Development and Operation of Speech-Oriented Information Guidance Systems, Kita-chan and Kita-robo

Hiromichi Kawanami*, Shota Takeuchi*, Rafael Torres*, Hiroshi Saruwatari* and Kiyohiro Shikano*
* Graduate School of Information Science, Nara Institute of Science and Technology, Japan
E-mail: {kawanami, shota-t, rafael-t, sawatari, shikano}@is.naist.jp

***Abstract*—The authors have been operating a real-environment speech-oriented information guidance system named "Takemaru-kun" daily at the Ikoma City North Community Center. The system introduces an example-based one-question-one-answer strategy, noise rejection and an adult versus child user discrimination mechanism. As a result, it realizes robust response generation since its initial operation in Nov. 2002. Following the success of "Takemaru-kun", two additional systems were developed and installed in a railway station, side by side, since Apr. 2006. Although the three systems share the same core software, "Takemaru-kun" and "Kita-chan" employ CG agents and "Kita-robo" has a robot-like body casing. At the railway station, users can talk to the system of their preference. In this paper, we introduce the two later information guidance systems and report their initial operational results including an analysis of user utterances according to age groups. An analytical result shows that children's utterances to the robot are twice as many as those to the CG agent. In addition to that, it is observed that children do not make difference in question topics between the systems, but adults do.**

I. Introduction

Since speech is one of the most familiar media of communication, various researches on developing practical spoken dialogue systems have been conducted in recent years. Automatic speech recognition (ASR) has been widely applied to dictation, Voice Search, and car navigation, to name a few. In this paper, we describe our real environment speech-oriented information guidance systems, *Takemaru-kun, Kita-chan* and *Kita-robo*. They are developed to realize a natural speech interface using ASR.

Takemaru-kun (Fig. 1), which was installed on Ikoma City North Community Center, is our first real-environment speech-oriented information guidance system and has been in operation since Nov. 2002 [1]. The task domain is not given beforehand. It is an example-based question answering (QA) system that is flexible to respond to user's questions on demand. An answer to a user's question is selected by referring to the question and answer database (QADB). *Takemaru-kun* employs a CG agent, which has been welcomed by the users, especially by children. However, a robot-style agent is supposed to be more attractive for children. Some researches focus on the effect of an agent form, that is to say, an on-screen agent and a robotic agent [2][3].

To investigate the effect of an interface, the authors developed two additional speech-oriented information guidance systems with different agent forms. One is named *Kita-chan* and has a conventional CG agent. The other is *Kita-robo*, with a robot-like casing. The two systems are implemented together, beside the gate of a railway station (Fig. 2), since the end of Mar. 2006 [4]. In this paper, we introduce the two information guidance systems and report their initial operational results including an analysis of user utterances according to age groups.

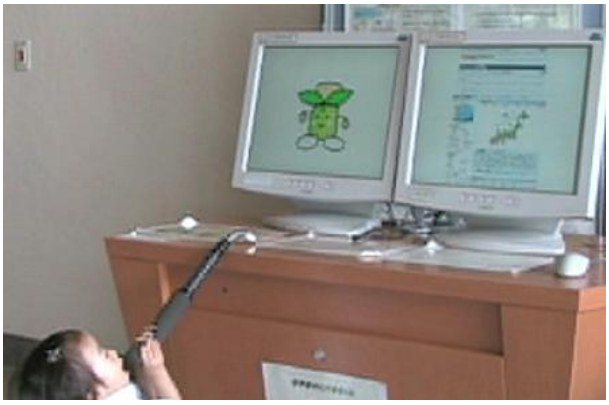


Fig. 1. Spoken dialogue system *Takemaru-kun* and a child user

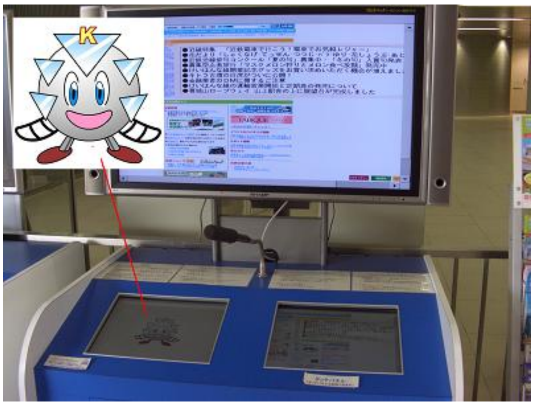
II. Real-environment Information Guidance Systems

