Implementation of Speech Recognition System Using JAVA Applet

Seungho Choi*, Kwangkook Choi*, Kyungnam Kim**, Jinyoung Kim***, Kijung Kim****
*Department of Information and Communication Eng., Dongshin University.
**Department of Information and Communications, Kwangju Institute of Science and Technology.
***Department of Electronics, Chonnam University.
****Department of Internet Information Technology Kwangyang College.
520-714, Daehodong, Naju, Chonnam, Republic of Korea.
Tel: +82-613-330-3194, Fax: +82-613-330-2909
E-mail: shchoi@dongshinu.ac.kr, choicomm@ttl.co.kr, knkim@geguri.kjist.ac.kr, kimjin@dsp.chonnam.ac.kr, kjkim@kwangyang.ac.kr

Abstract: In this paper, a word-unit recognition is performed to implement a speech recognition system over the web, using JAVA Applet and continuous distributed HMM. The system based on Client/Server model is designed. A client computer processes speech with Applet, and then transmits feature parameters to the server computer though the Internet. The speech recognition system in the server computer transmits the result applied by the forward algorithm to the client computer and the result is displayed in the client computer by text.

1. Introduction

Recently, the speech recognition system in the Internet has been implemented since the publication of JSAPI(Java Sound Application Programmer's Interface) in 1998.

If the multimedia data are implemented by the HTML, the JAVA O/S can directly access to them over the web. Otherwise, the multimedia data are integrated and processed using a helper application or a third party S/W application[1].

One of the Characteristics of JAVA is to be able to access and share the information on the Internet without limitations. The Java code and class do not affect other applications. Because the JAVA is appropriate to develop application programs which are in distribute computing structure and communication environment, new communication service can be created appling the JAVA to the speech recognition system.

In this paper, we have implemented the word-unit HMM Speech Recognition System in the web which can be applied to the Traffic Information System, the Hotel Reservation System, the Travel Information System and so on.

2. The Implementation of the Speech Recognition System using JAVA

2.1 JAVA Applet

Java Applet is designed to execute JAVA program on the Web Browser such as the Netscape Navigator or the Microsoft Internet Explorer. However there is a security problem related to user computers on the web, therefore the proposed speech recognition system adopts the SoundBite class of the third party program in Scrawl, Inc, USA(http://www.scrawl.com) to solve the problem and to recode speech signal[1].

2.2 JAVA Applet processing

Applet plays three roles. First, the speech recoding of users, second, the preprocessing for extracting speech feature parameters, and third, the socket generation for users to transmit feature parameters to the recognition server.

1) The design of the record and play

The speech record/play diagram using SoundBite class is depicted in Figure 1.

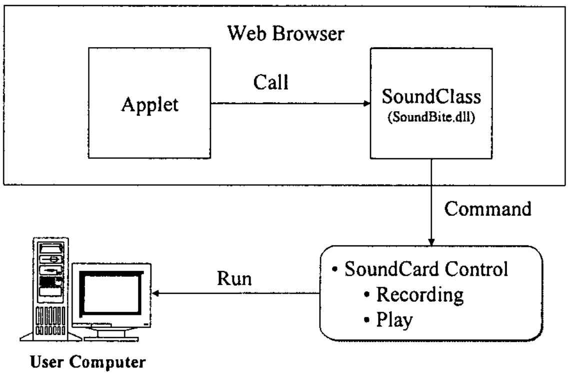


Figure 1. The speech record/play diagram.

2) Extraction of the Mel Frequency Cepstral Coefficient(MFCC)

Figure 2 describes the process to obtain the MFCC after the real-time speech extraction[2].

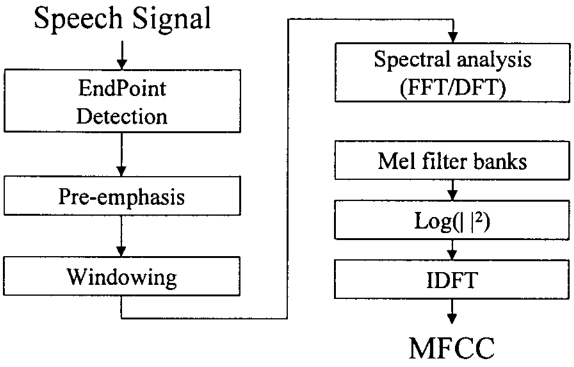


Figure 2. MFCC extraction.

3) The Design of socket generation

After seeking feature parameters, Applet generates a socket for transmitting data to recognition server. If socket is steadily connected, Applet transmits parameters to the recognition server and stands by until the recognition server transmits the recognized result to Applet. Besides, the socket connects the speech process program between a client computer and the server computer with TCP/IP. Figure 3 describes the socket connection between Applet and the recognition server.

