KEYWORD SELECTION, AND THE UNIVERSAL SPEECH INTERFACE PROJECT

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

The Universal Speech Interface project (a.k.a. “speech graffiti”) designs and tests standardized interaction styles for speech communication between humans and simple machines. In this paper we introduce the project and its goals, and describe in detail the process we used for one part of the design process: selecting universal keywords. Finally, we discuss current and planned developments for the project.

## The USI Project

The Universal Speech Interface (USI) project (a.k.a. “speech graffiti”) is an attempt to create a standardized interaction style for speech communication between humans and simple machines. As speech interfaces become more prevalent, our belief is that by designing, testing and promoting a unified look-and-feel, speech applications can become more usable and desirable tools.

The three common approaches to speech user interfaces – unconstrained natural language (NL), directed dialog, and command-and-control – each have their own strengths and weaknesses. NL systems in principle require no previous knowledge or training on the part of the user, but this flexibility obscures the functional limitations of the system and makes it difficult for the user to understand what the application can and cannot do (it also strains the recognition and understanding technology). Directed dialog and command-and-control systems solve this transparency problem by limiting what the user can say at any point, but this is done at the cost of user control. Directed dialog systems guide users through a specific dialog path, asking appropriate questions along the way, but these steps can be tedious for experienced users who want to get to their goal more quickly. Command-and-control systems give more control to the user, but require users to learn a specialized vocabulary and syntax for each application, which becomes infeasible as the number of applications grows.

We have addressed these issues in the USI project by developing a unified look-and-feel speech interface design which users can learn in order to enable them to explore and use any application compliant with that design. We avoid the problems of directed dialog systems by making it easy for expert users to go directly to their ultimate goals, while still allowing novice users to successfully navigate their way through the system. The USI system comprises a set of universal keywords and interaction guidelines, which alleviates the command-and-control problem of having to learn and remember new interaction protocols for different applications. Furthermore, the standardized interaction guidelines help make the system more transparent than NL interfaces.

The USI approach has other advantages as well. The semi-structured interaction and reduced vocabulary means higher speech recognition accuracy compared to NL interfaces (results forthcoming). The interface’s small footprint makes it appropriate for use in small devices where application size is a factor. Finally, standardization also enables the creation of application-

generators and developer toolkits, which dramatically speed up development of new applications.

For the philosophy underlying the USI project, see [1]. For more information about the project, see http://www.cs.cmu.edu/~usi.

## Choosing Keywords

A core feature of the USI is its set of standardized keywords. These keywords are used to address *interaction universals* – events, needs or situations which recur across a wide range of applications. Our original selection of keywords for the interface was based mostly on our own intuitions about which words had simple, unambiguous meanings and were at the same time relatively acoustically distinct. In early user studies, we found that most users were able to learn and use these keywords fairly well, but we thought it would be a prudent to investigate other potential keyword choices, and to see if others' intuitions matched ours. To do this, we designed and administered a web-based user survey. Our approach was strongly influenced by the studies presented in [2, 3], the goals of which, like ours, were to determine effective keywords for use in voice applications. However, neither of the previous studies covered the exact set of functionalities as the USI keywords do. Furthermore, the survey described in [3] elicited keywords from users but did not ask subjects to rate any keywords proposed by the researchers; the study in [2] used both approaches, but surveyed only a very small pool of subjects.

Our survey was conducted in two phases. For the first phase, 82 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 action in each situation. They were subsequently also asked to rate 3-6 responses pre-chosen by our team for each of the 17 contexts. Ratings were done on a four-point scale: "unacceptable," "not great," "acceptable," and "perfect." We then chose 5-9 of the best-rated responses for each context from Phase I, and for Phase II asked 50 different subjects to rate these responses on the same scale. The full results of Phase II are shown in figure 1. Note that a few of the 17 contexts used in the survey (e.g. #10 and #15-#17) correspond to functionalities that are not implemented in the current USI system, but we asked users about these since we intend to incorporate them in future versions.

To rank the final results of Phase II, we used the following formula: each "unacceptable" rating incurred a cost of five points, each "not great" rating incurred a cost of three points, and each "acceptable" rating a cost of one point. The points for each proposed option were then totaled ("perfect" ratings incurred zero points) and divided by the number of respondents who rated that keyword, producing the adjusted weight score shown in the right-most column of Fig. 1. The best-ranking choice for each context is listed in boldface, and items that were rated significantly higher than the next-best choice are listed in italics.