A. Systems overview

As we mentioned above, *Kita-chan* and *Kita-robo* are based on the preceding system *Takemaru-kun*. *Takemaru-kun* system is implemented at the entrance hall of the community center. The new systems are set in *Gakken Kita-Ikoma* railway station since Mar. 2006. Although the systems share the same core system, the agent characters, service contents and input/output devices are different.

Environmental noise is different between the community center and the station. The average noise levels are 50[dBA] in the community center and 60[dBA] in the station. In the station, noises from car engines, horns, sound of rain, wind, etc. are observed.

As mentioned before, *Kita-chan* employs a conventional CG agent like *Takemaru-kun*, and *Kita-robo* uses a robot-like casing. Fig. 3 illustrates the difference in the input/output devices



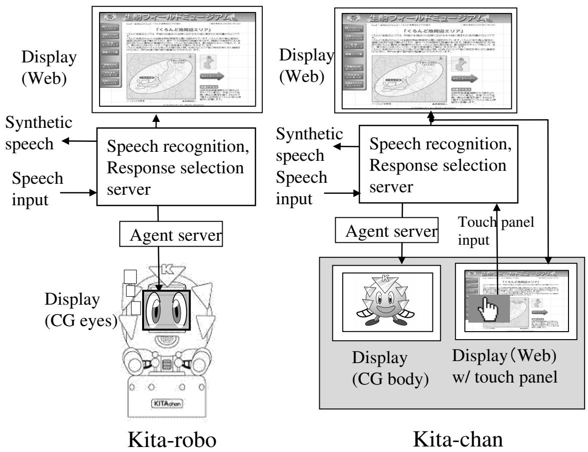


Fig. 3. Input/output devices of *Kita-robo* and *Kita-chan*

used in *Kita-chan* and *Kita-robo*. When a user utterance is input, the systems answer the required information using CG animation, synthetic speech and an Internet web browser. A touch panel display is also present only in *Kita-chan*. CG agent animation is used in *Kita-chan*, and CG eyes animation is used in *Kita-robo*.

The task domains of *Kita-chan* and *Kita-robo* are basically the same. They provide information about the station facilities, train lines, local information, city information, sightseeing, CG agent profile, and so on. In addition, users can activate an Internet search engine by saying "*KeNsaku Kaishi. (Start Voice Search mode).*"

The difference in the service contents between the systems is owing to an additional input device besides speech. As *Kita-chan* has a touch panel, the system answers using interactive web services such as *Yahoo transit* for some questions, but *Kita-robo* just displays static information using web pages.

B. Speech recognition and response generation

Kita-chan and *Kita-robo*'s dialogue management systems inherit the framework of *Takemaru-kun*: a one-question-one-answer dialogue strategy and example-based response generation. They also inherit speech/noise discrimination and adult/child classification of input speech. Speech/noise discrimination is conducted using five GMMs (Gaussian Mixture Models): GMM for adult speech, child speech, laugh sound, cough sound and other noises. In the preliminary test, typical noise sounds in the station were successfully rejected using the GMMs used in *Takemaru-kun*. Therefore the same GMMs were also introduced to the two systems.

The child or adult classification is conducted using acoustic likelihoods. As shown in Fig. 4, four decoding processes are run in parallel as a combination of acoustic models (adult or child) and language models (N-gram and network grammar). As a result, four recognition candidates are output with their acoustic likelihoods. If the highest likelihood is obtained from adult's models the input is classified as adult speech, and vice versa for child speech.

In the response generation module, an example-based approach is employed. A rule-based approach is subsidiarily introduced only when the likelihood of a decoding result using network grammars is the highest. In the example-based response generation, 5-best decoding results are used. The most similar example question to the decoding results is selected from the Question Answer Database (QADB). A similarity is calculated using shared word frequencies between the input and example questions [5][6]. Two kinds of QADBs are implemented for response generation. One QADB is used for adult users and the other is for child users.

