Auditory Menus Are Not Just Spoken Visual Menus: A Case Study of "Unavailable" Menu Items

Myounghoon Jeon
Sonification Lab, School of Psychology,
Georgia Institute of Technology
654 Cherry Street, Atlanta, GA, 30332 USA
mh.jeon@gatech.edu

Benjamin K. Davison
Sonification Lab, School of Interactive Computing,
Georgia Institute of Technology
654 Cherry Street, Atlanta, GA, 30332 USA
davison@gatech.edu

Siddharth Gupta
Sonification Lab, School of Interactive Computing,
Georgia Institute of Technology
654 Cherry Street, Atlanta, GA, 30332 USA
s.gupta13@gmail.com

Bruce N. Walker
Sonification Lab, School of Psychology & Interactive Computing,
Georgia Institute of Technology
654 Cherry Street, Atlanta, GA, 30332 USA
bruce.walker@psych.gatech.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
Auditory menus can supplement or replace visual menus to enhance usability and accessibility. Despite the rapid increase of research on auditory displays, more is still needed to optimize the auditory-specific aspects of these implementations. In particular, there are several menu attributes and features that are often displayed visually, but that are not or poorly conveyed in the auditory version of the menu. Here, we report on two studies aimed at determining how best to render the important concept of an unavailable menu item. In Study 1, 23 undergraduates navigated a Microsoft Word-like auditory menu with a mix of available and unavailable items. For unavailable items, using whisper was favored over attenuated voice or saying "unavailable". In Study 2, 26 undergraduates navigated a novel auditory menu. With practice, whispering unavailable items was more effective than skipping unavailable items. Results are discussed in terms of acoustic theory and cognitive menu selection theory.

Keywords
Auditory Menus, Auditory User Interface

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces – Evaluation/methodology, Interaction styles (e.g., commands, menus, forms, direct manipulation), Auditory (non-speech) feedback, Voice I/O

General Terms

Design, Experimentation, Human Factors, Performance

Introduction

For decades, an increasing awareness of the limitations of traditional visual interfaces has spurred research on sound as a viable mode of information display [1]. If implemented well, the use of sound can lead to a more universally accessible interface design [2] for users with temporary or permanent vision loss [e.g., 3, 4] and also for users with normal vision [e.g., 5]. The use of speech is the most obvious means of using sound instead of visual text. The most common auditory displays to date have been speech interfaces such as screen readers: JAWS, Window-Eyes, and Thunder. Also, most major operating systems of desktop computers have implemented voice-over for assistive purposes (often inspired by successes of the early MERCATOR [6] project). Considerable academic research has focused on speech interfaces for online systems including audio HTML [7], screen readers for visually impaired people [8], and online help systems [9]. However, relatively lower demand for this assistive technology for the blind results in lack of competition in the commercial market, which leads to slower revisions of products [10], and also slower research [11]. Therefore, considerably more research is still needed to explore even the most basic parts of a speech interface.

Designing auditory interfaces clearly may require different approaches than visual interfaces. In some cases, as Yalla and Walker [12] suggested, using analogy to visual display might be a starting point and an appropriate way of designing parts of an auditory display (e.g., icons vs. earcons and auditory icons; text+icons vs. spearcons and spindexes; scroll bar vs. auditory scroll bar). However, there are important visual interface concepts that are not so easily conveyed with sound. There are yet other cases where a common or default method for auditorily displaying functionality may not be optimal.

In this project, we have been investigating novel ways to make an advanced, more natural, and intuitive speech interface beyond just direct translation from a GUI (as in [6]). This paper presents two case studies looking at one particular (and very common) menu attribute, namely the status of a menu item being *available* ("active", "selectable") versus *unavailable*. We hope to show that both programming infrastructure and design decisions matter, and it is not always appropriate for an auditory interface to simply "say out loud" whatever the visual display "shows". To this end, the present study examined the simple idea of using 'whisper' and 'attenuated voice' for unavailable items in auditory menus. Study 1 compared these two approaches for the display of unavailable menu items in the context of an existing application (JAWS screen reader). We predicted that both of them, given their efficient and intuitive nature, would outperform the currently common method of appending the spoken phrase "unavailable" to menu labels. Study 2 compared the whisper implementation to another existing programming policy, namely just skipping unavailable items, with regard to learning the entire menu structure and forming a mental model for it.

