Tutoring in a Spoken Language Dialogue System

JAAKKO HAKULINEN, MARKKU TURUNEN, and KARI-JOUKO RÄIHÄ
Tampere Unit for Computer-Human Interaction
Department of Computer Sciences
33014 University of Tampere, Finland
Firstname.lastname@cs.uta.fi

Abstract

We have developed interactive multimodal software tutors to teach users how to use a spoken dialogue timetable by guiding users and monitoring their interaction. They feature a visual representation of the spoken dialogue to support error recognition and recovery and thus helping the users to learn the required interaction style. Two different versions of tutoring were compared to a static web manual in a between-subjects experiment (N=27).

1 Introduction

The challenges of designing spoken dialogue systems are well known, as are the usual solutions. How do the users know the functionality provided by a speech-based system? How do they know when to speak and what to say to the system? A well designed speech interface supports the users' natural way of speaking. However, in practice the interface must also guide users to speak in a way that the system is able to understand. Implicit and explicit prompts, hints, and tapering embed the guidance in the spoken interaction. (Yankelovich, 1996) When a system is used repeatedly, it is plausible that the users are willing to invest some effort into fully learning service.

How, then, are speech-based systems introduced to new users? When speech is an additional modality e.g. in the case of voice control systems in automobiles, the speech-based features can be described in the owner's manual or users can discover the voice control possibilities through the graphical part of the interface. Unimodal, telephone-based spoken dialogue systems need some auxiliary material to introduce them to the users. In addition to an introduction to the service, users are often provided with some instructions on how to use the system. Such a web-based tutorial can improve the user experience and users' perception of the system (Kamm, Litman, and Walker, 1998).

Another approach to introducing new applications to users is software tutoring. This is popular with graphical interfaces, particularly in video games, but it has been almost neglected in the case of speech-based applications. However, the tutorial type guidance can be embedded into a dialogue system, e.g., as a specific guided mode, which can make the system more transparent to users and thus help them, for instance, in knowing how to correct errors (Karsenty and Botherel, 2005). This kind of guidance can be extended by implementing a software tutor, a separate dialogue partner, which not only guides users but also monitors their interaction and makes sure that the users indeed learn to use the system. We have implemented such a tutor and found it reduced the amount of problems users have during the learning period (Hakulinen, Turunen, and Räihä, 2006).

Here we follow-up our previous work on unimodal tutoring by studying graphical tutoring in speech interface. The visual presentation can overcome the transient and linear nature of speech and its low output rate. The multimedia tutors are connected to the spoken dialogue system so that a user can try out the system under the supervision of the tutor. Different tutor concepts were developed (Hakulinen, Turunen, and Salonen, 2005) and two most promising ones were chosen for an experiment. The tutors introduce the spoken dialogue system to users, guide them through an elementary scenario, monitor users' interaction with a spoken

Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, pages 239–242,
Antwerp, September 2007. ©2007 Association for Computational Linguistics

dialogue system and provide guidance as necessary, for example, after recognition rejections.

We collected data on users' interaction with the tutor and the dialogue system and users' attitudes towards the guidance materials and the system. The data did not show significant differences in the task completion rates, but the most troublesome interactions occurred in the web guidance condition. The software tutor with more interaction possibilities was ranked highest in the subjective evaluations, while the other tutor was ranked the worst among the three conditions. Thus, the multimedia tutor can help in learning to interact with a spoken dialogue system, but only when designed properly. The graphical form used in the most interactive guidance helps users in understanding the functionality of the spoken dialogue system. The results point out the importance of constructing the guidance material in a manner that closely corresponds to the interaction model of the system: the interface is essentially a form-filling dialogue, and the highest ranked tutor is based on a graphical version of the form.

2 Guidance Materials

The tutors are graphical software applications run on a personal computer and they communicate with the spoken dialogue application running on a server. A web manual has been constructed based on the tutors by removing all interactivity and arranging the information into a static document. All material is in Finnish, figures and examples have been translated for the paper.

The spoken language dialogue system that the tutors guide users on is called Busman. It is a research prototype of a telephone-based service for Tampere area public transport timetables (Turunen et al. 2005). Typical utterances understood by the system include "Which line runs from University Hospital to the city center" and "When after six pm does a bus depart from Hervanta to university". The system uses form-based dialogue management. Implicit confirmations are used extensively and mostly the interaction is user initiative. System initiative prompts are used for obtaining missing information and after repeated error situations. A short and a rather exhaustive spoken help messages can be heard by giving respective commands.

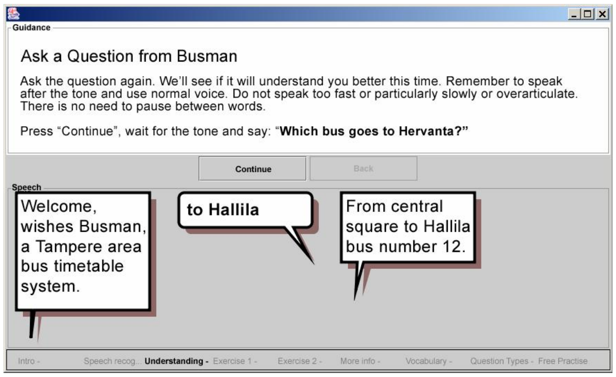
The system uses the Finnish language ASR (Philips SDK with unisex Finnish acoustic models, about 1500 words per grammar) and TTS(Mikropuhe by Timehouse). The system does not support barge-in but telephone keypad can be used to interrupt the system.

