

A SPEECH-OPERATED RAILROAD INFORMATION & RESERVATION SERVICE WITH MULTI-STAGE DIALOGUE

Ho-Hyun Jeon, Chang-Sun Ryu, Jae-In Kim, Myoung-Wan Koo

Spoken Language Research Team, Multimedia Technology Lab.
Korea Telecom

{hhjeon, csryu, jaeinkim, mwkoo}@kt.co.kr

Abstract

Dialog management is very important in telephone-based services. This paper describes KORIS (Korean national Railroad Information & reservation Service) system for access to rail travel information. The system allows to access the timetable by using telephone over the public network. It was developed to replace a conventional service like a ARS (Automatic Response Service) and to support an attendant service with which a user can get information. We use a CHMM (continuous hidden Markov model) for isolate word recognition and multi-stage grammar for dialog management. The grammar and some parameters for recognition depend on the service. It prevents wrong results that are not relevant to each stage. The paper will show the overview of the system and results of field trials.

1. Introduction

The technology of speech recognition has affected recently on our lives because of the progress of hardware and software. It has been applied in wide fields from conventional PSTN network to advanced Internet. It is reasonable that the providers of telecommunication service become interested in the application of the speech recognition because their customers are the subscribers of conventional voice services and have their own wired or wireless terminals.

The technology of speech recognition which can be substituted for a sort of "hands-busy, eyes-busy" task is useful[1]. Think of a following hands-busy task. It takes slightly longer than 5 seconds that a person remembers a 10-digit telephone numbers by heart and then keys in the number to dial while it takes about 25 seconds that a person forgets the number and then finds an appropriate web page with the corporate telephone directory, types in the name of the called party, retrieves the number from the directory, and keys it in with the phone and makes a call [3]. The speech recognition can be substituted for the hands-busy task and can spare time. Besides, the application such as the attendant system of voice recognition [2] may remove confusion caused by people's failure of their remembrance and provide even automatic dialling as well.

Also, RAILTEL based on an isolated speech recognizer was applied to the Italian Railway timetable over PSTN for automatically obtaining information about train departures and arrivals, train services, fare, and so on [7]. Like the former application, RAILTEL, the technology of speech recognition can be applied to improve an inconvenient element of the conventional automatic system. This paper describes an application system which strives to remove inconvenient elements of the conventional system in the South Korea.

At first, the overview of KORIS is described. The configuration and feature of the system is given a full detail of. And then the multistage dialogue is illustrated in chapter 3. Also, I review the eight-month statistical data of field trials to tell some impact for the service. Finally, I conclude the paper as I arrange it.

2. The Overview of KORIS

The office of Korean National Railroad had established ARS in the centre to provide information about the railroad. Making reservation, people must keep the code of start station and end station in mind or look up into the codebook of a code to the stations. If a person often makes trip to the resorting station, he/she may use ARS without difficulty. But, in hasty modern society with a lot of information codes, it is generally difficult that people always keep the code in mind. So, The office of Korean National Railroad wanted to apply the technology of the advanced speech recognition to the conventional ARS(Automatic Response System)of an information centre to lessen customers' difficulty down for customers have to remember the codes of the station names in mind.

The office had come to implement KORIS together with Korea Telecom. The scenario of the new system had come to follow the one of the conventional ARS for the office had not like to give its customers confusion from new scenario Then, Korea Telecom had completed to develop a recognition engine and applied the speech recognition technology to KT-STOCK system[4,5,6], which served information about Korea stock exchange market, and VDS(Voice Dialing System), which dialled automatically to the called party.

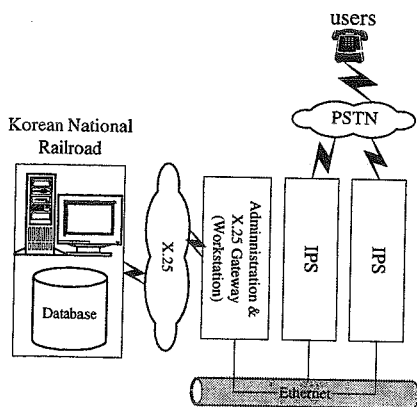


Figure 1. System configuration

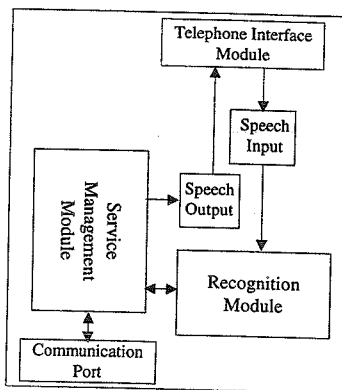


Figure 2. IPS (Information Processing System)