In some cases (e.g. #1 and #2), this survey provided strong validation for keywords we had been using in the current USI implementation, but in other instances it seems that our current keyword

*may* not be the best choice. This survey did not address acoustic dissimilarity for the keywords, and more importantly it did not ask users to rate the options in the true context of interacting with the system. Nonetheless, we believe this survey provides valuable "discount usability"[1] information about user preferences and intuitions, as creating a truly in-context survey for assessing such a wide variety of keywords would be a rather large undertaking. It is not clear to us, therefore, that indiscriminately switching all of our keywords to the best-ranked options from this particular study is the right thing to do for the interface. Instead, we have now chosen a revised working set of keywords influenced by the results of this study and are currently working on a larger-scale user study, which we hope will provide further insight into keyword choice issues.

Figure 1:

**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. You might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***repeat*** | 2 | 9 | 19 | 23 | 1.06 | repeat |
| *repeat that* | 2 | 11 | 26 | 13 | 1.33 | |
| what? | 13 | 9 | 18 | 12 | 2.12 | |
| again | 6 | 20 | 21 | 3 | 2.22 | |
| say again | 8 | 20 | 20 | 3 | 2.35 | |
| undo | 16 | 22 | 11 | 3 | 3.02 | |

**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. You might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***start over*** | 0 | 5 | 24 | 24 | 0.74 | |
| restart | 4 | 12 | 21 | 15 | 1.48 | start over |
| begin again | 0 | 17 | 24 | 9 | 1.50 | |
| reset | 5 | 20 | 22 | 5 | 2.06 | |
| clear all | 5 | 19 | 25 | 3 | 2.06 | |
| cancel | 9 | 19 | 22 | 2 | 2.38 | |
| delete | 16 | 28 | 8 | 0 | 3.31 | |
| erase | 18 | 28 | 5 | 1 | 3.44 | |

**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). You might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***undo*** | 2 | 14 | 26 | 9 | 1.53 | |
| back | 4 | 19 | 21 | 6 | 1.96 | scratch that |
| forget that | 6 | 16 | 23 | 6 | 1.98 | |
| cancel | 9 | 13 | 25 | 4 | 2.14 | |
| scratch that | 8 | 19 | 16 | 8 | 2.22 | |
| ignore | 8 | 21 | 19 | 3 | 2.39 | |
| delete | 12 | 20 | 18 | 2 | 2.65 | |
| erase | 12 | 24 | 12 | 2 | 2.88 | |

[1] *viz*., usability methods "that are cheap, fast, and easy to use" [4]

**4. You're in the middle of interacting with the system and want to find out what you can say at that point. You might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **options** | 1 | 7 | 37 | 8 | 1.19 | |
| help | 3 | 10 | 20 | 19 | 1.25 | now what? |
| menu | 3 | 14 | 28 | 7 | 1.63 | |
| choices | 3 | 12 | 31 | 4 | 1.64 | |
| what can I say? | 7 | 16 | 18 | 11 | 1.94 | |
| now what? | 11 | 22 | 13 | 3 | 2.73 | |
| what now? | 13 | 22 | 12 | 5 | 2.75 | |
| what's next? | 16 | 17 | 13 | 4 | 2.88 | |

**5. Say the system has read you the first set of items. When you want the system to read you the subsequent group of items, without referring to the items by name, you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***continue*** | 0 | 4 | 26 | 20 | 0.76 | more |
| next | 0 | 5 | 38 | 10 | 1.00 | |
| more | 2 | 14 | 24 | 11 | 1.49 | |
| go on | 2 | 11 | 20 | 9 | 1.50 | |
| skip | 18 | 22 | 11 | 1 | 3.21 | |

**6. If you want to hear further information about the current item, without referring to the item by name, you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **details** | 2 | 9 | 30 | 12 | 1.26 | more |
| more info | 2 | 9 | 31 | 10 | 1.31 | |
| tell me more | 4 | 13 | 25 | 9 | 1.65 | |
| additional information | 4 | 20 | 20 | 5 | 2.04 | |
| more | 11 | 24 | 12 | 3 | 2.78 | |

**7. If you want the system to go back to the initial item in the list, without referring to the item by name, you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***first item*** | 0 | 11 | 28 | 14 | 1.15 | first |
| go to beginning | 7 | 8 | 26 | 12 | 1.60 | |
| start over | 10 | 11 | 21 | 9 | 2.04 | |
| beginning | 5 | 20 | 18 | 4 | 2.19 | |
| go to top | 8 | 21 | 22 | 0 | 2.45 | |
| first | 5 | 30 | 15 | 2 | 2.50 | |
| list | 17 | 32 | 3 | 0 | 3.54 | |

