

DMB (DIGITAL MULTIMEDIA BROADCASTING) VOICE EPG APPLICATION

Bong-Ho Lee, So Ra Park, HeeJeong Kim, ChungHyun Ahn, Soo In Lee

Electronics and Telecommunications Research Institute

E-mail : leebh@etri.re.kr

ABSTRACT

Recently, mobile TV is becoming a mainstream service in mobile broadcasting scenario, where requires lots of mobile factors such as robust transmission, high performance and easy interface and so on. As mobile broadcasting services are being highlighted, the easy interface and appropriate application are increasingly in demand. In most mobile scenarios, where users may actually share their attention with other concurrent tasks and where highly integrated devices may have very limited physical characteristics, intuitive man-machine interfaces are key factors to successful applications. The EPG, an essential application in most digital broadcasting systems, is also not free from the easy interface and mobile factors. The conventional GUI driven EPG solutions are, sometimes, not appropriate to the mobile system aiming the mobile TV and rich interactive data services. In this paper we present voice enabled EPG application that features voice user interaction and dialog technology allowing the user to have a speech interaction with the terminal in navigating and searching any program or service. We illustrate an overall service framework addressing the content delivery and consuming architecture fitted to the DMB environment. Moreover, we propose and implement an agent platform by profiling the elements of VoiceXML and extending EPG related elements to enable the EPG functionalities.

1. INTRODUCTION

DMB(digital multimedia broadcasting) is a mobile broadcasting system targeting the mobile TV and interactive data services. It has built on the basis of Eureka-147 DAB system designed for the delivery of high-quality digital audio program and data services for mobile, portable and fixed terminals[1]. The state-of-the-art coding technology MPEG-4 along with MPEG-2 has incorporated into the well designed DAB system to enable the mobile TV service. Moreover, DMB has capabilities to deploy various data applications ranging from DLS(dynamic label service), slideshow, BWS(broadcasting web sites), BIFS(binary

format for scene), and even to EPG(electronic program guide).

With the advent of digital broadcasting, a multitude of programs become available at the user's terminal[2]. This is why the EPG is normally appreciated as an essential application in digital broadcasting. In order for the user to get access to a specific program what he wants, the EPG platform has to provide well-structured guide information and an easy and convenient interface to the user. Currently, there are plenty of EPG providing solutions and specifications, basically from the proprietary implemented solutions to the TV anytime metadata based enhanced solution[3].

For the DAB/DMB EPG service, EPG XML[4,5] is specified to form its data structure and deliver the program information available on DAB/DMB ensemble. This specification defines the XML schema for the DAB/DMB EPG. Hence, the EPG data being created by EPG XML can be sent either a file or binary bit stream format. It also can be used as an exchange format among the content creators or even the broadcasters.

Taking a look at it from the user interaction, it sophisticatedly specifies program relevant information by defining lots of tags and attributes. It is basically made for a simple text or GUI based EPG solutions. However, even if EPG XML is highly optimized for the description and delivery, it seems not insufficient for mobile EPG service. The insufficiency is on the user interaction that requires relatively an easy interface appropriate to the mobile environment. Especially, when the driver wants to select a program by browsing the EPG panels while he is driving, the text or GUI based platform might inconvenient and sometimes very difficult.

On the other hand, if the program information can be rendered into a speech format by synthesizing the text or playing an audio clip, this approach seems very convenient to the mobile user. To accomplish this goal, we present an overall service framework, an implemented platform (voice EPG platform) and the EPG related elements. In the following section, along with the overall service framework, this paper introduces the VoiceXML[6] designed to provide interactive voice applications that feature TTS(text to speech), ASR(automatic speech recognition) and man-machine speech interaction. As a next step, we define key

EPG associated elements to provide a way of access to the receiver resource for the EPG functions like program selection. Finally, we conclude this paper with the experimental result being made for the evaluation and with some future works.

2. VOICE EPG SERVICE FRAMEWORK

This voice enabled EPG service can have an overall service framework as illustrated in figure 1.