Study 1: Familiar Menu Navigation

As mentioned above, many current auditory menu systems add a word (e.g., "unavailable" to the end of a spoken menu item). However, this is an inefficient and inelegant approach. Based on our experience designing auditory menus for both sighted and visually impaired users, we felt that more acoustically distinct delineation of unavailable items such as whisper or quieter voice would be better than saying "unavailable" at the end of the menu items [13]. We set out to quantitatively investigate this hypothesis.

Participants

Twenty-three undergraduate students (8 female; mean age = 19.7 years) participated in this study for partial credit in psychology courses. They reported normal or corrected-to-normal vision and hearing; and signed informed consent forms.

Stimuli and Apparatus

We created a menu containing 84 menu items based on those found in the Microsoft Word menu on a desktop operating system (e.g., File, Edit, View…). To render the menu structure, an auditory menu was built in Java, leveraging the AudioPlusWidgets library [10]. The menu items were spoken in a female voice via text-to-speech (TTS) generated using the AT&T Labs TTS Demo program with the 'Crystal-US-English' female voice (http://www.research.att.com/~ttsweb/tts/demo.php). Menu items were not visible (see Figure 1). We implemented attenuated and whisper methods for unavailable items in Study 1. Attenuated speech sounds were created by attenuating the original TTS files by 10 dB in Cool Edit Pro 2.0. Whisper sounds were created by using "Morphoder", one of the audio plug-ins in Cubase SX 3.0. Overall, whispered TTS was about 20 dB quieter, but was actually 15 dB louder in the 4 kHz band, and the noise level was boosted by 20 dB from the original TTS. Additionally, we used one more condition in which, rather than changing the speech type to whisper or attenuated, the system added the word "unavailable" to the menu label. For example, the *Print Preview* menu item would become, "Print Preview…unavailable". The motivation for this comes from the existing screen reader software, JAWS, in which "unavailable" is used. On the other hand, Macintosh OS X appends the word "dimmed" to the item (e.g., "Print Preview…dimmed"). We consider this to be an even worse approach, given that "dimmed" describes the visual rendering of the menu item, and not the functional state. Therefore, in our experiment, we adopted the JAWS approach.

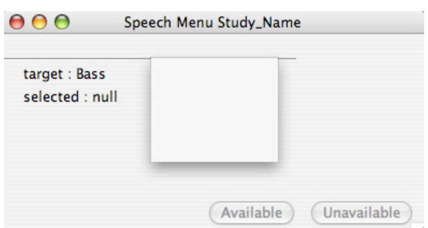

Figure 1. The menu structure for Study 1 & 2. Participants could not see the menu items. In Study 1, a familiar Microsoft Word-like menu was used. In Study 2, an acoustically similar, but semantically unfamiliar menu was used.

Stimuli were presented using a 17" iMac computer, running Mac OS X 10.5. Participants listened to auditory stimuli using Sennheiser HD 202 headphones.

Design and Procedure

There were three within-subjects conditions, based on unavailable item presentation type: Whisper, Attenuated, and Unavailable. The overall goal of the participants was to find a randomly assigned target menu item, by moving with the four cursor keys, and indicating whether it was available or unavailable. On every trial, participants were asked to press 'Option + F' key to start the auditory menu, which always started with the top left menu item. Menu navigation time was operationalized as the time between the first press of the arrow key to start moving, and the press of the

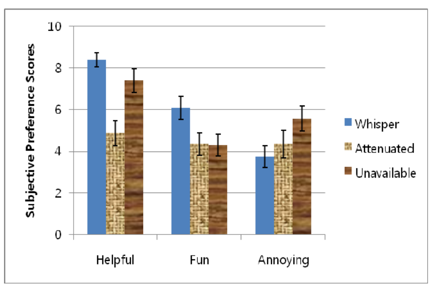

Figure 2. Subjective rating scores. Overall, participants favored the female/female-whisper over other conditions.

enter key. Two types of errors were also logged: Errors of target-selection and errors of type-match to identify the target as available or not. There were no practice trials. Each block contained 30 trials of different names as targets, all rendered in the same manner (e.g., all whisper). In every condition, 30% of the items were randomly designated as unavailable. After completing a block, the next block presented 30 more trials in a different condition, and so on. The order of appearance of the conditions was fully counterbalanced across participants. After three blocks, one for each of the conditions, participants filled out a short questionnaire consisting of 11-point Likert-type questions (e.g., "0" = not at all fun, "10" = very fun).

Results

Overall, objective performance was similar in the three conditions. We measured navigation time, errors in target-selection and errors in type-match (available vs. unavailable). These results were analyzed with a repeated measures analysis of variance (ANOVA), which showed no statistically reliable difference in speed or errors between the speech sound types.

