An Implement of Speech DB Gathering System Using VoiceXML

Dong-Hyun Kim\(^1\), Yong-Wan Roh\(^1\) and Kwang-Seok Hong\(^1\)

\(^1\)School of Information and Communication Engineering, SungKyunKwan University, 300, Chunchun-dong, Jangan-gu, Suwon, KyungKi-do, 440-746 Korea
Super1621@hotmail.com, elec1004@hotmail.com, kshong@skku.ac.kr

Abstract

In this paper, we introduce speech DB gathering system using VoiceXML. In general, speech DB is very important to speech recognition and synthesis system. But, in the present system which does not use voiceXML, compatibility between different kinds of systems needs much labors and expenses. While VoiceXML is a standard dialog mark-up language for the next generation voice applications. For this reason, nowadays, many company uses VoiceXML when they implement their systems for speech applications.

If we implement voice information services using VoiceXML, service developers can save their labors. And, service operators can manage their system easily. If there’s need to change their service scenarios, operators can change their service scenarios using Graphical Users Interfaces. It doesn’t matter to operators that whether they know technical problems or not. As the results of above mentioned, more and more companies will use VoiceXML when they implement their applications and importance of VoiceXML will be increased. In this paper, we developed speech DB gathering system using VoiceXML for the purpose of verifying technical problems.

1. Introduction

The voice portal market is exploding with enormous opportunities for service providers to grow business. Voice-based internet access uses rapidly advancing speech recognition technology to give users anytime, anywhere access to web-based information. And it uses that most universal form of communication and access—the human voice—over an office, wireless, or home phone. Because of various application fields, gathering of speech DB is required. Speech DB offers additional information and document to reuse speech relation developer at all times, and is collection of speech data that are composed of readable form by computer [1]. Speech DB is very important factor in the signal process like speech recognition, speech synthesis and so on. In general, speech DB gathering system of platform like a conventional IVR (Interactive Voice Response) hard to adjustment and maintenance repair and it is implemented to specific field and service suitably. So, it is necessary much time and many costs when we add new service. VoiceXML is designed to resolve these controversial points and be easiness for use and work of internet contents [2]. VoiceXML is proposed by Voice XML forum that is founded by AT&T, IBM, Lucent and Motorola. The VoiceXML is an interactive markup language, which supports an environment to develop applications that search web information via voice and phone. It is designed for creating audio dialogs that feature synthesized speech, recognition of spoken and digital tone multiple frequency (DTMF) key input, telephony, and mixed initiative conversations. Speech DB gathering system of platform like a conventional IVR was hard to add new function and required special technology for DB interface. Also, it has many problems that much manpower and time is required during the construction system. In this paper, we gather speech DB using VoiceXML in order to solve the problems. When we are gathering the speech DB using VoiceXML, Advantages of this system are followed. New function is easy to add and to interface to the DB using Oracle, SQL, Microsoft Access and so on, since the VoiceXML document independently exists form in the system. Also, VoiceXML document forms have many advantages of saving manpower and time because it has tag form involved in XML(Extensible MarkUp Language)[3]. We implemented the speech DB gathering system. The total gathered speech DB is 1568 units that are name of stocks and related to sentences of stocks transaction.

2. VoiceXML gateway

VoiceXML gateway connects telephone network and internet network, transmits URL to web servers like a web browser of HTML, analyzes VoiceXML documents that web servers send and performs function of rendering.

Figure 1 shows the composition of VoiceXML gateway, when a user calls with a wire or wireless phone the call is transmitted to the VoiceXML system through telephone network and the internal system receives the call signal of the phone and requests a VoiceXML document to the web server through the internet network by synchronizing an interpreter [3]. When the web server finds the corresponding VoiceXML document and transmits it to the VoiceXML system, parsing of the document is performed in the interpreter and a synthesized voice made by text-to-speech (TTS) synthesis is transmitted to the user by processing a scenario [4].
2.1. VoiceXML interpreter

The VoiceXML interpreter analyzes the VoiceXML document that the server sent and controls operation platform in order to perform interaction with the user.

![Image](image1.png)

Fig.2. VoiceXML architecture

Figure 2 shows a structure model of the VoiceXML and each function will be described as follows.

2.1.1. Document server

The document server processes requests from the VoiceXML Interpreter. The server produces VoiceXML documents in reply, which are processed by the VoiceXML Interpreter. The VoiceXML interpreter context may monitor user inputs in parallel with the VoiceXML interpreter.

2.1.2. VoiceXML interpreter

VoiceXML interpreter loads the VoiceXML document and performs processing of the document by analyzing the document. It is a core element for the environment of VoiceXML operation. It controls sequential flow of the document processing in accordance with 47 kinds of tags related to dialogue, grammar, event, audio output, call control, audio input and flow control by analyzing the voice application scenario, and sends command signals which are necessary for speech platform by determining the speech input and output. Also, it generally controls operations of the VoiceXML document, such as downloading of necessary resources through the document server or transferring the resources into another document.

2.1.3. VoiceXML interpreter context

VoiceXML interpreter context may be responsible for detecting an incoming call, acquiring the initial VoiceXML document, and answering the call, while the VoiceXML interpreter conducts the dialog after answer.

2.1.4. Implementation platform

The implementation platform generates events in response to user actions and system events. Some of these events are acted upon by the VoiceXML interpreter itself, as specified by the VoiceXML document, while others are acted by the VoiceXML interpreter context.

2.2. Speech recognition

Speech recognition is extracting phoneme, language information from acoustic information of included speech which is a process recognized and reacted by machine [5]. Speech has many advantages of easy and natural a means of proceedings delivery and is not necessary device of a high price in the input or delivery of speech process, and is recognized of effectiveness in many fields from man-machine application interface [6].