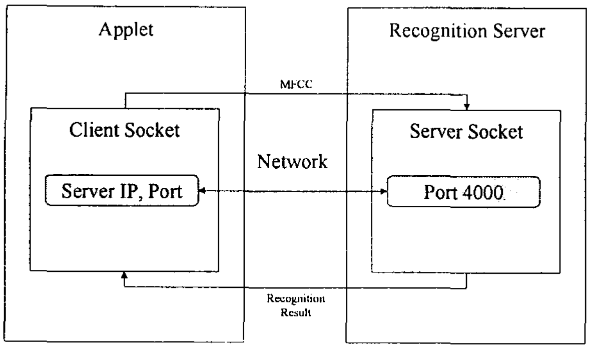


Figure 3. The socket connection between Applet and the recognition server.

2.3 The design of Java Native Interface

When designing application programs, JNI can solve the problem of the impassable representation with Java. It is designed to be executed with the library or application programs written by other languages such as C, C++, assembly inside the Java Virtual Machine(JVM) and Figure 4 describes the design process of the application program[3].

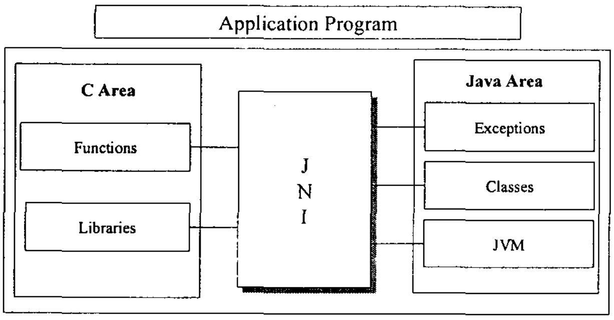


Figure 4. The design process of the application program.

Accordingly, in this paper, JNI is used in the connection of the sound card control part of client computer and the connection part between the recognition server and the recognizer. JNI, which is to control the sound card of a client computer on the web, uses SoundBite class and this class is designed to be able to use in the web browsers which support Java on platform of MS Window 95/98/NT. Figure 5 describes JNI configuration of SoundBite.

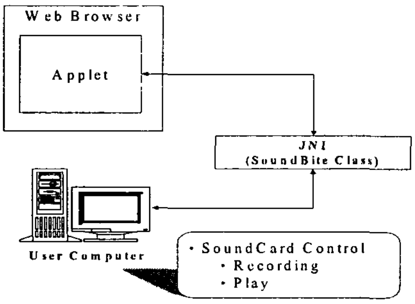


Figure 5. The JNI configuration of SoundBite.

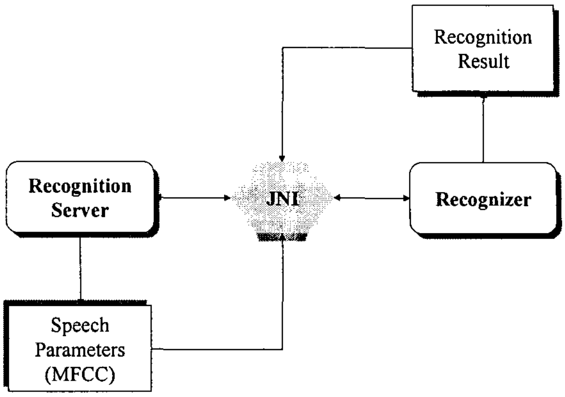


Figure 6. JNI Diagram between the recognition server and the recognizer.

2.4 Data processing of Applet and the recognition server

The socket of the recognition server activates specific port which can be connected by the Applet of a client computer, and generates the socket for the client. Applet is initialized with the IP address and the port number of the recognition server, and request the connection to the server. When connection is done properly, Applet generates packets and transmits to the recognition server.

The recognition server analyzes the packets and delivers the packets to the proper module, and then transmits the recognized result to Applet through the socket. When all processes are finished, the client socket is exterminated and the server socket is re-initialized for a new request.

Figure 7 shows the data processing of Applet and the recognition server.

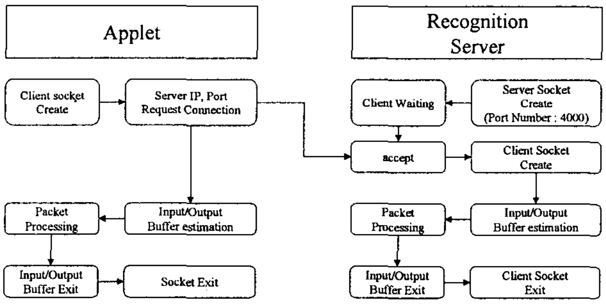


Figure 7. The data processing of Applet and the recognition server.

