ISCA Archive
http://www.isca-speech.org/archive

International Symposium on Chinese Spoken
Language Processing (ISCSLP 2002)
Taipei, Taiwan
August 23-24, 2002

PHONEAGENT: A CONVERSATIONAL INTERFACE FOR TELEPHONE EXCHANGE SYSTEM

Bin SHE , Mingxing XU, Wenhu WU

Center of Speech Technology,

State Key Laboratory of Intelligent Technology and Systems,
Department of Computer Science and Technology,
Tsinghua University, Beijing, 100084
[she, xumx, wuwh]@sp.cs.tsinghua.edu.cn

ABSTRACT

This paper presents a conversational interface of telephone – **PhoneAgent**, that allows users to find the right person they want using natural spoken language. The system integrates telephony technology, Client/Server technology and human language technology, including telephone-based speech recognition, robust language understanding, language generation, dialogue modeling. PhoneAgent consists of three parts: Tele-Controller, Dialogue Manager and Staff Client/Server system. Tele-Controller, which includes a computer with a telephone card connected with a telephone line serving as a call center, monitors the telephone line and processes the input and output of speech data. Dialogue Manager gets speech data from Tele-Controller and recognizes it using speaker independent telephone speech recognition engine, after that the recognition result is processed by the dialogue system which interacts with Staff C/S system to determine whether the person wanted is present or not and generate the proper answer to the user. Staff C/S system is a network system which manages the absence of staff and if a phone call comes for some one who is present, the network system will notify him coming to get the phone call. Now the system can be put to use in an office with not more than one hundred staff.

1. INTRODUCTION

Chinese Speech Telephony technologies based on the ASR engine and series of application develop tools make it easy for speech interface to be integrated into telephony systems. Services of human-computer interaction with speech technology are developing fast in many fields nowadays, for example: weather information query, call center, plane information query, etc.

Telephone-based interactions pose several research challenges. For example, telephone speech is often hard to recognize and understand due to the reduced channel bandwidth and the presence of noise. In addition, telephone interaction demands a high-quality verbal response. Furthermore, near real-time performance is necessary to avoid long delay over the phone which makes users impatient. PhoneAgent system is providing a high task accomplishment ratio which is more than ninety percent using robust speaker independent telephone speech recognition technology and spoken dialogue with no limit of sentence form, and a multi-channel technology is developed to make the interaction with users much easier and more efficient. Utterance verification and rejection in this system make it more stable to use. The near real-time recognition delay, which is less than one second makes the interaction more natural and pleasant.

2. SYSTEM OVERVIEW

To access PhoneAgent, a user calls a phone number specified in our lab. After a connection has been established, PhoneAgent speaks a greeting message. After the greeting, the user is free to engage in a conversation with PhoneAgent. If the person wanted is absent the user could leave a message or ask for another person. In any case, the user can force PhoneAgent to stop and the phone call will be processed manually.

PA:	这里是清华大学语音技术中心，请问您找谁？
User:	我找一下张国亮
PA:	请问您找张国亮还是张欣研？
User:	张国亮
PA:	张国亮现在不在实验室，请问您要留言吗？
User:	不用了，那李净在吗？
PA:	李净暂时不在座位上，请您稍候再拨。请问您还找谁？
User:	那，佘斌在不在？
PA:	请稍候，佘斌马上就来。

Fig.1 Example between PhoneAgentand a user

Fig.1 gives and example of interactions between PhoneAgent and a real user. After the last sentence spoken by PhoneAgent finished, a network notification will be sent to the man whom is wanted on the phone.

PhoneAgent system consists of telephone speech server, speech recognizer, speech generator, semantic analyzer, dialog manager, PhoneAgent (PA) network notification system(including online server and PA client). Fig.2 shows the modules of the system.

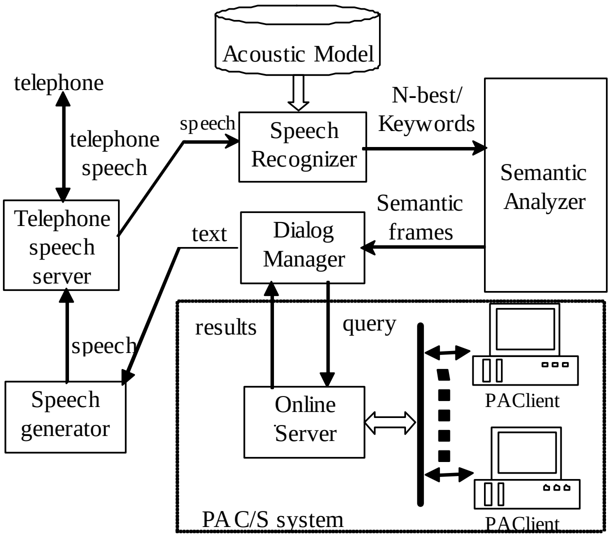

Fig.2 Architecture of PhoneAgent

After the user calls in the phone, the telephone speech server gets 8kHz and 8 bit speech signal data from the telephone line, from which the speech recognizer extracts features. The recognition results are sent to semantic analyzer which generates semantic frames for dialog manager. Response speech is generated according to the text response of the dialog manager after queries to the online server.