2.1 Tutor Design

The goal of the tutors is to introduce the Busman system to new users and teach them how to interaction with it. In five to ten minutes, users will learn the functionality of the system and use it by following the instructions given by the tutor.

The tutors were presented to users as application windows as can be seen in Figures 1 and 2. The only aural component in the tutors is a notification sound that directs users' attention from the application context to tutoring when necessary.

The snapshot of Balloon tutor during the hands-on exercise part can be seen in Figure 1.



Ask a Question from Busman

Ask the question again. We'll see if it will understand you better this time. Remember to speak after the tone and use normal voice. Do not speak too fast or particularly slowly or overarticulate. There is no need to pause between words.

Press "Continue", wait for the tone and say: **"Which bus goes to Hervanta?"**

Welcome, wishes Busman, a Tampere area bus timetable system.

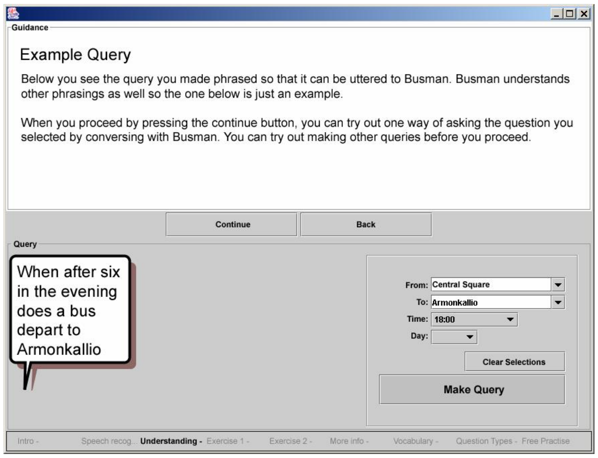
to Hallila

From central square to Hallila bus number 12.

Fig. 1: A screenshot of the Balloon tutor.

The Form tutor includes all the functionality of the Balloon tutor. And a form consisting of graphical user interface components, which users can use to create queries that can be asked from the Busman system. The GUI form can be seen as a visual representation of the timetable system. The benefit of a graphical form is based on a finding by Terken and teRiele (2001) that a multimodal interface with a graphical query interface provided a mental model that can be useful with a speech only interface. The Form tutor is shown in Figure 2.

Guidance in both tutors is organized similarly into six segments, each consisting of one text screen. In addition, there is a hands-on exercise in the middle of the tutoring where users try out Busman under the supervision of the tutor. This part consists of calling Busman and making three queries. In the end, there is free experimentation while the tutor is still active. The last text segment before the free experimentation is a summary.



Example Query

Below you see the query you made phrased so that it can be uttered to Busman. Busman understands other phrasings as well so the one below is just an example.

When you proceed by pressing the continue button, you can try out one way of asking the question you selected by conversing with Busman. You can try out making other queries before you proceed.

When after six in the evening does a bus depart to Armonkallio

From: Central Square
To: Armonkallio
Time: 18:00
Day:

Fig. 2: A screenshot of the Form tutor.

Speech balloons are used to visualize spoken dialogue, i.e., speech recognition results and system outputs, both during the hands-on exercise and the free experimentation. They are also used to display an example dialogue before the exercise. The balloons use bold face font to emphasize keywords in user utterances. Furthermore, the balloons provide a short dialogue history.

The tutors guide the users step by step during the hands-on exercise. The users are told exactly what to say when they call the timetable system for the first time. The tutors monitor the speech recognition results for errors and by comparing ASR results to the requested input, the tutors can spot errors with certainty but not deduct their reason. They do not guess but provide guidance on how to remedy the situation. If the recognition results do not match the required input closely enough, help is given, and the user is asked to try again maximum three times, simplifying the requested input if some information has already been given successfully. The help provided includes instructions on how to speak, such as to use normal voice and talk after a tone. By pointing out errors and providing relevant guidance, the tutors can help users in learning to detect, diagnose, and correct errors.

In addition to the two tutors, a web based version of the same material was created. It contains the same texts and graphics as the tutors as far as possible.

3 Experiment

There were three conditions (named web, balloon and form), one for each guidance material, with 9 participants each. Age of participants ranged from 16 to 41 years with an average of 26. Ten of them were male and 16 female. Most of them had never used spoken dialogue systems and the remaining had had random usage. Participant’s computer using skill ranged from inexperienced user to active hobbyist, most being common users. There were no significant differences on background variables between the conditions. The participants received a movie ticket for their participation. They were randomly assigned to the conditions.

The test consisted of a 15 minute learning period with the guidance and a 15 minute period for working with a set of 11 tasks without the guidance material. In the end, participants filled in two questionnaires where the timetable system and the guidance material were evaluated.