**8. If you want the system to go to the final item in the list, without referring to the item by name, you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***last item*** | 1 | 3 | 31 | 18 | 0.85 | last |
| end of list | 0 | 12 | 29 | 11 | 1.25 | |
| go to the end | 3 | 15 | 25 | 9 | 1.63 | |
| last | 6 | 17 | 24 | 5 | 2.02 | |
| bottom | 6 | 31 | 11 | 1 | 2.73 | |
| end | 15 | 21 | 12 | 4 | 2.88 | |
| finish | 25 | 21 | 5 | 1 | 3.71 | |

**9. If the system has just read an item to you, and you want to hear it again, without referring to the item by name, you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***repeat item*** | 2 | 6 | 24 | 21 | 0.98 | repeat |
| repeat | 4 | 8 | 22 | 18 | 1.27 | |
| repeat that | 2 | 12 | 23 | 15 | 1.33 | |
| again | 3 | 17 | 27 | 3 | 1.86 | |
| say that again | 8 | 15 | 22 | 7 | 2.06 | |
| say again | 9 | 17 | 22 | 4 | 2.27 | |
| what? | 16 | 15 | 14 | 7 | 2.67 | |
| restate | 16 | 26 | 8 | 2 | 3.19 | |

**10. If the system has just read item A to you and now you want to hear items B and C (without having to issue another command between them, and without referring to the items by name), you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***next two items*** | 1 | 6 | 29 | 16 | 1.00 | (n/a) |
| next two | 1 | 11 | 26 | 13 | 1.25 | |
| continue | 8 | 14 | 20 | 7 | 2.08 | |
| next, next | 9 | 25 | 14 | 3 | 2.63 | |
| forward | 9 | 33 | 9 | 0 | 3.00 | |

**11. If the system has just read item C to you and you want to hear item B again, without referring to the items by name, you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***previous item*** | 2 | 5 | 28 | 16 | 1.04 | previous |
| back one | 2 | 9 | 30 | 10 | 1.31 | |
| go back one | 0 | 14 | 25 | 12 | 1.31 | |
| repeat previous | 3 | 15 | 27 | 6 | 1.71 | |
| previous | 3 | 15 | 28 | 5 | 1.73 | |
| repeat last | 7 | 17 | 22 | 4 | 2.16 | |
| back | 7 | 22 | 17 | 2 | 2.46 | |
| back up | 10 | 23 | 13 | 4 | 2.64 | |

**12. Now imagine that you're using a specific system that provides movie information over the phone. You're interested in finding out when the movie Heist is playing at the Waterfront Theater. Assuming that the system can only accept fairly simple input, how do you think you might convey this particular request to the system? You might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **what time does heist play at the waterfront?** | 1 | 9 | 22 | 21 | 1.02 | |
| times for heist at waterfront theater | 1 | 8 | 30 | 14 | 1.11 | |
| list show times for heist at waterfront theater | 0 | 9 | 32 | 12 | 1.11 | |
| when is heist playing at the waterfront? | 2 | 9 | 24 | 16 | 1.20 | theater is the waterfront, movie is heist, show times are what? |
| waterfront theater, heist, show times | 9 | 16 | 21 | 5 | 2.24 | |
| heist, waterfront theater, times | 11 | 17 | 18 | 5 | 2.43 | |
| theater is the waterfront, movie is heist, what are the show times? | 17 | 19 | 13 | 3 | 2.98 | |
| theater is the waterfront, movie is heist, show times are what? | 22 | 22 | 8 | 1 | 3.47 | |

**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?**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **what did I say?** | 4 | 11 | 22 | 14 | 1.47 | restate |
| repeat request | 2 | 15 | 23 | 8 | 1.63 | |
| repeat input | 6 | 15 | 22 | 5 | 2.02 | |
| repeat command | 7 | 14 | 25 | 3 | 2.08 | |
| where were we? | 7 | 19 | 16 | 8 | 2.16 | |
| last command | 9 | 19 | 14 | 3 | 2.58 | |
| repeat | 16 | 20 | 9 | 4 | 3.04 | |
| restate | 13 | 26 | 7 | 2 | 3.13 | |

**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. For this you might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **ok** | 3 | 12 | 21 | 15 | 1.41 | go |
| proceed | 2 | 12 | 31 | 7 | 1.48 | |
| go ahead | 4 | 13 | 29 | 6 | 1.69 | |
| continue | 5 | 10 | 31 | 4 | 1.72 | |
| do it | 3 | 19 | 18 | 10 | 1.80 | |
| execute | 6 | 20 | 19 | 7 | 2.10 | |
| go | 5 | 26 | 16 | 4 | 2.33 | |
| correct | 9 | 19 | 18 | 4 | 2.40 | |