Figure 3 shows configuration of speech recognition system that is classified into preprocessing department and recognition department. Preprocessing department extracts feature vector for recognition from speaker uttering a speech. Recognition department acquires recognition result in comparison with trained standard pattern. Final recognition result is generated by language process using language model when we recognize speech of complex structure [7] [8].

![Image](image2.png)

Fig.3. Speech recognition system

2.3. Speech synthesis

Text information is transmitted by synthesizing as voice having clear and natural features which are similar to those of actual human speech [9]. Text-to-speech (TTS) synthesis system is a system that reconstructs input sentences as a speech-centered symbol string through language processing department and speech signal processing department then synchronizes it. Figure 4 shows a general TTS system.

![Image](image3.png)

Fig.4. Speech synthesis system

Basic structure of the speech synthesis system is shown in figure 4. The system has an overall structure that makes synthesis sounds through a language processing department and speech signal processing department about text input.
3. Speech DB gathering system based on VoiceXML

Speech DB gathering system based on VoiceXML was designed as an interactive speech DB gathering system. It has a talk with the user by using speech recognition and speech synthesis engines.

3.1. Entire system module

Figure 5 and 6 are an entire block diagram and a flow chart corresponding to the system designed in the present paper. In figures 5 and 6, a Dialogic board receives the call signal through PABX and the telephone network during the user call. After that, the board requests the VoiceXML document from the web server through the internet by synchronizing the interpreter. When the web server finds the corresponding VoiceXML document and transmits the document to a VoiceXML gateway, the document is parsed in the interpreter in case an announcement is needed, it is transmitted to the user through the TTS. The user hears the announcement and inserts his or her ID and password. In case the ID and password information is identical as those registered in the database after searching the ID and password from the user DB, a speech list is brought from the DB and the speech is sent to the user by converting the list to the TTS.

When the user speaks after hearing the sound from the phone, the speech is heard by the user again and the phone asks the user whether to store the speech. The user determines whether to store the speech or not in the last step. The system gathers the entire speech DB through a series of processes. The process will be described through each module in detail.

3.2. New user module

Only IDs and passwords which are combinations of 3 digits are set to be valid. The ID and password of a new user are set in advance and stored in the user DB. The system searches the user DB for checking whether the user is one of the users stored in the user DB after receiving the ID and password when the new user logs in as in figure 7. In case the user is not a user stored in the user DB, an announcement to check the ID and password is made and the process is restarted. When the user is determined as the user stored in the user DB, a folder named after the ID of the new user is generated.

![Fig.7. New user module](image)

3.3. Registered user module

Registered user module is shown in figure 8, if folder of user ID is not made or if the password is not identical after receiving inputs of the ID and password, the user is not a stored user. In this case, an announcement to check the ID is made and then the process is returned to the initial. In case the user is a user stored in the user DB, an announcement of User name is called and “nice to meet you.” is made. The process moves into the folder of user ID name. Then the speech list is converted to the TTS after being brought from the speech list DB. This login processing is performed in order to generate only one folder for an ID of each user and store the user's own speech DB only in the folder without storing speech DB of another user.

![Fig.8. Registered user module](image)

3.4. Speech recording module

A speech recording module is shown in figure 9. The speech list is brought from the speech list DB and the user hears a speech sound by converting the list to the TTS.
The user speaks after hearing the sound from the telephone and the user hears the sound to check whether the user speak properly. Then an announcement of “Will you store it?” is made. If the user thinks the speech is not proper, the user says “no” and the process is returned to the same speech list step and repeated. In case the speech is proper, the user says “yes” and the count of the speech list DB increases and it goes on to the next speech list. The system was implemented so that files are stored in the folder of user ID name by using ASP (Active Server Page) in the storing step and accordingly files of each user are stored in the corresponding folder.

4. Analysis of implemented system

4.1. Development environment

In developing the above system, windows 2000 professional was used, and IIS (Internet Information Server) 5.0 which is only for window NT was used as the web server. The system was built by using ASP (Active Server Page) for login processing, folder generation and file transmission, and Microsoft Access was used for storing the speech list in the DB. Intel Dialogic 41JCT/LS used for call control and HUVOICE 1.0 of KT used as its Interpreter, speech recognizer and synthesizer.

4.2. Results

Figure 10 shows a waveform view of speech gathering files gathered with the present system. The speech can be received by 8-bit quantization in 8 kHz sampling rate. Figure 10 shows a waveform and spectrum corresponding to a speech of “Sungshinyanghoesamwoobee”. Advantages of this system

are followed. New function is easy to add and to interface to the DB using Oracle, SQL, Microsoft Access and so on. Because the VoiceXML document independently exists from the system. Also, advantage of the format of the VoiceXML document is saving manpower and time, because the format has a tag format implied in XML (extensible markup language).

5. Conclusion

Speech recognition and speech synthesis technologies have been expanded to many fields of service applications, such as voice portal service. The more expand to various kinds of service applications, the more necessity for gathering speech DB occurs. In the traditional IVR systems, speech DB gathering functions implemented using C/C++ languages or exclusive development tools.

In this paper, we implemented speech DB gathering system through wired/wireless telephone network using VoiceXML. For these reasons, we reduced development time for application services. Principle features of implemented system using VoiceXML are as follows. There’s no need to know technologies of speech recognition and speech synthesis system. In other words, we can change the flows of provided service with changing several documents of VoiceXML. Also, the announcement speech of the steps can be changed by modifying the VoiceXML documents since we implemented document server independently from the system. And other features like strengthened security functions, user interface through IP network, easy reuse of implemented resources and so on can be said.

6. Reference