III. User Speech Database

Like *Takemaru-kun*, every system input has been collected since the operation started. The first ten months' (Apr. 2006 to Jan. 2007) inputs were manually transcribed with noise tags. The total number of inputs to *Kita-chan* and *Kita-robo* are 64,214 and 94,237 respectively. Noise inputs are included in the counts.

Age group and gender information were also labeled by hearing. Age group consists of 5 categories: pre-school child, lower-grade student, higher-grade student, adult, elderly. Gender labels are male and female. Fig. 5 illustrates the number of user utterances according to age group. The figure shows that the main age group for *Kita-chan* is adult and that for *Kita-robo* is lower-grade student.

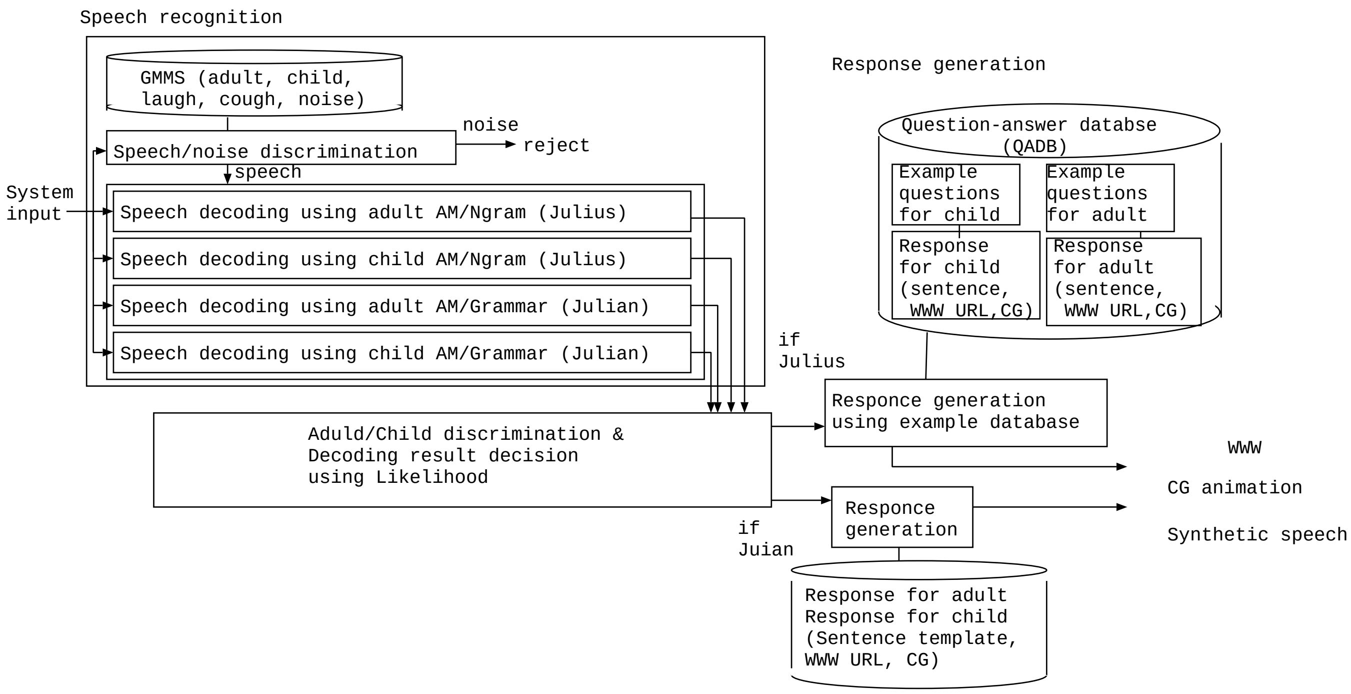


Fig. 4. Process flow of speech recognition and response generation

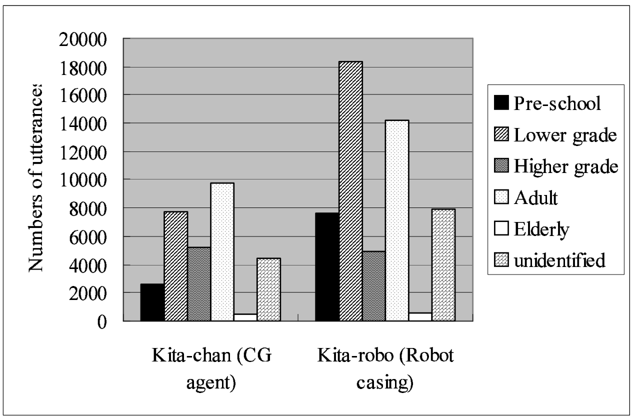


Fig. 5. User utterances classified by age group (ten months)

Topics of user questions were also categorized manually. The categories are (a) Greetings and chat, (b) Agent profile, (c) Station facilities, (d) Local information (e.g. post office, restaurants), (e) Sightseeing, (f) Ikoma City, (g) General (e.g. news, weather forecast, Web search), (h) Train lines, (i) Maps, (j) Web browser navigation. These category labels are exclusively attached to each user utterance.

IV. Topic Category Analysis

A topic category analysis of utterances to *Kita-chan* and *Kita-robo* was conducted concerning two age groups: adult and child.