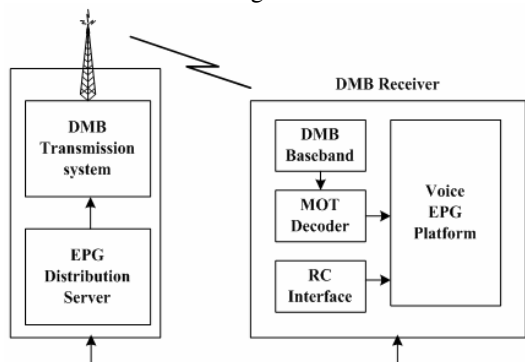


Figure 1. Voice EPG Service Framework

The transmission side is briefly illustrated of EPG distribution server and its main DMB transmission system. The key role of the EPG distribution server is to feed the EPG contents through the DMB network and interact with the receiver via the interactive network. The DMB transmission system composing of several DMB sub-systems such as media encoder, ensemble multiplexer, and data encapsulator like MOT[7] encoder is to transmit the given DMB signal into the air.

At the receiver side, the DMB baseband, MOT decoder and RC(return channel) interface modules are integrated into the voice EPG platform to acquire and eventually execute the EPG contents. In terms of the DMB baseband, it takes responsible for the baseband processing such as tuning, ensemble de-multiplexing and sub-channel decoding. The sub-channel data from the baseband module is fed into the MOT decoder for decoding so that the data can be restored to the original EPG contents. Whereas, RC interface module is activated when there is demand for the interactive EPG contents customized to the user preference specifically. In the end, those restored and retrieved EPG contents via the unidirectional and bidirectional network are consumed on the voice EPG platform.

3. VOICE EPG PLATFORM

This proposed voice EPG platform is designed based on the open platform, VoiceXML, which is approved as a voice web standard. This VoiceXML is designed for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, etc. It major goal is to bring the

advantages of web-based development and content delivery to interactive voice response applications[6].

Basically, the main feature enabling the voice enabled EPG application is to define a conversational state between the user and the machine. This is commonly called “dialog” and it is embodied by XML tags. Additionally, the interaction means both system output and the corresponding user input need to be well structured with the speech dialog. Moreover, a dialog control flow mechanism is also required to ensure the seamless execution of a multitude of dialogs since the interactions are normally processed according to the link from dialog to dialog, dialog to document and document to document, for instance. Lots of additional things including above two key features are well specified in the VoiceXML[6].

However, this relatively pervasive VoiceXML platform is required to be profiled for the mobile usage by imposing constraints on the unnecessary elements and functions, ultimately to make a mobile version of VoiceXML. In this section, we elaborate not only on the profiling, but also on the extending EPG associated elements.

3.1. Voice EPG platform

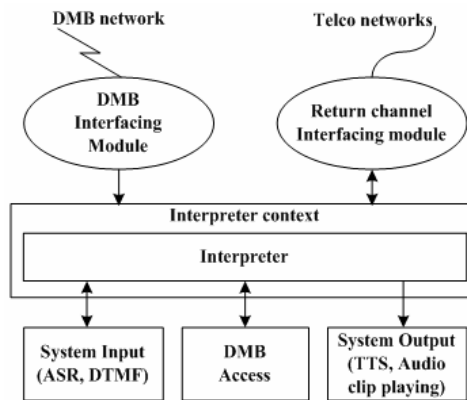


Figure 2. Voice EPG platform

Figure 2 illustrates the general blocks of voice EPG platform. The DMB interfacing module is to decode the MOT packet and restore the relevant EPG data from the DMB network. Precisely, it is responsible for the retrieval and aggregation of the voice EPG data as well as the signaling and binding of the contents to the actual execution engine, the interpreter.

Similarly, the return channel interfacing module is involved in the return channel side of data acquisition and binding as well. The interpreter context plays a key role in operating and managing of the overall execution of applications. The interpreter is a core part of voice EPG platform, and hence it performs lots of key tasks such as ranging from the document loading, interpreting, handling of FIA(form interpretation algorithm) and interfacing with the integrated three platform components, together with its enclosing interpreter context.

