COMBINING SPEECH AND EARCONS TO ASSIST MENU NAVIGATION

Maria L.M. Vargas

University of North Dakota
Department of Computer Science
Grand Forks, ND 58201
vargas@cs.und.edu

Sven Anderson

Bard College
Computer Science Program
Annandale-on-Hudson, NY 12504
sanderso@bard.edu

ABSTRACT

Previous research on non-speech audio interfaces has demonstrated that they can enhance performance on menu navigation tasks. Most of this work has focused on tasks in which the menu is not spoken and visual representation of the menu is accessible throughout the task. In this paper we explore the potential benefits that earcons, a type of structured sound, might bring to spoken menu systems for which a visual representation is not available. Evaluation of two spoken menu systems that differ only in whether they also employ earcons, indicates that the use of earcons improves task performance by reducing the number of keystrokes required, while also increasing the time spent for each task.

1. INTRODUCTION

The need for efficient, auditory-based, navigation of information structures has grown with the ubiquity of telephone-based automated information services. Increasingly, telephone users are presented with a large number of choices that are structured and presented as a hierarchical menu. Telephone-based interfaces usually present information acoustically and are therefore inherently serial. Spoken entries in telephone-based menu systems include short words or phrases grouped together under related headings. Menu options at the current level are spoken in sequence and users press buttons on their phone's numeric keypad to select from a small number of choices.

In additional to their relatively slow serial presentation, auditory menus are difficult to navigate. Rosson [1] investigated hierarchical telephone-based interfaces and found that positional information was often implicit in the speech token used to represent a menu item. A user must, therefore, understand the hierarchy's category/sub-category relationships, since no feedback concerning change in hierarchy level is provided.

This paper describes a technique that uses earcons to convey additional locational cues during menu navigation. Earcons are short, non-speech sounds, that can be constructed to impart additional information. The use of earcons to represent hierarchically arranged auditory menus has been studied by several researchers [2, 3]. Results from a number of studies indicate that earcons can significantly increase user efficiency during navigation of a visual menu system [3, 4, 5]. Leplâtre and Brewster (2000) found that when navigating a cellular telephone interface participants made 17% fewer keystrokes when earcons were included in the interface. Additionally, the sonified interface significantly decreased errors and required no more time to navigate.

The present research investigates whether the same advantages exist when earcons are added to spoken menu systems like those encountered during telephone-mediated database access. The processing of speech in humans is known to rely on information processing mechanisms that differ from the processing mechanisms used to processes non-speech acoustic signals (see [6] for a review). If speech and non-speech sound can be processed in parallel with limited interference, we hypothesize that earcons may further enhance a spoken interface, much as they are known to enhance graphical menu systems. To test this hypothesis, we evaluated the usefulness of earcons in the context of a novel interface to automobile accessories, such as headlights and windshield wipers. The safe operation of an automobile demands significant attention to visual cues, and therefore precludes use of extensive visual menus. An auditory interface for control of automobile accessories might permit the driver to change the state of various accessories without attending to a visual display.

2. A SONIFIED AUTOMOBILE INTERFACE

We explored the sonification of a spoken interface using a novel automobile simulation that would be completely unfamiliar to experiment participants [7]. The control of many existing automobile accessories (e.g., the radio) requires a driver to redirect her attention away from the road; for example, a driver may need divert attention from the road while tuning the current radio station. Direct visual feedback from such controls can divert visual attention from driving and should be minimized. An auditory interface to automobile accessories would permit drivers to maintain full visual attention on the driving task, and may therefore be safer. In addition, an auditory interface may be practical for drivers who have physical limitations that make it necessary to engage complex controls via a small set of buttons attached to the steering wheel. In this instance, a driver might adjust accessories by using the set of buttons to navigate an auditory interface.

