Keywords for a Universal Speech Interface

Stefanie Shriver and Roni Rosenfeld
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 USA
+1 412 268 8669
{sshriver, roni}@cs.cmu.edu

ABSTRACT
In this paper, we describe an internet survey conducted to help choose keywords for a universal speech interface. We present the background of and motivation for this study, and discuss its results and implications for our project.

Keywords
Spoken dialog systems, speech, user interface, standardization, user survey.

INTRODUCTION
The Universal Speech Interface (USI) project is an attempt to create a standardized interaction style for speech communication between humans and simple machines. Our belief is that by implementing and promoting a standard interface, speech applications can become more usable and desirable tools. The interface that we have been developing in this project comprises a set of keywords and interaction guidelines that users learn in order to enable them to explore and use any application that is designed using the USI standard.

One of our main design principles for the USI interface is that it should be more restricted than natural, conversational language (although at the same time it should be less restricted than typical application-specific command-and-control interfaces). The proposed benefits of this design are

- Less complex grammars and vocabularies can be used, promoting lower speech recognition error rates;

- The overall system is less complex than for natural language interfaces; this allows the system to be used in small devices and, with standardization, supports the creation of application-generating toolkits;

- Using a specific speaking style encourages the user to view the system as a tool, which should make the user less likely to overestimate its capabilities.

Other proposed advantages to speech interface standardization include

- The user will know in advance what keywords and syntax work with the system;

- The user will know how to discover the functionality and limitations of the system;

- The skills learned on one application will be transferable to others;

- The user will form correct expectations about system output, which can help make navigation and error correction more transparent and therefore easier.

Initial user studies have been done on a current version of the USI system, and a description of this interface and preliminary results can be found in [2]. We are planning to conduct larger-scale, intensive user studies to test several of the proposed advantages of the USI approach, both on its own and as compared to natural language speech interfaces.

MOTIVATION
Before proceeding with larger studies however, we wanted feedback on the keywords currently used in the interface. Our original keyword selections were based on our own assumptions and intuitions about what made memorable, acoustically distinct, relatively non-technical keywords. Hoping to either validate our choices or to find better ones, we decided to conduct a survey to see how our keyword choices fared and whether users had other suggestions that were more desirable than our own.

Related Work
In this research, we were strongly influenced by the studies presented in [1, 3], the goals of which, like ours, were to determine effective keywords for use in voice applications. The USI system makes use of a set of eight core keywords, and we chose to conduct our own survey as neither of the previous studies covered this exact set of keywords. Furthermore, the survey described in [3] elicited keywords from users but did not ask users to rate any keywords proposed by the researchers; the study in [1] used both approaches, but used a very small pool of subjects.

SURVEY DESIGN
Our survey was conducted in two phases. In Phase I, we created an internet questionnaire with two parts, similar to that used in [3]. In the first part, subjects were presented with 17 contexts and were asked to provide a single word or short phrase that they felt would most successfully perform the desired command in that situation. The 17 contexts

Situation	Top-rated keywords	Wtd score
1. The system just said something, but you were distracted or otherwise didn't hear it very well. You would like the system to play the last thing it said again.	repeat	1.12
	repeat that	1.39
	what?	2.12
2. Your interaction with the system isn't going very well. You would like the system to forget everything you've said so far so that you can try your request again.	**start over**	0.76
	restart	1.45
3. You would like to get rid of the last thing you said (for instance, because you mis-spoke, or changed your mind, or the system misunderstood you).	**undo**	1.54
	back	1.98
4. You're in the middle of interacting with the system and want to find out what you can say at that point.	help	1.22
	options	1.24
5. Say the system has read you the first set of items from a longer list & you want the system to read you the subsequent group of items.	**continue**	0.79
	next	1.02
6. You want to hear further information about the current item in a list.	details	1.30
	more info	1.33
13. Assume that you successfully provided some information to the system (like a movie title and a movie theater), but then you were distracted for a moment – now you're not sure what you've already told the system. How do you think you might ask the system to tell you what you've already said?	what did I say?	1.54
	repeat request	1.67
	repeat input	2.09
	repeat command	2.13
	where were we	2.23
14. To make sure your request above was correctly understood, the system just repeated it to you. You're satisfied with the confirmation. Now you want to tell the system to actually carry out your request and retrieve the information from the database.	ok	1.38
	proceed	1.51
	continue	1.68
	go ahead	1.73

