

Study of Humanoid Robot Voice Q & A System Based on Cloud

Tong Chunya and Zhong Qiubo*

School of Electronic and Information Engineering, Ningbo University of Technology, Ningbo, Zhejiang, 315016, China

Abstract: As voice technology receives wide application in our daily lives, Question-and-answer service is taking shape in the research of humanoid robot. This paper designs a humanoid robot Voice Q & A system based on cloud. After the robot collects voice signals, it would first process the signal to understand the question, and then searches the Internet for the best answers and tell it to the user with voice. In this system, dominant cloud technology is introduced and robot technology, speech recognition technology and Internet searching technologies are combined, providing people with a new way of human-computer interaction through natural language. Results show that with the application of cloud technology, the efficiency of speech recognition and the accuracy of answers have been enhanced substantially.

Keywords: Human-computer interaction, humanoid robot, question-and-answer service, voice cloud.

1. INTRODUCTION

Voice is the most common way of communication. According to statistics, 70%~80% of information is passed down through dialogue. 25%~50% in office involves face-to-face conversation. And 10%~20% time is spent on the phone [1]. Thus, the application of speech recognition technology in human-computer interaction has always been a hot issue. It has a promising prospect.

As robot technology and artificial intelligence are gaining momentum, more and more smart robots come into our lives. But current interaction between human and robot remains the traditional way of order, such as through button or switch. Such communication is dehumanized. In order to talk with robots more convenient and more natural, the human-computer interaction based on natural language becomes increasingly important. In Oct. 2011, on Apple Inc.'s press conference for iPhone 4S, the private voice assistant Siri attracted much attention. The birth of Siri, a "talkative" intelligent robot, indicates that "speech recognition" is no longer a distant dream. And intelligent robot interaction technology based on natural language leads a new wave of strategy.

This paper proposes a humanoid robot Voice Q & A system according to the speech recognition technology of Xunfei Voice Cloud and Baidu's search platform of Question-and-Answer Service Cloud. Question-and-Answer Service Cloud. The system collects voice questions from users and sends them to speech recognition engine of the Cloud. The words are then sent to Baidu Zhidao for answers through the robot and the text answers finally return to the robot.

2. DESIGN FRAMEWORK OF VOICE Q & A SYSTEM

Considering that the native libraries of the speech recognition engine in the robot is complicated and far from

complete. The Voice Q & A system employs the speech recognition technology in the Cloud to process data. The robot would automatically connect to the search engine in the Cloud. The design framework of the system is shown in Fig. (1) (next page). It consists of three modules: NAOrobot, Xunfei Voice Cloud and Baidu Question-and-Answer Service Cloud.

Step 1: Use the voice sensor on the robot to collect voice data;

Step 2: The robot visits Xunfei Voice Cloud and sends the voice data to the Voice Cloud for processing and speech recognition. Output text questions to the robot;

Step 3: The robot receives text questions, visits Baidu Question-and-Answer Service Cloud for answers, and sends back the best answer;

Step 4: The robot receives the text answers, processes texts, integrates voices and speaks the answer out.

2.1. Voice Data Collection

This system involves NAOrobot, an intelligent humanoid robot with four microphones [2]. Voice data are collected by NAOrobot through C++ with the support of Windows, Linux.

The ALSoundExtractor module in the robot can produce a voice processing function. First, set up the format of sound, including sample frequency, channel settings, etc. Then, collect sound clips and send them to the buffer for processing. Finally, after the complete processing and end voice data collection.

2.2. Reception and Transmission of Voice Data

After the robot collects voice data, send the voice data to the speech recognition engine in the Cloud under open network through access point of Xunfei Voice Cloud and send the text data back to the robot through the Internet. The robot will visit the Question-and-answer service Cloud. With the

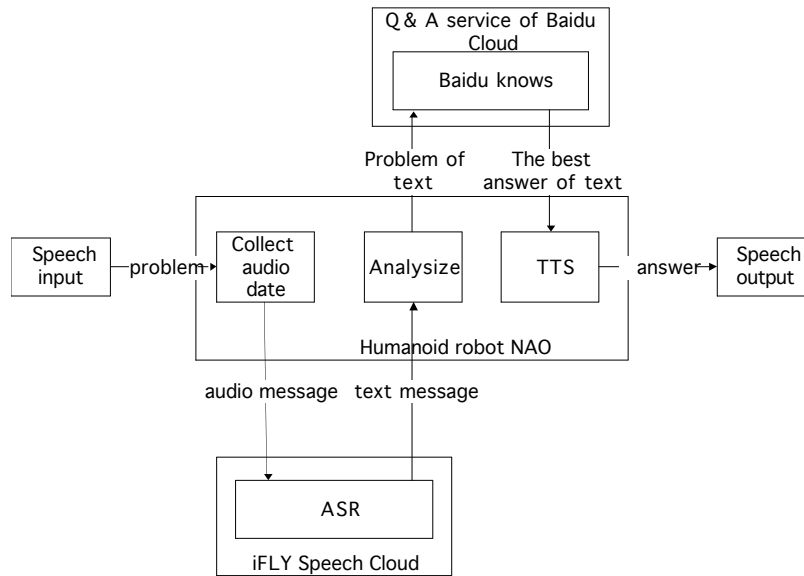


Fig. (1). Design framework of voice Q & A system.