A SASSI questionnaire (Hone and Graham 2000) was used to gather opinions on the Busman timetable system. A set of questions developed by Hassenzahl et al. (2000) was used to gather opinions on the guidance. Both used seven-item Likert-scale questions and an additional field for open comments. The guidance questionnaire also included scales on the length, amount, and consistency of guidance. The questions were in Finnish.

In the tasks the participants were asked to find a bus line number for a given route and a departure time that was near a given time.

3.1 Results

Task completion rates were similar in all conditions. The telephone calls reveal a wider variety of error rates in the Web condition. Questionnaires and general observations made during the experiments raise the Form tutor as the most highly ranked guidance type and provide some insights into differences between different kinds of users.

Interaction with the System Users’ interaction under the guidance of tutors seems to be more consistent while some users of a static manual do just fine and others have serious problems. While there were no statistically significant differences in the error rates between the conditions, the variances of utterance level error rates (i.e., percentage of utterances that did not result in correct system response) between the three conditions were significantly different (Bartlett test of homogeneity of variances, df = 2, p < 0.05). The Web condition had the highest variance in error rates while the Balloon condition had the lowest. When the training part, i.e., the direct effect of the condition is removed, and only the interaction during the tasks is considered, the error rate distributions become more similar.

Questionnaires The guidance evaluation questionnaire resulted in different overall evaluations for the guidance materials. The differences are highly significant (Friedman rank sum test (of evaluation medians), df = 2, p < 0.001). Rank sums (higher value – better evaluation) were 55.5 for the Web condition, 39.5 for the Balloon condition, and 73.0 for the Form condition. There were no statistically significant differences between the conditions within single guidance evaluation questions.

There was no significant difference between the conditions on the SASSI evaluation of the Busman system. However, participants' backgrounds correlate with some evaluations. Computer skills is a variable that highly significantly correlated (Pearson's product-moment correlation df = 25, p < 0.01) with answers to five questions. In all cases more experienced computer users considered the timetable system worse, i.e., less pleasant and more irritating. Speech user interface experience correlates also with computer skills (Pearson's product-moment correlation, df = 25, p < 0.05). However, computer skills did not correlate with error levels or task completion rates. Furthermore, the correlations of computer skills were only with system evaluations. There was no significant correlation with the guidance evaluations, which suggests that the tutors, while not equally necessary to, were equally accepted by the different users.

In guidance questions age correlated (Pearson's product-moment correlation, df = 25, p < 0.001) negatively with answers to the question "Guidance was too long", i.e., younger participants considered the guidance too long more often than older ones.

4 Discussion

In this study, we compared different guidance materials to teach to use of a spoken dialogue system. The results indicate that interactive tutoring helps especially those people, who would have most problems learning the use with static guidance materials. While some users can learn to use a system just fine with just a static manual or even without any guidance, others have many problems in learning the style of interaction required in human-computer spoken dialogue. Unlike static guidance, tutors were able to take care of all users. It is worth mentioning, that especially those, who felt more insecure on using the system, reported that they felt comfortable when they received support from the tutor in the beginning. Tutoring can support users who could not learn the system otherwise, but not all users should be forced to use one.

5 References

Jaakko Hakulinen, Markku Turunen, and Kari-Jouko Räihä 2006. Evaluation of Software Tutoring for a Speech Interface. *International Journal of Speech Technology*, 8, 3, 283-293.

Jaakko Hakulinen, Markku Turunen, and Esa-Pekka Salonen. 2005. Software Tutors for Dialogue Systems. *Proceedings of Text, Speech and Dialogue, LNAI 3658*, Springer, 412-419.

Marc Hassenzahl, Axel Platz, Michael Burmester, and Katrin Lehner, 2000. Hedonic And Ergonomic Quality Aspects Determine a Software's Appeal. *Proceedings of CHI2000*, ACM Press, 201-208.

Kate Hone, and Robert Graham, 2000. Towards a Tool For The Subjective Assessment of Speech System Interfaces (SASSI), *Natural Language Engineering*, 6, 3 & 4, September 2000.

Candace Kamm, Diane Litman, and Marilyn Walker, 1998. From Novice to Expert: The Effect of Tutorials on User Expertise with Spoken Dialogue Systems. *Proceedings ICSLP*, ASSTA, 1211-1214.

Laurent Karsenty, and Valérie Botherel, 2005. Transparency Strategies to Help Users Handle System Errors. *Speech Communication*, 45, Pp. 305–324.

Jacques Terken, and Saskia te Riele, 2001. Supporting the Construction of a User Model in Speech-Only Interfaces by Adding Multi-Modality. *Proceedings of Eurospeech 2001 Scandinavia*, ISCA, 2177-2180.

Markku Turunen, Jaakko Hakulinen, Esa-Pekka Salonen, Anssi Kainulainen, and Leena Helin, 2005. Spoken and Multimodal Bus Timetable Systems: Design, Development and Evaluation. *Proceedings of SPECOM 2005*, 389-392.

Nicole Yankelovich. 1998. How Do Users Know What To Say? *Interactions*, 3, 6, 32-43.