We created a simulated interface for common automobile accessories as shown in Figure 1. The simulated interface was implemented in Java 1.3 using the standard Application Programmer's Interface (API). None of the various graphical controls is active; instead, the state of various accessories is changed by using key presses to navigate an acoustically presented menu of subcategories corresponding to the lights, the windshield wipers, the ventilation system, and the radio.

Figure 2 is a graphic representation of part of the menu that was used in this experiment. It depicts the top-level of the menu and all entries within the lights submenu. Due to the size of menu, submenus for the wipers, ventilation, and radio are not shown. Each top-level menu item has a depth ranging from three to five, yielding a total of 133 different menu items. Users traverse the tree

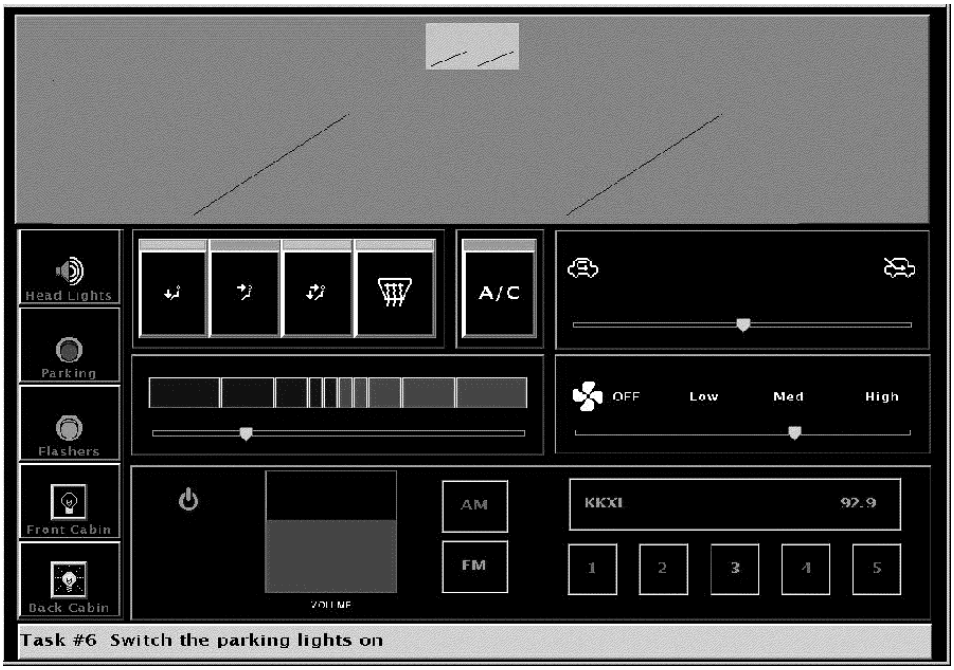

Figure 1: *The simulated automobile interface. Original interface is in color and occupies the entire 17-inch display.*

to locate the accessory they wish to select. Changes to the state of an accessory only occur when a leaf node is selected. For example, the lights-cabin-rear-on/off node allows the user to toggle the state of the rear cabin lights between off and on.

Up-arrow and down-arrow keys are used to change level. The right and left arrow keys are used to traverse the current level, which is linked to form a circular list. Menu items are acoustically presented immediately after navigation keys are pressed. The currently spoken item corresponds to the node in the menu that is being visited. This control strategy permits users to hear the list of items, one-at-time, as many times as desired in the forward or reverse order. The home key returns the user to the root node of the tree, the top-level (i.e., lights-wipers-ventilation-radio). Non-terminal nodes are selected using the down-arrow. Entries at the terminal nodes are selected using the enter key. Selection of an item results in a "click" sound followed by the indicated change of state on the simulated dashboard display.

Limited visual feedback is provided on the simulated windshield and dashboard. For example, when the rear windshield wipers are on they appear to move in the rear-view mirror.