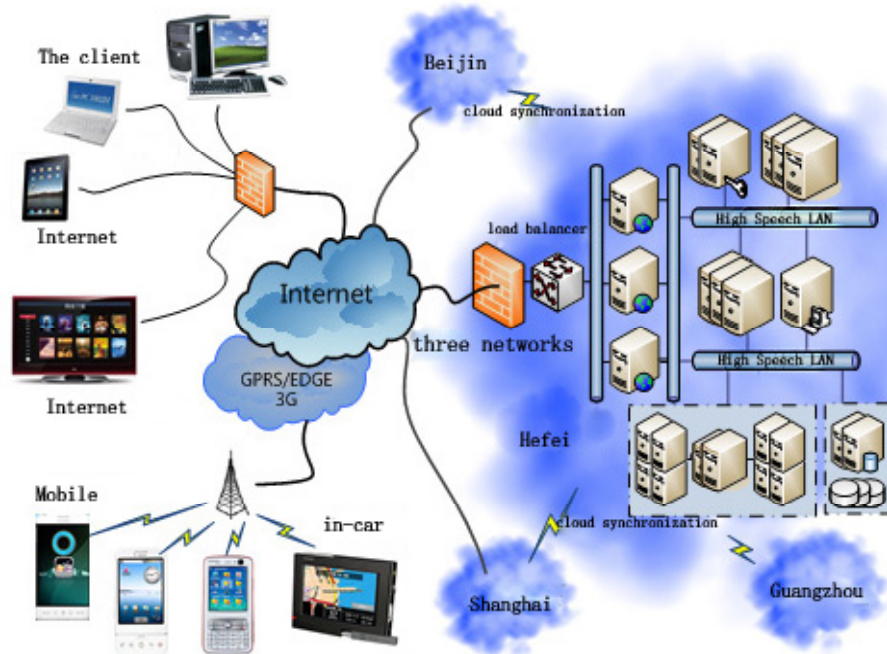


Fig. (2). The network structure of Xunfei voice cloud.

speech recognition engine in the Cloud, the accuracy is largely enhanced.

3. SPEECH RECOGNITION, VOICE INTEGRATION AND QUESTION-AND-ANSWER SERVICE

3.1. Speech Recognition Technology

Speech recognition technology, also named as automatic speech recognition (ASR), is key to human-computer interaction technology. The robot can transmit the voice signal to corresponding texts or orders after recognition and understanding. It equals to put “ears” on the robot and makes it audible. A robot with ears can better fulfill human-computer interaction in a natural way [3]. Now, speech recognition

telephone and speech recognition notebook spring up in the market, such as Voice Organizer [4] of US, VPTC and Parrot [5] of France. Speech recognition system is in essence a multiple recognition system, according to the process of speech recognition by human [6]. Speech recognition should be based on acoustics, phonetics and linguistics in order to have automatic recognition of high quality [7].

In this paper, Xunfei Voice Cloud is introduced to help with speech recognition. The network structure of Xunfei Voice Cloud is shown in Fig. (2) [8]. The main function of Xunfei Voice Cloud is to transmit voice to the Cloud Server and check whether the recognition result is given. If so, send the recognized texts to the local. The process of speech recognition is shown in Fig. (3).

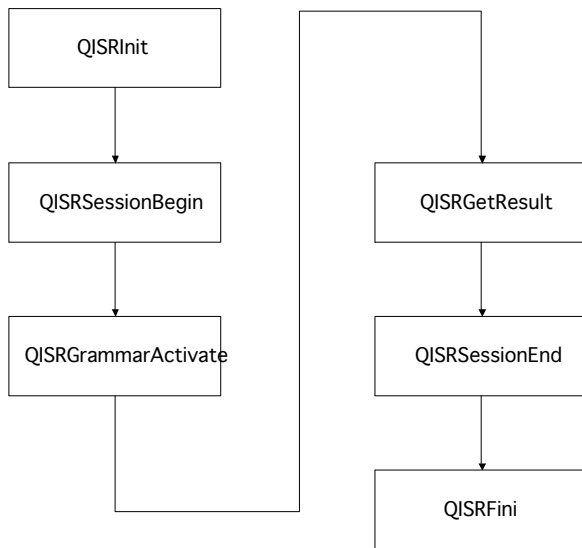


Fig. (3). The process of speech recognition.