In contrast to the objective performance, the results of the subjective ratings showed significant differences between conditions, as detailed below. Figure 2 shows the results of subjective rating scores. Participants favored the Whisper over other conditions. Repeated measures ANOVA showed a statistically significant difference between speech sound types for 'functionally helpful' rating values, $F(2, 44) = 14.673$, $p < .001$, $\eta_p^2 = .400$; and for 'fun' rating values, $F(2, 44) = 3.546$, $p < .05$, $\eta_p^2 = .139$. For the multiple comparisons among the speech sound types, we conducted paired-samples t-tests. On the 'functionally helpful' scale, participants rated the Whisper $t(22) = 5.273$, p < .001, and the Unavailable $t(22) = -3.710$, p < .01, significantly higher than the Attenuated. Also, on the 'fun' scale, the Whisper showed higher scores than the Attenuated $t(22) = 2.220$, $p < .05$ and the Unavailable $t(22) = 2.138$, $p < .05$. Even if 'annoying' rating values showed only marginal difference $F(2, 44) = 2.489$, $p = .095$, $\eta_p^2 = .102$, users tended to rate the Unavailable condition very high on the 'annoying' scale.

Study 2: Novel Menu Navigation

Study 2 compared the *whisper* version of speech menus with a menu system that would be plausibly created when developers just accept typical menu item behavior. In particular, the default menu behavior in a GUI (such as Java Swing and the "disabled" setting for menu items) is often to simply skip over unavailable items when navigating through the menu with the cursor keys. For sighted users, the unavailable items are typically displayed in grey text (compared to black text for available items). However, a visually impaired user who is navigating an auditory version of the menu with a keyboard would not know about those unavailable, greyed-out items since they are skipped over. That should hinder learning of the menu items, and the overall menu structure. Developing an auditory menu likely requires a different programming approach. If so, it would be important to make this functionality available to developers, and communicate to them the importance of such considerations. In order to obtain more objective data about navigation performance with respect to system learning, we created an unfamiliar menu item for participants to learn [see e.g., 14].

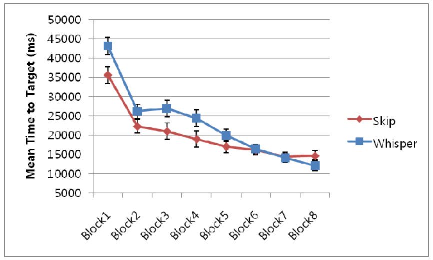

Figure 3. Overall mean time to target (ms). Lower times indicate better performance. Error bars show standard error of the mean.

Participants and Stimuli

Twenty-six undergraduate students (10 female; mean age = 20 years) participated in this study for partial credit in psychology courses. None had participated in Study 1. The apparatus was the same setup as was used in Study 1. Given that we intended to look at learning rates in this study, it was likely that any familiarity with the MS Word-like menu items used in Study 1 could contaminate the results. Thus, we created a new 2-dimensional menu structure. The new menu had an identical layout to the MS Word-like menu used in Study 1, however its labels were replaced with unfamiliar names in order to minimize any effect of users' previous knowledge of the MS Word menu. Instead of MS Word menu titles such as File and Insert, we adopted unfamiliar titles such as Hills and Islands. We were careful to maintain as much acoustic (or at least syllabic) similarity as possible, without any semantic similarity. For example, Insert became Islands, which both have two syllables. As in Study 1, participants could not see the menu items.

Design and Procedure

We used a between-subjects design in Study 2 given the focus on learning effects. Thus, there were two conditions based on speech type: Skipping unavailable items and whispering unavailable items. All participants experienced the same procedure as Study 1 except that the target was randomly selected among available items only. After completing all eight blocks, participants filled out a short questionnaire.

Results

Figure 3 shows mean time to target (i.e., "search time", in ms) for each of the speech types. The mean time to target in the skip condition was lower than that in the whisper condition in early blocks. However, in Block 6, the whisper condition reached the same level and after that, search time in the whisper condition was lower than in the skip condition. These results were analyzed with a 2 (Speech type) x 8 (Block) repeated measures ANOVA, which revealed a statistically significant difference between blocks, $F(7, 168) = 60.554$, $p < .001$, $\eta_p^2 = .72$. While overall there was no difference between speech types, the interaction between speech type and block was statistically significant, $F(7, 168) = 2.649$, $p < .05$, $\eta_p^2 = .10$. This interaction reflects the fact that the practice effect was greater in the whisper condition than in the skip condition. For selection errors, there was no statistically significant difference between speech types. Also, for the subjective rating data, neither 'helpfulness' nor 'likability' scores had statistically reliable differences.