All spoken menu items were prerecorded tokens collected from an adult male speaker of American English. Structured earcons were created to represent each menu item using the guidelines presented in [8]. Each top-level menu item is represented by a particular simulated instrument (timbre) and motif (chord). The lights family timbre is piano; the windshield wipers family uses a chorus; the ventilation family uses bells; and the radio family uses horns. Arpeggios and chords were alternated at the top level to increase distinctiveness of the earcons.

All items beneath a top-level entry inherit the instrument and notes of the top-level motif. Within each node, earcons share timbre and motif and are therefore differentiated on the basis of melody and rhythm as described in [5]. Each earcon begins with a very brief (10-25 msec) percussive sound that indicates the depth of the item in the tree. We experimented with simultaneous playback of speech and earcon, but found that this made it difficult to hear the either the speech or earcon clearly. Therefore, the playback of earcons precedes speech in these experiments. Both earcon and speech playback can be interrupted simply by pressing any of the navigation keys. We found that once participants were familiar with the menu, they did not wait to hear entire menu items, but quickly moved to the next item. The Java Sound API has playback latency that is much greater than 100 msec, which would have slowed the interface appreciably and compromised the experiments. Audio playback therefore made use of the Tritonus Java Sound API, which exhibits a playback latency of about 20 msec.

3. METHODS

Thirty six participants were recruited and randomly placed in the Speech Only Group or Earcon and Speech Group. During the experiment participants completed a list of tasks, one at a time (e.g., "turn on the headlights"). Participants in the Speech Only Group heard spoken menu items with no earcons. Participants in the Earcon and Speech Group heard an earcon preceding each spoken menu item.

Prior to each experiment, the simulated automobile interface and auditory menu were explained to participants by reading from a prepared script. Participants were shown a graphic representation of the menu and permitted to become familiar with the software for five minutes during which time they performed five practice tasks and explored on their own initiative. The graphic menu was then removed and the participant began to perform the 43 tasks of the experiment.

All participants completed the same set of tasks by navigating the menu and making selections using a small subset of keys on the

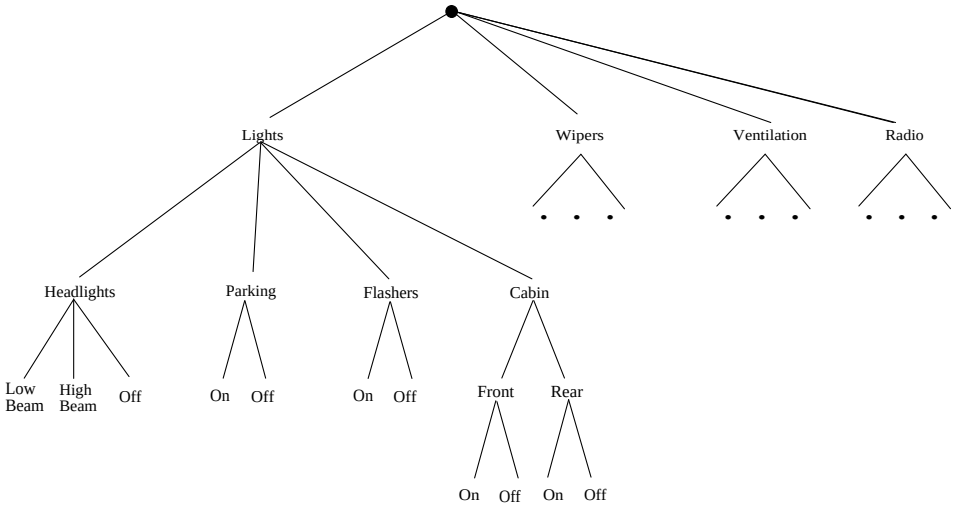

Figure 2: *Automobile interface options menu.*

Table 1: Sample of tasks used for the experiment.

