correctly. It seems unlikely that error-free ASR will be available in the foreseeable future.

However, highest precision is not really required for our approach. The goal is not to obtain a correct transcript, but simply to gather enough semantic information to generate a characterization that the system can employ to find relevant content. The interface offers primary the user the original audible content from recordings, because audio is doubtless a much richer medium of communication. Voice quality and intonational characteristics are lost in transcription, and intonational variation has been widely shown to change even the semantics of the simplest phrases. Hence, the presentation of texts is intentionally limited in contrast to (Whittaker, 1999).

An advantage of our system, also respective to the previously mentioned problem, is the low complex but efficient MSD. It enables us to monitor up to 30% more channels with still a good accuracy compared to a system with MSD of higher complexity. In any case MSD and ASR often lead into major difficulties while modern broadcast uses background music for spoken amounts.

During an evaluation time of one month we were able to process up to four radio channels at the same time and integrate the obtained information automatically into our database for instant use On the other hand the monitoring of several data services is possible without any problems. The limitation of ASR could be avoided by splitting up the task by parallel processing which can reduce the lag of time between the recording and the end of the indexing process. The current limitations of the introduced system have to be handled by more efficient speech recognition subsystems, sophisticated semantic retrieval algorithms, and a higher degree of parallel processing. Furthermore, prospectively a more natural communication style using a combination of speech, gesture and contextual knowledge should be possible. Therefore, a system able to interpret the semantics of speech is inevitable.

7 CONCLUSIONS

The Digital Radio was extended with the capability to systematically search for contents in DAB/DMB audio and data services; no major obstacles exist to extend the principles also on HD Radio^{TM}, internet services, and podcasts etc. The functional enlargement of a digital receiver significantly adds value by promoting the evolution towards an embedded device providing innovative functionalities:

- Interactive search for content from audio and data information sources,
- Speech-based output of content,
- Conversion of highly accepted internet services into the broadcast environment.

The information extraction and retrieval process of broadcast information delivers a newspaper-like knowledge base, while web services provide an encyclopedia-like base.

Even if ASR engines could supply accurate transcripts, they are to this day faraway from comprehending all that speech has to offer. Although recognizers have reached a reasonable standard recently, there are other useful information which can be captured from audio recordings in the future: language identification, speaker tracking and indexing, topic detection and tracking or non-verbal ancillary information (mood, emotion, intonational contours, accentuation) and other pertinent descriptors (Magrin, 2002).

Furthermore, the prospective capability of devices to adapt to preferences of specific users offers an enormous variety of augmentations to be implemented. In case of the functionality of directly searching for content elements as described in this paper, there is the possibility of especially selecting respectively appropriately sorting those elements that are conforming to the preferences of the current user.

Hence, the development of radio usage from passive listening towards an interactive and individual dialog is strongly supported and the improved functionalities render the radio to be an appropriate device to satisfy much more multifarious necessities for information than before. As a result users are capable of selecting desired audio contents more systematically, with higher concentration and with higher density of information from current and past programs.

REFERENCES

ComVision, 2009. Audioclipping. http://www.audioclipping.de. [Acc. 10.03.2009].

Emnett, K. and Schmandt, C., 2000: Synthetic News Radio. *IBM Systems Journal* vol. 39, Nos. 3&4, pp. 646-659.

ETSI, 2004. Broadcast and Online Services: Search, select, and rightful use of content on personal storage systems TVA; Part 3 In: *ETSI TS 102 822-3-1*.

ETSI, 2005. Digital Audio Broadcasting (DAB); XML Specification for DAB Electronic Programme Guide (EPG). In: *ETSI TS 102 818*.

Glavitsch, U. and Schäuble, P., 1992. A system for retrieving speech documents. In: *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.168-176.

Herrmann, F. et al., 2007. The Evolution of DAB. In: *EBU Technical Review*. July 2007, pp. 1-18.

Magrin-Chagnolleau, I. and Parlangeau-Vallès, N., 2002. Audio-Indexing: what has been accomplished and the road ahead. In: *Sixth International Joint Conference on Information Sciences - JCIS'02*, pp. 911-914.

Minker, W. et al. 2006. Next-Generation Human-Computer Interfaces. In: *2nd IEE International Conference on Intelligent Environments*, 2006.

Nathan, D. et al., 2004. DAB Content Annotation and Receiver Hardware Control with XML. *Computer Research Repository (CoRR)*, 2004.

Porter, M. F., 1980. An Algorithm for Suffix Stripping. *Program-Automated Library and Information Systems*, vol. 14, July 1980, No. 3, pp. 130–137.

Sawhney, N. and Schmandt, C., 2000. Nomadic radio: speech and audio interaction in nomadic environments. In: *ACM Transactions on Computer-Human Interaction*, vol. Volume 7, pp. 353 - 383.

Schatter, G. and Zeller, B., 2007. Design and implementation of an adaptive Digital Radio DAB using content personalization. *IEEE Transactions on Consumer Electronics*, vol. 53, pp. 1353-1361.

Schatter, G., Bräutigam, C. and Neumann, M., 2006. Personal Digital Audio Recording via DAB. In: *7th Workshop Digital Broadcast.*, Erlangen, pp. 146-153.

Schäuble, P. and Wechsler, M., 1995. First Experiences with a System for Content Based Retrieval of Information from Speech Recordings. In: *IJCAI-95 Workshop on Intelligent Multimedia Information Retrieval.*

TVEyes, 2009. Podscope - The audio video search engine. http://podscope.com/. [Acc. 10.03.2009].

Whittaker, S. et al.., 1999. SCAN: Designing and Evaluating User Interfaces to Support Retrieval. In: *Proceedings of ACM SIGIR '99*. pp. 26–33.

Zaykovskiy, D., 2006. Survey of the Speech Recognition Techniques for Mobile Devices. In: *11th International Conference on Speech and Computer*, St. Petersburg.