Table 1 shows the number of utterances for each topic category. Here, age groups from pre-school child to higher-grade student are treated as child, and the rest as adult. Figs. 6 and 7 are pie graphs of the data from adults and children, of *Kita-chan* (CG agent) and *Kita-robo* (robot-like casing), classified by topics.

As shown in Fig. 6, greetings and questions about the agents (e.g. “How old are you?”, “Where were you born?”) account for two-thirds of the children’s utterances in both systems. The utterances in these topics can be considered as utterances for entertainment. Most of the rest consists of questions on Station facilities, General information and Maps. As Fig. 5 illustrated, children’s utterances to *Kita-robo* are twice as many as to *Kita-chan*; however, the rates of each topic in the two pie graphs are similar between the systems. On the other hand, in adults’ utterances, the rates of topics differ between systems. The rates of greetings and questions about the agents are only 42% on *Kita-chan* and 57% on *Kita-robo* in total, and the rest of utterances belong to various topics. It is interesting that a similarity between the pie graph of adults’ utterances to *Kita-robo* and that of children’s utterances can be observed.

These figures indicate that the difference in the agents form does not affect children from the viewpoint of which topics they would speak about. On the contrary, adult users tended to use *Kita-robo* for entertainment and *Kita-chan* for obtaining serious information.

V. Conclusions

The real-environment information guidance systems, *Kita-chan* with a CG agent and *Kita-robo* with a robot-like casing, were introduced. Comparative analysis of ten months’ user utterances was conducted from the viewpoint of age groups