1. Turn the radio on.
2. Tune in to FM - KKXL.
3. Turn the AC on.
4. Set the temperature to cold.
5. Set FM - KKXL to preset3.
6. Switch the parking lights on.
7. Set the volume to 8.
8. Turn the front intermittent wipers to slow.
9. Switch the vent to mixture.

keyboard as described in the previous section. After completing the current task, participants moved to the next task by pressing a labeled key. If the user did not end the task on the correct menu item, the software automatically made the necessary corrections to the graphical state and currently selected item. This allowed each user to start each task from the same item in the menu. The software automatically logged all user keystrokes as well as their time of occurrence. After completing the experiment, participant perceptions of task workload were measured using the computer based version of the NASA Task Load Index (TELEX) [9].

A total of 36 participants completed the experiment. One participant had significant difficulty using the interface and required 50% longer to complete the tasks than any other participant from either group. This participant's data was excluded from further analysis. Therefore, the results reported in this paper are based on 17 participants in the Speech Only Group and 18 participants in the Earcon and Speech Group.

4. RESULTS

4.1. Time

One measure of the efficiency of an interface is the time it requires. The average time spent to complete a task was 11.5 seconds for the Speech Only Group and 13.6 seconds for the Earcon and Speech Group, an additional task time of 18%. The extra time spent by members of the Earcon and Speech Group was significant ($t_{33} = -2.32, p = 0.027$). One explanation is that the auditory items (earcon plus speech) takes longer than the speech alone. On average, the earcons plus speech take approximately 90% longer than speech alone. However, the 18% increased time spent by the Earcon and Speech Group is substantially shorter than this increase in the duration of the items they heard, indicating that participants did not simply slow their responses to listen to the longer stimuli.

4.2. Keystroke Count

A second measure of efficiency is the number of keystrokes needed to complete a task. Total keystrokes also reflects participant familiarity with the organization of the menu, since extra keystrokes indicate the subject is searching for the correct item. The minimum number of keystrokes necessary to complete all of the tasks is approximately 250. The mean number of keystrokes for the Speech Only Group was 496.8; it was 431.0 for the Earcon and Speech Group. Thus, the Speech Only Group required 15% more keystrokes ($t_{33} = 2.17, p = 0.037$), suggesting that the Earcon and Speech Group participants benefited from information conveyed by the earcons.

The average number of keystrokes made for each task is plotted in Figure 3. Tasks are arranged in the order in which they were presented to all subjects. Note that the Speech Only Group requires more keystrokes for nearly all tasks. We might expect that the Earcon and Speech Group participants would also use the extra acoustic cues to learn their task more quickly. In this case we would expect the between-groups difference in the number of keystrokes to increase as the experiment progressed. In fact, no significant change in keystrokes as a function of task number was observed for the two groups.

4.3. Task Completion and Errors

Only three participants were able to complete all 43 tasks. The average number of completed tasks was slightly greater for the Earcon and Speech Group (40.3 vs. 39.5), but this difference was not significant.

Successful completion of a task does not necessarily mean that subjects were able to select the correct item with their first selec-

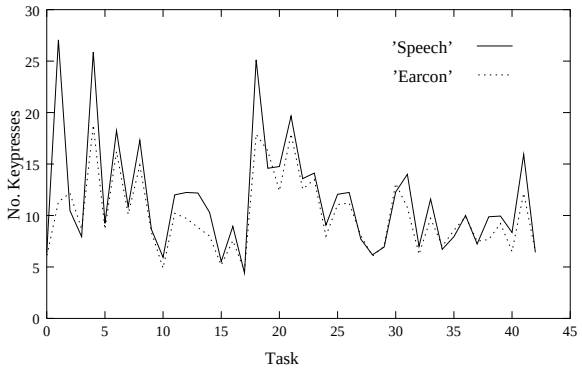

Figure 3: *Number of keystrokes to complete each task.*

Table 2: Average workload scores for the two groups. The final column lists the probability of a t-test indicating whether the differences of the means are significant.