In the meantime, the underlying platform components deal with the user interaction and the control of receiver. The user commands are recognized and the results are thrown to the interpreter via the user input component having ASR and DTMF button interface module. Meanwhile, composing of TTS or audio playing module the system output component output the EPG information into speech either synthesizing or playing. Besides, the DMB access component plays a main role in enabling the application to gain access to the DMB hardware resource to get some DMB data and mainly control the behavior of the receiver.

3.2. The profiling of VoiceXML

Since VoiceXML is made for the interactive voice response applications that mainly target the call center applications and server based voice portal services, it needs to be profiled in some elements or functions which seem to be unnecessary for the voice EPG application.

Firstly, the telephony control is eliminated because this voice EPG application is based on broadcasting scenario. Therefore, the <transfer> and <disconnect> elements are not supported.

Secondly, the <record> element is also eliminated because it is normally used when the user input is required to be recorded to transmit the recorded file to the server. This application scenario is not so practical in this voice EPG application.

Thirdly, <log>, <meta> and <metadata> elements are not supported. Concerning the speech grammar where are two equivalent languages, XML form and augmented form of the W3C SRGS(speech recognition grammar specification), the former is only allowed for the voice EPG applications in order to have a relatively lightweight solution. In fact, XML form is a normative language rather than the augmented one recommended as an optional way and additionally the latter case requires a proprietary parser for the interpretation of it.

With above profiling, we can have relatively mobile version of VoiceXML engine. And therefore we are able to seek out or select any program just by saying speech keyword and by taking advantage of the voice user interface.

3.3. VoiceXML Extensions for the voice EPG

Despite the profiling of VoiceXML, extensions are necessary to offer application an access to the hardware resource for the controlling of DMB EPG properties like ensemble and programs. The controlling means a normal EPG function such as tuning an ensemble. So, this section describes the extensions of VoiceXML in order for the application to get access and let it command and perform EPG related operations to the hardware resources.

The mainly required things for this purpose are to offer methods of tuning of ensemble and selecting, reserving and

recording of specific service or programme being scheduled in a certain ensemble or other ensemble.

For the former case, we define <tuning> element to embody the ensemble tuning when it is needed as given in table 1. It has two attributes of “ensembleid” and “frequency”. The “ensembleid” is a unique identifier for a specific ensemble and the “frequency” is an actual frequency of it.

With regard to the service or program selection, there could be three behaviors such as the selection, reservation and record. Literally, above behaviors are to let the user get tune into a specific program he wants, reserve a program scheduled to come, and record it promptly or at the time of available, respectively. These three behaviors are embodied into each element as listed in table 2.

Table 1, <tuning> element

ensembleid	a unique identifier. i.e, “e1.ce15”
frequency	an actual frequency (kHz). i.e, “201072”

Table 2, <watch>, <reservation> and <record> elements

elementtype	It represents the type of EPG contents having either “service” or “programme” as attribute values. i.e, “e1.ce15.c221.0”
elementid	It indicates each unique identifier of signaled EPG content in the type attribute, having “service id” and “programme id” as attribute values.

With above extended elements, the application can indirectly access to the underlying hardware resource especially for the EPG relevant operations. From the execution perspective, when the interpreter encounters one of above elements which is normally situated in <form> elements, the interpreter should send out the designated commands and parameters to the underlying DMB access module as illustrated in given figure 2. In response to the interpreter’s command, the DMB access module will perform the operation in a proprietary way by using the receiver specific APIs, for instance.

In the case of alternative input mean DTMF, it is also extended to enhance user interaction available of button by adding navigation keys to the existed DTMF keys. The extended keys are “main”, “exit”, “forward”, “backward”, “previous”, “next” and “refresh”. Here is an example of voice EPG content that shows a common program guide that allows the user to tune in or reserve any program.

```

<form id="program_21">
  <field name="input">
    <grammar type="application/srgs+xml"
      src="/grammar/user_command.grxml"/>
    <prompt timeout="86400000">
      This entertainment tonight will guide
      you to Hollywood world today!
      Please say “watch” or “reservation”
      if you want to tune in or reserve it.
  