**15. Say you've used this movie info system to find out what movies are playing at the Squirrel Hill Theater. While the system is reading you the list of movies, you ask about the show times for one of the movies, and so now the system is reading you a list of times. What do you think you might say to have the system move back and continue reading the movie list?**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **back to movies** | 1 | 9 | 27 | 14 | 1.16 | (n/a) |
| return to movies | 2 | 11 | 28 | 13 | 1.31 | |
| continue movie list | 2 | 14 | 25 | 11 | 1.48 | |
| movie list | 2 | 13 | 31 | 5 | 1.57 | |
| more movies | 3 | 16 | 29 | 3 | 1.80 | |
| next movie | 7 | 12 | 27 | 5 | 1.92 | |
| movies | 3 | 25 | 22 | 1 | 2.20 | |
| back | 14 | 21 | 10 | 0 | 3.18 | |
| return | 15 | 26 | 7 | 0 | 3.33 | |

**16. Say you're using a system like this to control a device like a stereo, and so the system is always in a kind of "standby" state waiting to hear voice commands. What do you think you might say to alert the system that you would like to start interacting with it?**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| ***stereo on*** | 1 | 8 | 22 | 17 | 1.06 | (n/a) |
| hello stereo | 4 | 16 | 23 | 9 | 1.75 | |
| stereo | 7 | 14 | 20 | 8 | 1.98 | |
| system on | 8 | 24 | 15 | 6 | 2.40 | |
| wake up | 12 | 18 | 18 | 2 | 2.64 | |
| start | 16 | 21 | 9 | 4 | 3.04 | |
| ready | 15 | 22 | 11 | 2 | 3.04 | |
| hello | 20 | 17 | 8 | 5 | 3.18 | |
| system | 13 | 30 | 7 | 0 | 3.24 | |

**17. Say you wanted to decrease the volume on the stereo. You might say**

| | unacceptable | not great | acceptable | perfect | adjusted weight | original USI keyword |
|---|---|---|---|---|---|---|
| **volume down** | 1 | 6 | 32 | 11 | 1.10 | (n/a) |
| lower volume | 1 | 11 | 31 | 9 | 1.33 | |
| decrease volume | 3 | 11 | 25 | 12 | 1.43 | |
| turn it down | 6 | 17 | 19 | 11 | 1.89 | |
| less volume | 5 | 20 | 22 | 4 | 2.10 | |
| quieter | 10 | 20 | 17 | 5 | 2.44 | |
| softer | 10 | 19 | 19 | 3 | 2.47 | |
| lower | 13 | 31 | 6 | 1 | 3.22 | |

## Current and Planned Developments

*Generalization to other application types:* Our universal design was informed by on-paper analysis of speech applications of diverse types: information access, data entry, transactions, device control, and more. However, so far we have only implemented information access applications, and thus only that part of the design has been tested and refined. We have recently started implementing device- and gadget-control applications, and are planning to move on to other application types as well.

*More comprehensive user studies:* So far we have performed limited user studies, which upheld the learnability of the basic design as well as skill retention. A more comprehensive study is planned for Spring 2002, where the USI approach will be compared directly with a natural language approach.

*Automated interactive tutorial*: Using USI applications require a one-time, 5-minute training session. While this has proven successful in a face-to-face setting, for ultimate widespread dissemination an automated method must be developed and tested. We have started the development of an interactive tutorial that can be taken over the telephone, with or without web access.

*Application Generator:* One of the advantages of the USI approach is that the semi-structured nature of the interface makes possible a development toolkit that can dramatically accelerate development time for new applications while enforcing compliance with the universal design (much like the Macintosh

developer's toolkit did for Macintosh-style GUI applications). In the case of speech applications, which typically require significant expertise to develop, such a toolkit also reduces the expertise barrier. We have recently developed a web-based application generator [5] which allows people with no speech technology background (in fact, even non-programmers) to create new USI applications in as little as 15 minutes. We hope to demonstrate this tool at the conference.

## REFERENCES


1. Rosenfeld, R., Olsen, D., and Rudnicky, A. "Universal Speech Interfaces." *Interactions 8*, 6 (2001), 34-44.
2. 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.
3. Telephone Speech Standards Committee. "Universal Commands for Telephony-Based Spoken Language Systems." *SIGCHI Bulletin 32*, 2 (April 2000), 25-29.
4. Nielsen, J. "Heuristic Evaluation," in *Usability Inspection Methods*, J. Nielsen and R. L. Mack, Eds. New York: Wiley, 1994, pp. 25-62.
5. Toth, A., Harris, T. K., Sanders, J., Shriver, S., and Rosenfeld, R. "Towards Every-Citizen's Speech Interface: An Application Generator for Speech Interfaces to Databases," submitted to ICSLP 2002.