The configuration of the KORIS system is shown in Fig 1 and Fig. 2. Fig. 1 depicts that the system consists of IPS modules and administration module. One IPS may have 4 through 16 recognition ports. The system can be expanded. The packet data from IPS is sent to and received from the database of Korean National Railroad over X.25 network. The packets include the information about reservation of tickets, fares, and so on. The administration module is in charge of the following three roles. First, it takes a role of X.25 gateway and monitors packets between the database and IPS to find errors if they occur. Second, it initializes IPS, monitors the recognition module of IPS, and saves collected voice data from IPS. Lastly, it also includes the automatic generator of pronunciation dictionary and a grammar if there is a need to update them. It is very convenient that an administrator monitor IPS in GUI(Graphical User Interface) environment and she/he access IPS at a remote place.

Fig. 2 shows the block diagram of IPS, which has the recognition module and the service management module. The service management module is in charge of the proceeding of a KORIS

scenario, communication with the database of the Korean National Railroad over X.25 network, collection of statistical data about KORIS, and collection of user's voice data. IPS has RS-232C port, which is connected to the console terminal. The module is also related with a recording guidance module where guidance for the information of the Korean railroad is recorded by the office over PSTN. The recognition module may have at most 4 speech recognition cards and one card has 4 DSPs, which comprises a total of 16 channels per module.

3. Dialog Management

KORIS has 3 service menus on the top of the scenario. The service has an inquiry menu for the seats available, a reservation menu for only membership, and a guidance menu for railroad information. The membership of the railroad can access all the three menus but non-membership cannot access the reservation menu. People access commonly to inquire of the seats available and a fare. Also, they can get information about tours and a guide telephone number to a station through the guidance menu.

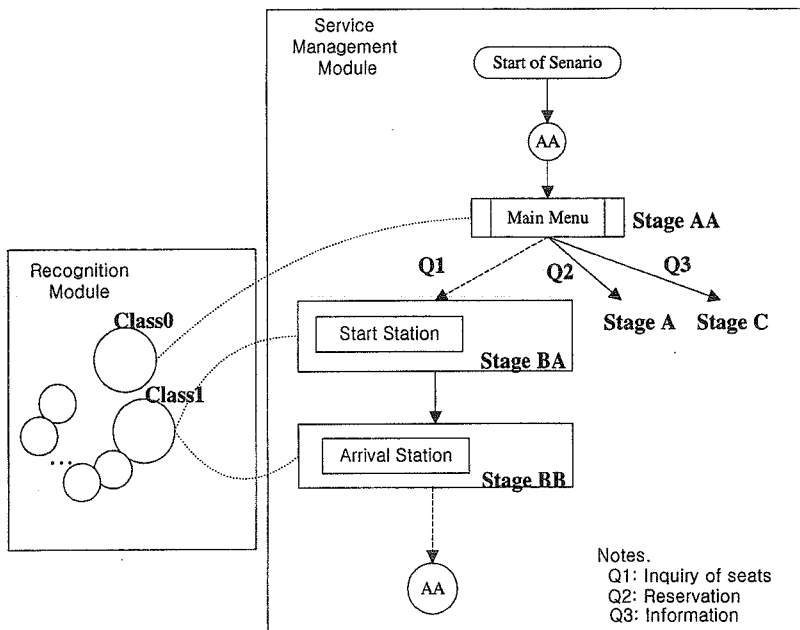


Figure 3. The relationship between classes and multi-stages

Fig. 3 shows concisely the procedure related to the speech recognition according to the inquiry menu for the seats available. The inquiry menu is one of three top menus and is belonged to stage AA. When users come in the stage, they had better speak out one of the candidate words in class1 because the stage is charge of only the words belonged to the class. An initial process of the recognition system is proposed to make an initialization of a multi-stage dialogue. The initial process of recognition program can decide the relevant class to the coming stage when one stage makes a transition to another stage according to the scenario. So, the recognition system has a multistage dialogue in the way of the methodology.