Table 1. Phase II survey results for core USI keyword contexts.

covered the eight core USI keywords, some additional navigation keywords, a syntax issue, and a few additional situations for which we have been considering adding keywords. In the second part, subjects were asked to rate 3-6 pre-selected responses for each of the 17 contexts. These responses were selected by the experimenters as utterances that users might be likely to use in each context and were combined with other responses generated by having four pilot subjects answer only the first part of the questionnaire. 82 subjects completed Phase I.

After reviewing the results of both parts of Phase I, we created a one-part, second internet questionnaire to be used as Phase II. This survey reproduced only the keyword rating format of part 2 from Phase I, but this time added responses which had been frequently suggested in Phase I/part 1, and dropped some which had not done well in Phase I/part 2. This resulted in 5-9 responses to be rated for each of the 17 contexts. 50 respondents (none of whom had participated in the Phase I) completed Phase II. In both phases, we attempted to describe the situations as carefully as possible to avoid biasing the answers. We also explained to subjects the concept of these keywords existing as part of a simple,

"mini-language" which would make it easier for a computer to understand the speaker. In the ratings section of each questionnaire, users were asked to rate each response as one of "unacceptable," "not great," "acceptable," or "perfect." The actual questionnaires used and full results for each phase of the survey can be accessed at www.cs.cmu.edu/~sshriver/exp/kwds1001/mainsurvey.html.

RESULTS

The results of Phase II for the core USI keywords are shown in Table 1 (see above URL for full results for both phases). To assess the appropriateness of each keyword, we multiplied each "unacceptable" rating by 5, each "not great" rating by 3, each "acceptable" rating by 1, and each "perfect" result by 0, and averaged the scores for each keyword. A lower score therefore suggests a keyword that overall is likely to be more acceptable. Subjects were not required to answer every question; each item had an average of 48 responses. Table 1 lists each situation and its top-rated keywords; those listed in bold were rated considerably better than the next best option. Weighted scores are also listed as an indicator of 'strength of concept' for cross-situation comparison.

IMPLICATIONS

In some cases ("repeat," "start over"), this survey has validated keywords we have been using in the current USI implementation, while in other instances it is obvious that our current keyword is not the best choice (e.g. situation 13, for which our current keyword "restate" was rated lowest among all the options). Some situations are a bit less clear – for instance, in context 14, the keyword users rated highest ("ok") seems to us to suggest a keyword more appropriate for the context *confirm this action* than for what we actually intended, viz. *execute this action*. Also, we have not taken issues of acoustic clarity into account in this survey, which could be an issue with keywords like "undo" (particularly as compared with our current keyword "scratch that").

We now plan to implement several of the suggestions generated by this survey for use in our large-scale user studies. We expect that testing in this environment will provide further validation both for specific keywords in this and other speech interfaces and for the USI approach in general.

REFERENCES

1. Guzman, S. J., Warren, R., Ahlenius, M., and Neves, D. Determining a Set of Acoustically Discriminable, Intuitive Command Words, in *Proceedings of AVIOS '01* (San Jose CA, April 2001), 241-250.

2. Shriver, S., Rosenfeld, R., Zhu, X., Toth, A., Rudnicky, A., and Flueckiger, R. Universalizing Speech: Notes from the USI Project, in *Proceedings of Eurospeech '01* (Aalborg, Denmark, September 2001).

3. Telephone Speech Standards Committee. Universal Commands for Telephony-Based Spoken Language Systems. *SIGCHI Bulletin 32*, 2 (April 2000), 25-29.