Discussion and Conclusion

We introduced whispered and attenuated TTS sounds as alternative design for the unavailable menu items in auditory menus. The use of whispered and attenuated sounds in unavailable items was compared with the current application (saying "unavailable") and the default policy for the visual menu implementation (skipping unavailable items when manipulating arrow keys). Results were either subjectively (in Study 1) or objectively (in Study 2) positive in favor of the whisper approach, and never significantly negative. Study 1 showed users preferred the whisper over both the attenuated speech version and the version that speaks out "unavailable" for those items. In Study 2, the whisper version outperformed the skipping version, after a moderate amount of practice.

Unlike visual menus, auditory menus are typically presented serially in time. This makes temporally shorter auditory displays an important way to reduce the selection time. Applying whispered sounds for unavailable items might be more intuitive and better than adding "unavailable" to the item label. On the other hand, for short rendering, skipping unavailable menu items also showed hazardous results. Unlike the visual menus which users can scan over the greyed-out menu items, auditory menus do not allow for such scanning. Accordingly, while the skipping of unavailable items can initially obtain efficiency in auditory menus, it seems to prevent users from forming the desirable cognitive layout [15] for the entire menu structure in the long run. Further, generating a different order of menu items according to context (such as in the Microsoft Office 2007 "ribbon") can diminish trust and familiarity with the auditory system.

We focused on the assertion that designers should go beyond a naïve translation from text into speech when creating auditory interfaces. Only through empirical evaluations, one can determine which alternative methods lead to better preference and performance. Through the careful consideration of auditory-specific characteristics, we might be able to expand the application of auditory interfaces and provide universally accessible and acceptable interactions.

References

[1] G. Kramer, "An introduction to auditory display," in G. Kramer (Ed.), *Auditory display: Sonificaiton, audification, and auditory interfaces*, MA: Addison-Wesley, 1994, pp. 1-77.

[2] M. A. Nees, and B. N. Walker, "Auditory interfaces and sonification," in C. Stephanidis (Ed.), *The Universal Access Handbook*, New York: Lawrence Erlbaum Associates, 2009, pp. 507-521.

[3] A. D. N. Edwards, "Soundtrack: An auditory interface for blind users," *Human-Computer Interaction,* vol. 4, pp. 45-66, 1989.

[4] D. McGookin, S. A. Brewster, and W. Jiang, "Investigating touchscreen accessibility for people with visual impairments," *in NordiCHI 2008*, Lund, Sweden, 2008, pp. 298-307.

[5] M. Jeon, and B. N. Walker, ""Spindex": Accelerated initial speech sounds improve navigation performance in auditory menus," in *Human Factors and Ergonomics Society (HFES 2009)*, San Antonio, TX, 2009, pp. 1081-1085.

[6] E. Mynatt, and W. Edwards, "Mapping GUIs to auditory interfaces," in *the 5th Annual ACM Symposium on User Interface Software and Technology*, Monteray, CA, USA, 1992, pp. 61-70.

[7] F. James, "Lessons from developing audio HTML interfaces," in *the Annual ACM Conference on Assistive Technologies*, Marina del Rey, CA, USA., Marina del Rey CA USA, 1998, pp. 27-34.

[8] C. Asakawa, and T. Itoh, "User interface of a home page reader," in *the annual ACM conference on Assistive technologies*, Marina del Rey, CA, USA, 1998, pp. 149-156.

[9] A. Kehoe, and I. Pitt, "Designing help topics for use with text-to-speech," in *the 24th annual ACM International Conference on Design of Communication*, Myrtle Beach, SC, USA, 2006, pp. 157-163.

[10] B. K. Davison, and B. N. Walker, "AudioPlusWidgets: Bringing sound to software widgets and interface components," in *the International Conference on Auditory Display*, Paris, France, 2008.

[11] E. P. Glinert, and B. W. York, "Computers and people with disabilities," *ACM Transactions on Accessible Computing,* vol. 1, no. 2, pp. 1-7, 2008.

[12] P. Yalla, and B. N. Walker, "Advanced auditory menus " No. GIT-GVU-07-12, Georgia Institute of Technology GVU Center, 2007.

[13] M. Jeon, and B. N. Walker, "A preliminary investigation on availability issues in auditory menus," Georgia Institute of Technology Sonification Lab Technical Report, December 2009.

[14] L. Findlater, and J. McGrenere, "A comparison of static, adaptive, and adaptable menus," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 89-96.

[15] K. Norman, *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*, 1991.