Fig. 3, in the end, shows the relational map between stages and classes. This methodology groups closely connected words belonged to a class so that parameters of the Viterbi beam search algorithm are set on each stage and then the beam search may obtain efficient performance. Laboratory trial of simulation says that 84.60% out of 4,343 attempts succeeds in responding only for 550 station names in a stage about an input of the departure/arrival station name. Another advantage for the methodology prevents users from detracting from the scenario because they cannot get the result unrelated completely with the current stage.

The service has been implemented to dialogue on hierarchical multistage with a user especially when a customer wants to make reservation of a ticket. It asks membership where the customer will leave and arrive, what a class of train the customer wants to use, and what a class of seat the customer wants to use, and so on. After this dialogue of multi stage, a packet of reservation is generated, and then the system sends the packet to the host system of the office. The host retrieves the customer in the reservation database, and then responses the result to the system. If the host permits the customer to make reservation, the recognition system soon requests for ticketing.

The recognition system makes barge-in with the energy-based detection so that users may proceed on the scenario faster when they can adapt themselves at it. And for the same purpose users are proposed to push the special button-the button "*" only when they refuse a response of the system for their input speech. In case of their acceptance, they do not respond and wait for the next stage. Also, the recognition system was developed to holds the processing of a DTMF(Dual Tune Multi Frequency) signal from a telephone prior to the speech recognition to support conventional customers who had used the ARS system.

4. Field Trials

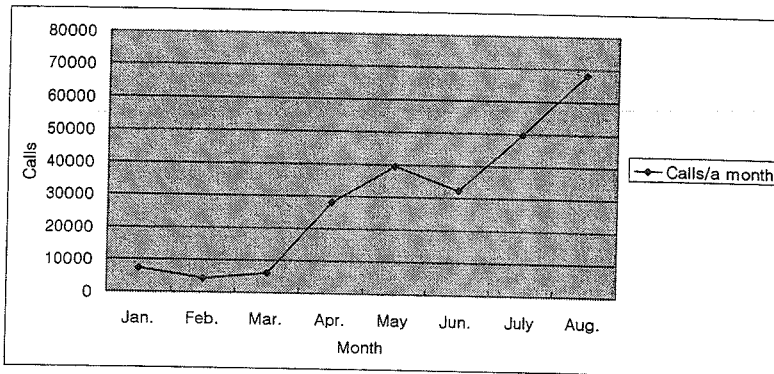


Figure 4. Calls per a month

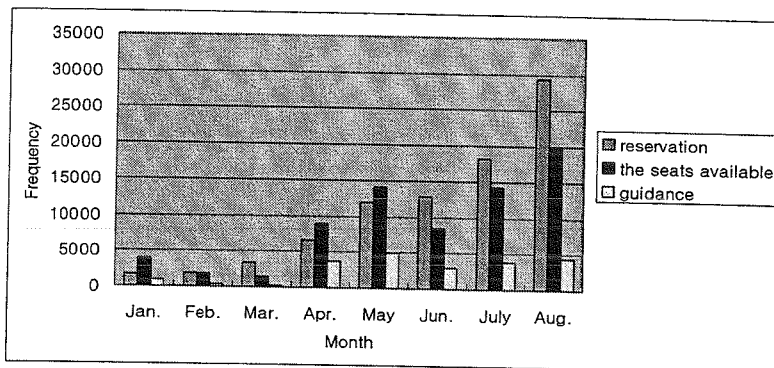


Figure 5. Frequency per a month according to a service

The system of KORIS established in Seoul features that people may access it through the unique service number, 700-1188, throughout the country. They don't have to push a local number. This service can accept wireless subscriber as well as wired subscribers. The service has provided over the telephone network since January 2000.