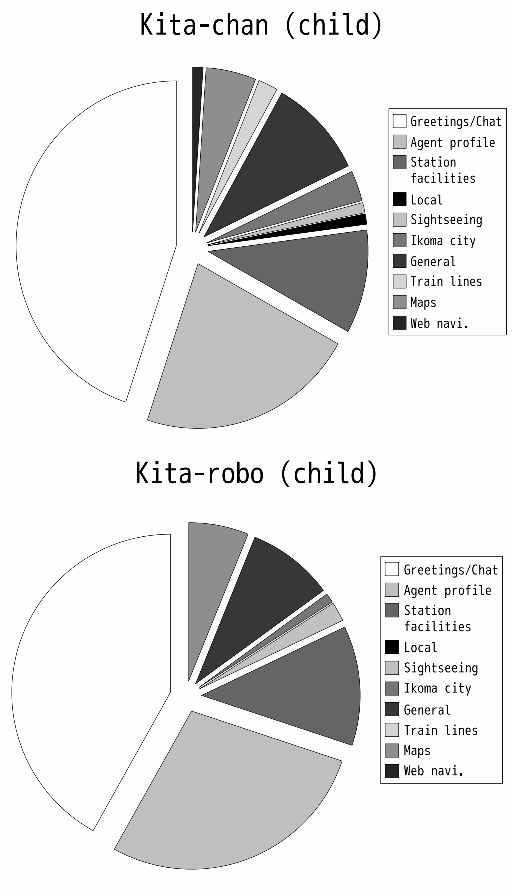


Fig. 6. Rates of topics in children's utterances

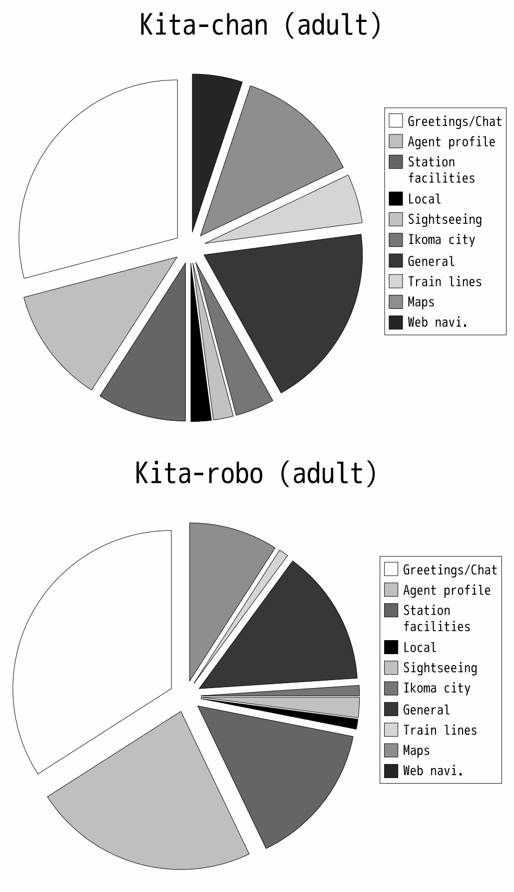


Fig. 7. Rates of topics in adults' utterances

TABLE I
TOPIC CLASSIFICATION OF USER UTTERANCES

	Child		Adult	
	Kita-chan	Kita-robo	Kita-chan	Kita-robo
Greetings and Chat	3172	6075	1473	2414
Agent profile	1569	3898	635	1706
Station facilities	717	1688	479	1123
Local info	44	15	101	40
Sightseeing	74	219	98	165
Ikoma City	228	135	217	93
General info	729	1225	996	1039
Train lines	112	66	284	90
Maps	377	824	687	691
Web navi.	87	15	239	19
Total	7109	14160	5209	7360

and topic categories. The utterances to *Kita-robo* were twice as many as to *Kita-chan*. It was also observed that two-thirds of children's utterances consisted of greetings and questions about the agents: children use both systems for entertainment without consciousness about difference on topics. Adult users tended to chose systems by question topics. They used mainly *Kita-robo* for entertainment and *Kita-chan* for obtaining serious information. These cues will be helpful to design new spoken dialogue systems.

Further detailed analysis on the agent forms and interfaces will be conducted. Analytical results will clarify how the system agents affect on keeping users' interest.

REFERENCES

[1] R. Nisimura, et al., "Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability," *Proc. ICASSP2004*, vol.1, pp. 433–436, 2004.

[2] A. Powers, et al., "Comparing a Computer Agent with a Humanoid Robot," *Proc. 2nd ACM/IEEE International Conference on Human-Robot Interaction*, pp.145–152, 2007.

[3] T. Komatsu, et al., "Can Users React Toward an On-screen Agent as if They are Reacting Toward a Robotic Agent?," *Proc. 4th ACM/IEEE International Conference on Human-Robot Interaction*, pp.217–218, 2009.

[4] H. Kawanami, et al., "Development and Operational Result of Real Environment Speech-oriented Guidance Systems, *Kita-robo* and *Kita-chan*," *Proc. Oriental-COCOSDA2007*, pp.132–136, 2007.

[5] H. M. Meng, et al., "Natural Language Response Generation in Mixed-Initiative Dialogs using Task Goals and Dialog Acts", *Proc. INTERSPEECH2003*, pp.1689–1692, 2003.

[6] H. Murao, et al., "Example-based Spoken Dialogue System with Online Example Augmentation," *Proc. INTERSPEECH2004*,pp.3073–3076, 2004.