2.5 Speech recognizer

A speech recognizer estimates maximum probability of training DB and preprocessed feature parameters in Applet using the forward algorithm. It transmits the recognized result to Applet as text and then performs word recognition using continuous distributed HMM.

The 22 words, which are used in a car navigation system, were spoken by the 52 male speakers in the 20's in the silent room and was stored in DB.

The collected data in DB are automatically detected by the end-point detection algorithm and are stored in the form of wave-file with 8kHz sampling frequency, 16bit quantization level, and mono channel.

Each word divided into a frame per 25msec and then each frame extracts the 26 feature parameters like

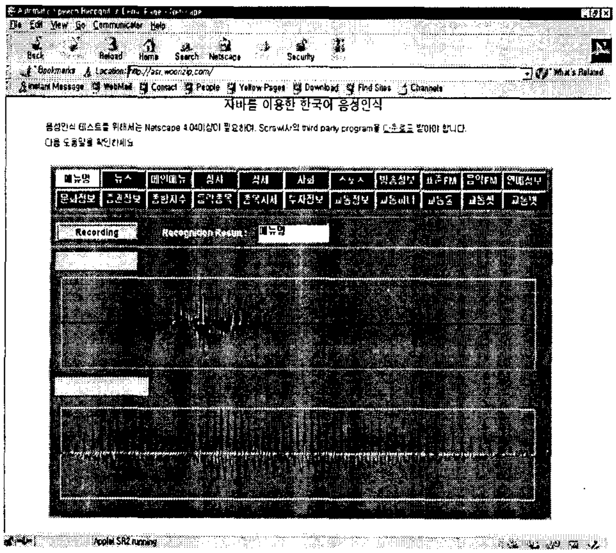
12 MFCC, 12 Delta-MFCC, on log energy, and on Delta log energy[4].

HMM topology sets the initial value of all words as the left-to-right model. HMM initial output values update the parameters until the threshold value converges into 0.0005 using the Gaussian mixture density function, forward-backward, and Baum-Welch re-estimation algorithm. The data are stored in training DB and are used.

3. Design of Graphic User Interface

The graphic user interface was realized over the web using 2D graphics class of Applet. First, when a recording button is clicked, the system records the spoken speech for 3 seconds and displays on the botton of Applet, and the system changes the button color, wich is recognized, for easy check for a user.

Figure 8 shows the graphic user interface on web based on the speech recognition system[5].



자바를 이용한 한국어 음성인식

음성인식 테스트를 위해서는 Netscape 4.04이상이 필요하며. Scraw사의 third party program을 다운로드 받아야 합니다.

다음 도움말을 확인하세요

Recognition Result : 메뉴명

Figure 8. GUI of the speech recognition system

4. Experiment Environment

Experiment environment needs a server computer and a client computer at local area network or Internet, and Intranet. Table 1 and Table 2 shows hardware and software environment of the server computer and the client computer.

Table 1. Environment of H/W

Computer / Item	Server	Client
Processor	450Mhz	150Mhz
Main Memory	128Mbyte	32Mbyte
Network	10Mbps LAN	10Mbps LAN
Sound	-	16Bit Support
Microphone	-	Dynamic 60

Table 2. Environment of S/W

Computer / Item	Server	Client
Operation System	MS NT 4.0	MS Window 98
Web Server	IIS 3.0	-
JDK	1.3Beta Version	1.3Beta Version
C Compiler	Visual C++ 6.0	-
Audio Capture	SoundBite 1.0	SoundBite 1.0
Web Browser	Netscape 4.6	Netscape 4.5

5. Conclusion

The recognition experiment was performed by 5 people who were not participated in the training. Each person spoke 22 words twice and the 220 test data were used in this experiment. In the result, the recognition error rate was 9.1%.

This result shows that new information and a communication service using voice is possible in the future over web.

References

[1] ZhemlnTu, Philips C. Loizou, “Speech Recognition Over the Internet Using Java,” Proc. ICASSP 99, pp. 2267-2370, 1999.

[2] C. Becchetti, Prina Ricotti, *Speech Recognition*, John Wiley & Sons, 1999.

[3] SUN Web Site: http://www.java.sun.com/

[4] Steve Young, The HTKBook(for version 2.2), Entropic Ltd., 1999.

[5] K. K. Choi, J. W. Lee, C. kim, S. H. Choi, “Speech Recognition using HMM over the WWW,” Proc. ASKC, pp. 77-80, 1999.