```

```

</prompt>
<filled>
  <if cond="input == 'watch'">
    <watch elementtype="service"
      elementid="e1.s1.p21.0"/>
  <elseif cond="input == 'reservation'">
    <reservation elementtype="programme"
      elementid="e1.s1.p21.0"/>
  <else/>
</filled>
</field>
</form>

```

4. EXPERIMENTS

We implement this proposed voice EPG platform in order to verify the performance and to make sure of extended EPG associated elements and its functionalities. This verification platform incorporates Korean and English available ASR and TTS engines into the voice EPG engine under the system environment of Windows XP and C++ language. Figure 3 shows an example of voice EPG contents having three navigation layers from ensemble page for the general guide information to each characterized pages for specific information in conjunction with 4 service channels. These contents are delivered through the DMB network using MOT protocol.

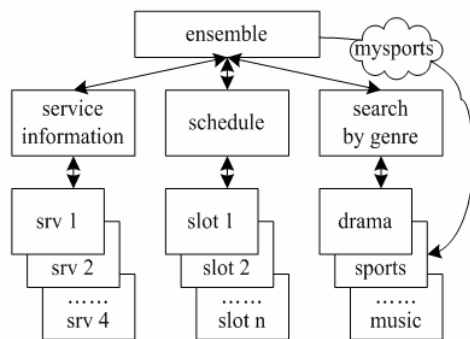


Figure 3. An example of Voice EPG contents

At the receiver side, when the platform is initialized the ensemble document is loaded and interpreted first to offer common ensemble and program information. Normally it gives general information of ensemble and its navigation information to the sophisticated program or even the genre information. The service information document is to provide the program information of each service. On the other hand, the schedule document offers program information pertinent to any time slot. As an intelligent EPG service, search document listing programs per genre is to enable the searching capability through the speech input. Though these search pages, the user can get into any genre page directly by saying registered keyword and eventually can reach any program what he needs.

Navigation point of view, through the uttering of speech keyword or pressing of a button while listening to the narrations from the ensemble page, the user can browse to any specific document layer to layer or sometimes spontaneously to a specific document.

One opportunity of this platform is that the user can have intelligent browsing experience simply by registering the special keywords having global scope of grammar in order to let the user jump to any specific program information. So, for example, if the “mysports” is registered as a global keyword having public grammar scope as illustrated in figure 3, then the user can directly jump to a document tailored for the sports programs from every document where he interacts.

5. CONCLUSION

Work performed has led to the proposal of overall service framework, extension of elements and to the implementation of platform to verify it. With this platform we can convince that this voice enabled EPG application can be useful and can offer lots of conveniences to the mobile user in browsing the EPG contents. Consequently, this platform has a couple of future works to be enhanced of concurrent representation with visual, user interaction reliability, remote interactivity and efficient delivery and so on.

6. REFERENCES

- [1] ETSI EN 300 401 “Radio Broadcasting Systems; Digital Audio Broadcasting(DAB) to mobile, portable and fixed receivers” May 2001.
- [2] Hee-Kyung Lee, Han-Kyu Lee, Jeho Nam, Beetnara Bae, Mulchul Kim, “Personal Contents Guide and Browsing based on User Preference”, *Proceedings of the AH'2002 Workshop*, Spain, May, 2002.
- [3] TV Anytime : Specification Series : S-3 on Metadata, SP0003V1.1, August 17, 2001.
- [4] draft ETSI TS 102 818 V1.3.1 (2005-09), Digital Audio Broadcasting (DAB); XML Specification for Electronic Programme Guide (EPG).
- [5] draft ETSI TS 102 371 V1.2.1 (2005-09), Digital Audio Broadcasting (DAB); Transportation and Binary Encoding Specification for Electronic Programme Guide (EPG).
- [6] Voice Extensible Markup Language(VoiceXML) version 2.0, W3C Working Draft, 23 October 2001, <http://www.w3.org/TR/voicexml20/>.
- [7] ETSI EN 301 234, “Digital Audio Broadcasting (DAB); Multimedia Object Transfer (MOT) protocol”.