Telephony Based Voice Portal for a University

Łukasz Brocki, Danijel Koržinek, Krzysztof Marasek

Polish-Japanese Institute of Information Technology, Warsaw
{lucas,danijel,kmarasek}@pjwstk.edu.pl

ABSTRACT

This paper describes a telephony-based automatic voice portal implemented in the Polish-Japanese Institute of Information Technology. The system is based on a single server with a Dialogic telephony voice board. It is integrated with the school's existing telephony network. The system uses Automatic Speech Recognition and Speech Synthesis to communicate with users. The dialog features: call routing, bulletin-board style information and individual student information.

1. Introduction

Voice portal described in this paper was developed in order to improve the existing telephone network and provide an additional interface for the systems already used in the school. The school in question is a medium sized university and receives over a hundred phone calls per day at peak times of the year. This voice portal allows automating some of the calls that would normally have to be handled by the school secretaries, mainly call routing and student information.

The telephone network in the university is a standard switched network based around a Central-Office (CO), a device that maintains and routes all the calls within the organization. The CO has many features that are programmed using a special console. It also has several connections with the public telephone network and many internal telephone extensions. Any telephone device within the network can connect to any other by dialling a special 3 digit code. Any device can also route a conversation in progress to a different extension by using the special "flash" key. In order to retain the integrity of the system, it was decided to connect the dialog system to the CO like a normal telephone device. The CO was reprogrammed to route all incoming external calls to one of the extensions occupied by the voice portal.

The voice portal is also integrated with the existing computer system within the university. This includes the news and bulletin-board service, also used on the school's

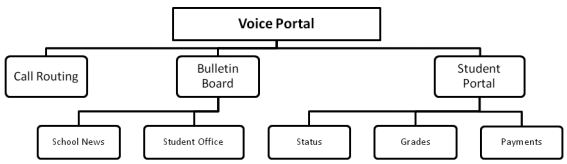

Figure 1. Voice Portal dialog hierarchy.

website, and the student database, which contains such information as grades, payments and various other study related information.

The whole system is thoroughly described on the website [1]. The release of the system was also supported by posters and e-mails describing it to the users.

2. Dialog Design

The dialog is split into three main categories, the main one being call routing. A person that calls for the first time is greeted by a short description that explains how to use the system. To connect to a person or place it is enough to say their name. Then, the system asks for a confirmation and finally it makes the connection. Sometimes, the system will also present a list of names if it is not sure who the user wants to connect to. For example, if the user says a common last name, like "Kowalski", the system will produce a list of all people named Kowalski that work in the school.

The other feature of the system is the news and bulletin-board section. This is basically a list of news items. The system reads the titles of individual items and when the user says the number of the item he wants to hear about, the system reads the whole article.

The last feature is the student portal. This is a password protected area which allows individual students to access information about the progress of their studies. The database is already accessible through a special website, however some information might be necessary even when users don't have access to the Internet. In fact, some students would frequently call the school staff to acquire this information. The dialog is a menu based and controlled by easy to remember voice commands. For example, saying "my grades" will give the user a list of last ten grades. Following that by "previous grades" will generate another list with older grades. The whole menu is protected by a special password that can be activated through the website [1].

3. System Architecture

The voice portal is a distributed system over a network. It consists of 3 basic subsystems, which communicate over a local network: Gateway, TTS (text-to-speech) Server and ASR (automatic-speech-recognition) Server.

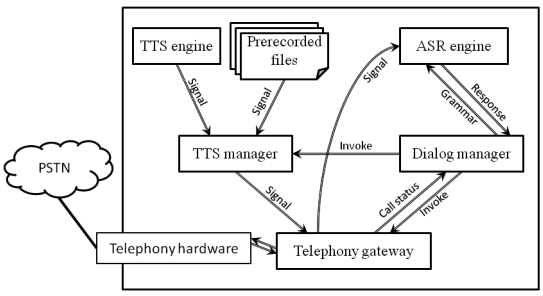

Figure 2. Voice Portal system architecture.