Workload Category	Speech Group	Eancon-Speech Group	t-test
Mental Demand	41.5	52.8	0.153
Physical Demand	11.2	11.7	0.833
Temporal Demand	32.4	39.7	0.272
Effort	43.2	36.9	0.382
Performance	31.5	23.3	0.265
Frustration Level	27.4	24.4	0.692
Weighted Workload	36.9	35.4	0.761

tion. Very often incorrect items are selected prior to selection of the correct item. The Speech Only Group made an average of 5.6 errors on completed tasks, whereas the Earcon and Speech Group made only 1.7 errors. This difference did not reach significance ($t_{33} = 1.66, p = 0.115$), since the mean difference was largely attributed to one task.

4.4. Workload

The NASA Task Load Index is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration. The mean scores for the TELEX are shown in Table 2. Temporal and mental demands were ranked higher by the Earcon and Speech Group, though effort was lower. No differences attained significance.

5. DISCUSSION

It is interesting to compare our findings with those obtained by Leplâtre and Brewster (2000), who employed earcons in a visual, non-spoken cellular phone interface. Both experiments involved menus of similar size and used similar principles for earcon construction. The reduction in the number of keystrokes is in good agreement (15% vs. 17%). Leplâtre and Brewster found no significant difference in the time per task, whereas in the present experiment participants who heard earcons required 18% more time per task. This difference may be attributed to the additional time required to present earcons in conjunction with spoken words. When combining earcons with a visual menu, earcons and menu items can be displayed in a completely parallel manner.

Subjects in the Speech Only Group took less time but made more keystrokes. As a result, the rate of keystrokes for the Speech Only Group was 1.0 key per second, whereas it was only .75 key per second for the Earcon and Speech group. Despite this difference, the TELEX subjective workload assessment did not reveal significant differences in mental demand for the two groups.

Experiment participants spent approximately 15-20 minutes working with the simulated automobile menu system. This corresponds to an interface that is familiar but certainly not well-known. We anticipate that after users become very familiar with such an interface, the possibility of making earcons that are much shorter than spoken words might lead to an interface that can be navigated more quickly than current spoken interfaces.

The results of the current experiment suggest that earcons can be added to spoken menu systems and thereby convey additional information via the auditory modality. Their inclusion may decrease the number of keystrokes and errors involved in the completion of such tasks without making appreciable changes to the overall perceived workload. The explanation for this finding is that earcons provide additional cues about the the relative location of the current menu item for spoken and visual menus alike. Future research will focus on how to optimize and automate the construction of earcons for use with speech.

6. REFERENCES

[1] M.B. Rosson, "Using synthetic speech for remote access to information," *Behaviour Research Methods, Instruments and Computes*, vol. 17, 1985.

[2] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, vol. 4, pp. 11–44, 1989.

[3] Stephen Brewster, V.-P Raty, and A. Kortekangas, "Earcons as a method of providing navigational cues in a menu hierarchy," in *Proceedings of BCS HCI*. 1996, pp. 169–183, Springer.

[4] Stephen Brewster, Adrian Capriotti, and Cordelia Hall, "Using compound earcons to represent hierarchies," *HCI Letters*, vol. 1, pp. 6–8, 1998.

[5] Grégory Leplâtre and Stephen Brewster, "Designing non-speech sounds to support navigation in mobile phone menus," in *Proceedings of the International Conference on Auditory Display*, 2000, pp. 190–199.

[6] A. M. Liberman, Ed., *Speech: A Special Code*, MIT Press, Cambridge, MA, 1996.

[7] Tom E. O'Neil, "Adding some audio to the visual component library," in *Proceedings of the Midwest Instructionand Computing Symposium (MICS 2000)*, St. Paul, Minnesota, 2000.

[8] Stephen A. Brewster, Peter C. Wright, and Alistair D.N. Edwards, "Guidelines for the creation of earcons," in *Adjunct Proceedings of HCI*, 1995.

[9] NASA Human Performance Research Group, "Task load index v1.0 (computerized version)," Tech. Rep., NASA Ames Research Center, 1987.