Online server manages all PhoneAgent clients' online status. All who want to use the PhoneAgent system should run PA clients on their PCs and PA clients can run and connect to the online server automatically and transmit the telephone notification. The online status of a PA client can be set manually.

3. SYSTEM IMPLEMENTATION

No special devices are needed to access PhoneAgent system. Users can uses telephone or mobile phone to access PhoneAgent. To run PhoneAgent system, we need a PC as a server which has windows series OS and PhoneAgent software. We also need a telephone line, a telephone card and LAN with PA clients installed in each PC that wants to benefit from PhoneAgent system.

3.1 Telephone speech server

The telephone speech server is an interface of PhoneAgent system which monitors all channels of the telephone car. It keeps a status machine for each channel to manage the low level events. And a speech detector is added into the telephone speech server, which will discard the telephone data if it finds no human speech data in it.

3.2 Speech recognizer

The recognizer is based on HTK and has a keyword table of 79 keywords. For acoustic modeling, the current PhoneAgent configuration makes use of context-dependent triphone Intial/Final models. And 13 MFCC with energy were computed for each frame together with their delta coefficients, creating a 42-dimensional feature vector. The recognition will generate 10 best keyword results for the semantic analyzer.

Channel normalization module is added as a front-end to deal with the telephone signal and it can give a better performance - about 5% higher in recognition accuracy.

3.3 Semantic analyzer

PhoneAgent is a domain specific dialog system and we could just fill in the semantic slots with the recognition results. Default value and history stack is introduced to resolve the omission in spoken language dialog which can use the context information of the dialog.

If there is no context, every empty slot will use its default value configured by developer. If the context exists, the empty slot will use the top of its history stack instead of default value. And when topic shifts, the history stack will be cleared.

3.4 Dialog manager

The dialog manager processes with the semantic frames and interacts with the online server to get the query results and generates the text response which will be transferred to speech response by speech generator and played to the end user through telephone line.

Keyword confirmation policy is used in PhoneAgent system which can lead the interaction with the user and reduce the error probability. Also the system will be friendlier.

Focus expectation policy sets the conversational focus expected making use of the context of the interaction between PhoneAgent and the user according to the characteristics of natural spoken language dialogs.

3.5 Online server

Online server manages the status of all users connected to it and responses to the query from dialog manager. It also send telephone notifications to PA clients.

In LAN environment, everyone has his own PC with different IP. With this precondition, PhoneAgent system configures the IP for each user at server side instead of password policy so that the user doesn't need to remember any password to login on the PhoneAgent online server.

Online server keeps data structure for each client, such as name, IP, online status, activity value, etc. Activity value is an attribute for each client to make the system more robust. There is a system timer at the server side and it decreases the activity value of each client connected. Each client sends online messages to server via network and when the server receives those messages, it will set the activity value of that client to max value. If the activity value falls down to zero, the system will kick the client offline. And this kind of design ensures the consistency when network problem occurs or clients have any exception.

When online server receives a query from the dialog manager and the user wanted is online, a notification will be sent to the user through network.

3.6 PA client

Installation of PA client on client PC is needed to take use of PhoneAgent system. Each time the PC starts up, PA client can run automatically and connects to the server. After the connection established, the PA client will hide as a trayicon at the right-bottom of the desktop. When the telephone notification comes from the online server, the PA client will pop a window confirming that the user will go to get the phone call and play sound. If the user has no response for more than a certain period, PhoneAgent system will take the user for absent and deal with the phone call itself.

PA client sends online status messages to server every second to update its status at server side. This can keep the consistency of the online status at server side with the reality.

A mouse detection function is added into PA client to change the user's online status automatically. After the mouse of the user has not been moved for more than a period, the system will change the status of this user to 'Leave', and if it detects that the mouse moves again, the status will be changed back to 'Online'.

4. DISCUSSION

Robustness is the key point in real environment speech recognition applications, especially when we deal with telephone channels with no visible interface. There are some policies running in PhoneAgent system:

1) Noise cancellation and channel normalization front-end can increase the acoustic recognition performance efficiently;

2) Context depended semantic frame can resolve simple omission in natural spoken dialog;

3) Keyword confirming and focus expectation will make the interaction friendlier;

4) Network robustness technologies and other technologies like activity value and mouse detection are needed in real application.