The distributed architecture allows for easy scalability of the system. When used to support 4 telephone lines, all 3 programs can be installed on a single server. In case of 120 lines implementation the system must use 8 computers: 1 working as Gateway, 1 as TTS and 6 working as ASR. In the latter case, each ASR server can conduct 20 dialogs, which gives 120 parallel conversations at any given time. Each of these programs is multi-threaded and fully utilizes the multi-processor and multi-core aspect of modern computers. More information about this system is available here [5].

3.1. Gateway

This is a program that serves as an interface for the specialized telephone voice board used in this implementation. These voice boards have DSP processors to cope with different aspects of noise in telephone lines, most notably echo. Depending on the type, each board can also work with many lines at the same time.

The Gateway server supports both analog and digital telephone cards. For the standard telephone lines it uses an analog card but when more than a few lines need to be maintained, digital telephone cards are recommended.

The main purpose of the Gateway server is to forward the signal from the telephone lines to the ASR server and likewise with the signal from TTS to the telephone lines. It also contains procedures to control the different aspects of the hardware, e.g.: DTMF generation and recognition, flashing, accepting and disconnecting calls.

3.2. ASR Server

This is the core part of the voice portal system. Its purpose is to maintain the dialog through a dialog manager (fig. 2) and recognize speech using the ASR engine developed by the authors [2]. The engine features an acoustic adaptation module that allows re-training of the neural networks. The built in speech detection module dramatically improves performance by disengaging the main recognition subsystem if the user isn't speaking. The architecture of the program allows conducting any number of dialogs on a single server. This is of course limited by the amount of processing power of the computer.

3.3. TTS Server

The purpose of this subsystem is to provide speech responses to the user by utilizing either the speech synthesis engine or prerecorded audio files. Most of the dialog was prerecorded in a studio by a professional speaker, however many of the steps still need to be synthesized, e.g.: news, person names, subject names, etc. These items are also cached on the hard drive so there is no need to synthesize them more than once.

4. User Reactions

By the end of the school term, during the exam term, the voice portal managed over a hundred calls a day on average – often 4 calls at the same time. It was during this time that the system reliability and user reactions were being tested. The breakdown of this is noted in table 1.

It was observed that many employees like to use the portal to connect to other people in school – it keeps them from having to remember or look up internal extension numbers.

Table 1. Breakdown of dialogs during the testing phase. Size is an estimated number of different utterances the system has to recognize. The statistics are counted on a representative sample of 500 dialogs

Dialog type	Size of dialog	% Dialogs
Call routing	~1500	80
News and information	~10	2
Student information	~10	18

Students call to check their latest grades and to check tuition status. A lot of people that call the school for the first time are impressed by the presence of such a system.

Unfortunately, there are also people that don't grasp the technological aspect and simply hang up without saying. Only after calling the second or third time, they say what they need.

5. Conclusions

The voice portal in the Polish-Japanese Institute of Information Technology improved the already existing telephone system. Many employees are pleased with the call routing feature and the portal became a new communication channel between the school and its students.

Voice portals and speech recognition are a novelty on the Polish market and as much as any new technology they face resistance in being accepted by their users. Just like in the authors' previous experiment with the Warsaw Transport Authority [3] the biggest challenge is breaking the ice with new users.

The authors are convinced that such systems could improve the operation of many different organizations. The system described in this paper should help assess the usability and performance of such systems in the future.

Acknowledgments. We would like to thank the management of the Polish-Japanese Institute of Information Technology for their support in accomplishing this project. This research was partially funded by the Luna (IST 033549) project.

BIBLIOGRAPHY

[1] http://speech.pjwstk.edu.pl.

[2] D. Koržinek, Ł. Brocki. 2007. Grammar based automatic speech recognition system for the Polish language. In: Recent Advances in Mechatronics.

[3] D. Koržinek, Ł. Brocki, R. Gubrynowicz, K. Marasek. 2008. Wizard of Oz Experiment for a Telephony-Based City Transport Dialog System. In: Proceedings of the International IIS'08 Conference.

[4] De Mori R. *Spoken Dialogues with Computers*.

[5] http://www.primespeech.pl.