The graphic data during 8 months are shown in a few following figurations. Fig. 4 shows that users have accessed somewhat higher on April and the usability has continued to increase to October. The April increase was due to the advertisement for membership. The continual increasing till October proves this service to improve user's inconvenience. Fig. 5 shows transition according to each service. During the first May of the service period, the frequency for the inquiry of the seats available was higher than the one for the reservation but it's transition has started to reverse from June. The frequency for guidance has kept low over and over. Fig. 5 explains that KORIS started to be used as the method of reservation by membership of the office of railroad. Fig. 6 compares the coefficient of utilization for the speech recognition with the one for the processing of DTMF signal from a telephone when membership makes reservation. The result represents that the speech recognition cannot be substituted absolutely yet for all the merits and the demerits about the conventional system. That is, keying the known code number of a station name might be even more convenient than speaking and waiting for response. It is essential that membership grows to speak a station name and wait for response when making reservation.

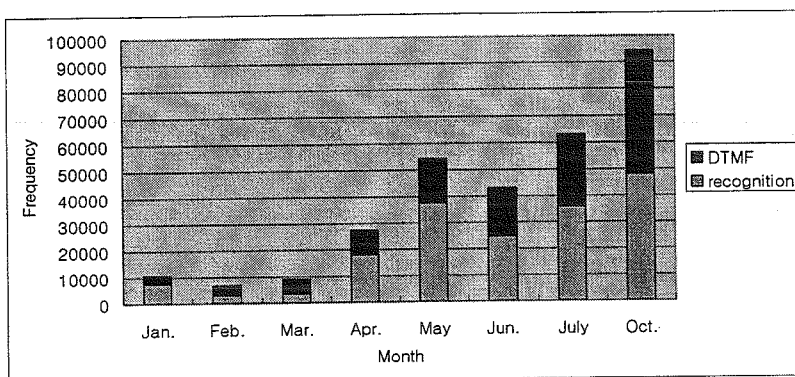


Figure 6. Speech recognition .vs. processing a DTMF signal in station names

5. Conclusion

KORIS had been developed to remove the codebook of the station names in the conventional ARS and then provide convenience to users. The multistage dialogue recognition system of KORIS features that it recognizes a group of relevant words to a transitional stage according to the scenario so that it can get high performance. Another advantage for the methodology of the dialogue prevents users from detracting from the scenario. Especially, the service can recognize user's speech in barge-in, so that users may proceed on the scenario faster when they adapt themselves at it. Also, the automatic generator triggered on the GUI supports the easy and convenient editing which can add or delete a station name.

The recognition system established in Seoul, the capital of The South Korea, features that people may access the service through the unique service number, 700-1188, throughout the country. This service accepts wireless telephone as well as wire telephone subscribers.

The graph of statistical data been collected for eight months explains that KORIS will give increasingly convenience for users while the speech recognition cannot be substituted absolutely yet for all the merits and the demerits about the conventional system.

As it used to be, the system needs to improve forward a conversational style system like a user asks really a guider of making reservation over the telephone. We have a plan to apply the speech recognition of numbers to KORIS system later.

References:

- [1] C. S. Ramalingam, et. al., "Speake-Dependent Name Dialing In A Car Environment With Out-of-vocabulary Rejection," proc. Of ICASSP99, Paper No. 1780, May 15-19, 1999.
- [2] K. Koumpis, et. al., "Design And Implementation Of Auto Attendent System For The T.U.C. Campus Using Speech Recognition," Proc. Of ICASSP98, vol.2, pp.845, May 12-15, 1998.
- [3] K. Laurila, et. al., "Name Dialing-How Useful Is It?," Proc. Of ICASSP2000, VI-3731.
- [4] M. W. Koo, et. al., "KT-STOCK: A Speaker-independent, Large-vocabulary Speech Recognition System Over The Telephone," in Proc. 1994 IEEE Int. Conf. Spoken Language Processing, Sep. 1994.
- [5] M. W. Koo, et. al., "A Stock Information System Over The Telephone Network," Proc. Of Int. Conf. On Signal Processing Applications & Technology, pp. 2039-2043, Boston, Oct. 1995
- [6] S. G. Park, et. al., " Development of Multichannel Information Retrieval System Using Speech Recognition," Proceedings of ICSP'97, August 26-28, Seoul, Korea.
- [7] R. Billi, et. al., "Field Trial Evaluations of Two Different Information Inquiry Systems," pp. 129-134, in Proceedings of IVTTA, 1996.