3.2. Voice Integration Technology

Voice integration technology outputs the voice signal according to texts and following voice processing rules. So that the robot or the computer can read the texts fluently and people can understand when they hear. It plays the role of a “mouth” and enables the robot to speak [10]. The technology that transmits texts to voice is also named as Text to Speech (TTS).

This system employs voice integration module in the NAOrobot to finish the speaking. But it is worth of noticing that this module supports the text coding of utf-8. However, the default text coding of vs 2010 in the windows system is GBK. So, before the NAOrobot reads the texts, there is a necessity to transmit utf-8 to GBK through Unicode.

3.3. Question-and-Answer Service Cloud

Natural language processing based on large-scale corpus is a traditional way of semantic recognition technology. But the difficulty lies in the establishment of the corpus. As the Internet develops rapidly, it has become a huge and global information space that can serve as the corpus of Question-and-answer service. Google, Baidu, Sogou, SoSo and other search engines are all available. This paper selects “Baidu Zhidao” as the Internet corpus. The most important reason is that among all social question websites, Baidu Zhidao account is for 80% answers. It also enjoys 95% of users. Given the large population of users, the answers provided by Baidu Zhidao can be reliable and can meet the demand of users [9].

Baidu Question-and-Answer Service Cloud is mainly used to search for answers for text questions on the Internet and copy them. Search steps are described as below:

Step 1: Search for the question; through the search, it can get many linked addresses relevant to the question. Rank the addresses according to their correlation. The first link has the largest correlation. Open it.

Step 2: Figure out the best answer; Open the link address in Step 1 and read all answers. Obtain the webpage where the best answers locate.

Step 3: Confirm the text answers; Obtain the source code of the webpage where the best answers locate and analyze the webpage in order to get the text answers.

4. RESULTS AND DISCUSSIONS

Tests are carried out to see the accuracy of speech recognition of Xunfei and relevance of the answer to the question. The Internet should be connected and smooth during the test. This paper selects 10 Chinese (without gender difference) randomly. There are 10 conditions, in which two are related to the environment (quiet and noisy) in the Fig. (4), three are related to the speed of speech (fast, moderate, slow) in the Fig. (5), three are related to ways of question (one proposes 10 prepared questions, each one proposes 10 prepared questions, and each one proposes 10 questions randomly) and two are related to Question-and-Answer Service Cloud website (Sougou Wenwen, Baidu Zhidao).

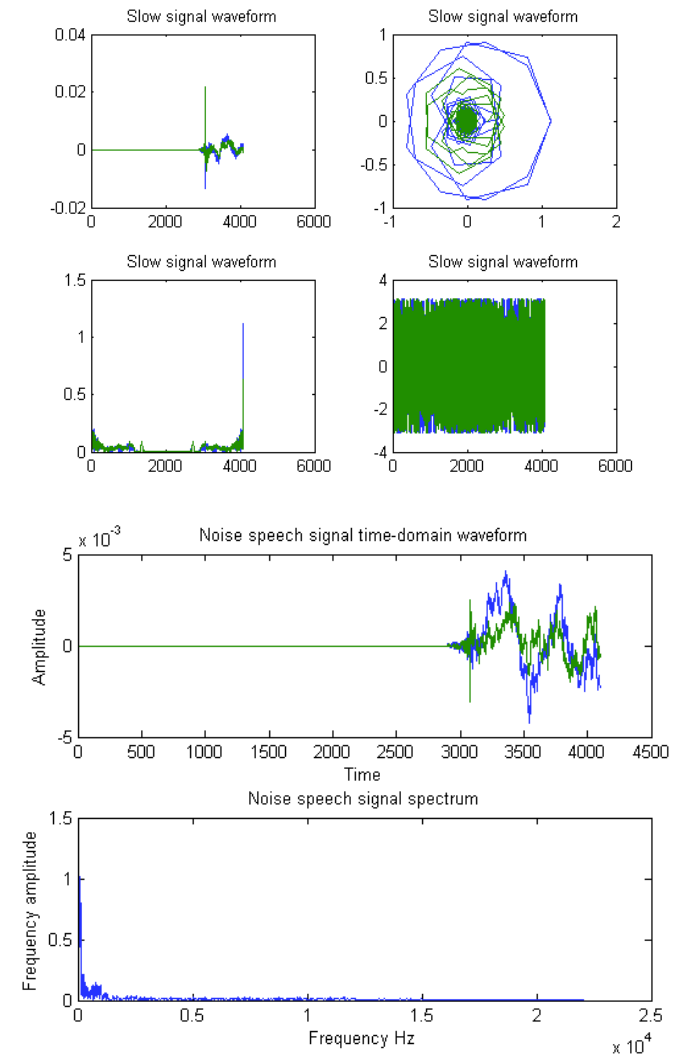


Fig. (4). Voice signals in two environments: “What is the programming language of the computer?”.

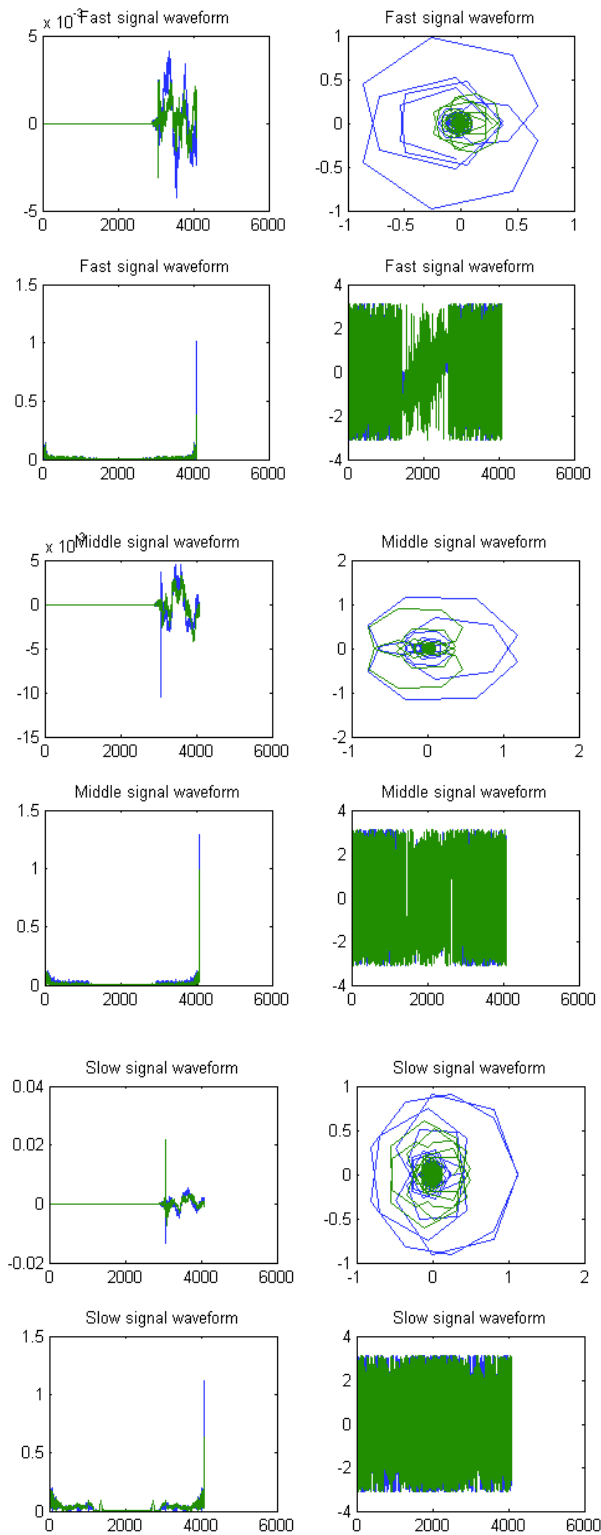


Fig. (5). Three speech speed of voice signals: “What is the programming language of the computer?”.

From Tables 1-4, we can see with the access to the Internet, if the questioner speaks Mandarin, or if the question is reasonable, the speech recognition accuracy and the correlation of the answer can meet the expected requirement. It can be applied to humanoid robot and fulfill human-computer interaction based on natural language.

Table 1. Speech recognition test: each one proposes 10 prepared questions, two environments and three speech speed.

Test Status	Quiet	Noisy	Fast	Moderate	Slow
Accuracy of speech recognition	90.4%	76.1%	79.0%	90.4%	81.7%

Table 2. Test result of one person raising 10 prepared questions to the robot.

Users' Question	Text Recognition	Successful Recognition Rate	Correlation
What's your name?	What's your name	100%	0%
What can you do?	What can you do	100%	0%
Where is Ningbo?	Where is Ningbo	100%	100%
What are interesting places in Ningbo?	What are interesting places in Ningbo?	100%	99.5%
What are the programming language of the computer	What are the programming language of the computer	90%	87%
Where is the capital of U.S.?	Where is the capital of U.S.?	100%	100%
When is the Christmas?	When is the Christmas?	100%	100%
Which country does Santa Claus come from?	Which country does Santa Claus come from?	100%	100%
Who is the author of The Old Man and the Sea?	Who is the author of The Old Man and the Sea?	87.5%	100%
What does it mean by "I am drunk"?	What does it mean by "I am drunk"?	90%	94%

Table 3. Test result of each one raising 10 random questions to the robot.

Performance	Speech Recognition	Correlation
Success rate	0.91	0.95

