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

MYOUNGHOON JEON, SIDDHARTH GUPTA, BENJAMIN K. DAVISON, AND BRUCE N. WALKER
(mjeon@mtu.edu) (s.gupta13@gmail.com) (davison@gatech.edu) (bruce.walker@psych.gatech.edu)

Sonification Lab, Georgia Institute of Technology, Atlanta, GA, USA

Auditory menus can supplement or replace visual menus to enhance usability and accessibility. Despite an increase of research on auditory displays, more research is still needed to optimize the auditory-specific aspects of these implementations. Here, we report on three studies to determine how best to render the important concept of an unavailable menu item. In Study 1, 10 undergraduates and 4 visually impaired adults compared various speech pairs for available and unavailable menu items. Participants preferred a female voice over a male voice and preferred our new implementations such as whispered and attenuated voices for unavailable items. In Study 2, 23 undergraduates navigated an MS Word-like auditory menu with a mix of available and unavailable items. Using a whispered voice for unavailable items was favored over an attenuated voice and saying "unavailable." In Study 3, 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, cognitive menu selection theory, and user interface accessibility.

0 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 (often, but not necessarily, in addition to visuals) can lead to a more universally accessible interface design [2] for users with temporary or permanent vision loss [e.g., 3, 4] and for users with normal vision [e.g., 5, 6]. Many auditory interfaces are implemented as menus in which the menu items are spoken to the user. The most common of those are speech-based applications such as the screen readers, JAWS, Window-Eyes, and Thunder. There have also been some efforts to make systems based on or enhanced by *non-speech* audio, such as Soundtrack [3], SonicFinder [7], Mercator [8], AudioDesktop [9], and Earpod [10]. Whether speech or non-speech, the relatively lower demand for assistive technology has resulted in less competition in the market, which has led to slower revisions of products [11] and also a slower pace of research [12]. Therefore, considerably more research is still needed to explore even the basic and fundamental parts of the auditory interface.

Yalla and Walker [13] suggested that using the analogy of a visual display might be a good starting point and an appropriate way of designing an auditory display. However, in that translation process, care is needed for optimal implementation of auditory interfaces. For example, simple conversion of text into speech is not enough [3]. What is required are mechanisms for fundamentally modifying the display in conjunction with the manner in which information is conveyed. For this, proper reflecting on an interface's auditory characteristics is required [12].

This paper presents a series of three studies looking at one particular (and very common) menu attribute, the state of a menu item being available ("active," "selectable") versus unavailable. We hope to show through this example that both programming infrastructure and design decisions matter, and it is not always appropriate for the auditory display to simply "say" whatever the visual display "shows."

1 AUI (AUDITORY USER INTERFACE) BY ANALOGY WITH GUI (GRAPHICAL USER INTERFACE)

1.1 Auditory Representation of Text

The use of speech is the most obvious means of using sound instead of visual text. Most auditory user interfaces to date have focused on using speech alone. There have been speech interfaces for online systems including audio HTML [14, 15], screen readers for visually impaired people

[16–18], and online help systems [19]. Also, operating systems of desktop computers have implemented speech interfaces (e.g., Voice Over, http://www.apple.com) for assistive purposes. Recently, navigation devices in which voice and sounds are the most important interaction modality have adopted diverse types of voices such as different languages, genders, synthesized text-to-speech (TTS) and real human voice, and even celebrity voices (http://www.garmin.com, http://www.tomtom.com). Even household appliances have incorporated partial or full speech technologies for universal design. For example, some washing machines use a human voice in the help function (http://www.samsung.com) and air conditioners and refrigerators employ TTS and human voice prompts (http://www.lge.com) to guide novice users and older adults. Moreover, the Apple iPod Shuffle is a broadly adopted consumer product that has no visual display at all, and employs only speech interaction. All these recent trends encourage further research into speech in user interfaces.

1.2 Auditory Analogies for Icons

Just as text is not the only feedback in a visual system, speech is not the entirety of an auditory user interface. Recently, non-speech auditory display has been more actively studied as a way to compensate for the weaknesses of speech, such as overall slowness of speech-based interfaces. For example, auditory icons [7] are audio representations of objects, functions, and events that bear an analogic relationship with the objects they represent. They are easily learned because the relation between a source of sound and a source of data is more natural than others. Adopting these advantages, Gaver [20] created an auditory icon-enhanced interface. Also, some researchers have attempted to convert GUIs to nonvisual interfaces using auditory icons [21, 22]. However, it is sometimes difficult to match all the functions of the devices such as “save” or “unit change” with proper auditory icons. As a result, there have been relatively few systematic uses of auditory icons in auditory interfaces, and certainly fewer in auditory menus in particular. Earcons [23], on the other hand, use short sequences of musical sounds as symbolic representations of actions or objects. While this arbitrary mapping means that earcons can be applied to any type of menu, the flexibility requires users to have some training. Earcons have been successfully applied in a desktop computer interface [24] and a mobile device [25, 26]. Also, subsequent hierarchical menu experiments showed that the systematic nature of earcons makes them promising for displaying hierarchical information [44, 45].

1.3 Auditory Analogies for Text + Icons

Modern GUI menus include both text and icons. As a relatively new analogy of this, spearcons [27] and spindex [6] were introduced into speech menus to overcome the shortcomings of either text-only (speech) or icon-only (auditory icons or earcons). Spearcons can be produced by speeding up spoken phrases, even to the point where the resulting sound is no longer comprehensible as a particular word [27]. These unique sounds blend the benefits of text and icons because of the acoustic relation between the spearcons and the original speech phrases. Accordingly, spearcons are easier to learn than other audio icons (e.g., auditory icons or earcons 28, 29). The use of spearcons has enhanced navigational efficiency on the spoken auditory menus of two dimensional interfaces [5] as well as one dimensional menus [30, 31].

A spindex [6] is created by associating an auditory cue with each menu item, in which the cue is based on the pronunciation of the first letter of each menu item. For instance, the spindex cue for “Apple” would be a sound based on the spoken sound “A.” The set of spindex cues in an alphabetical auditory menu is analogous to the visual index tabs in a reference book (e.g., a large dictionary). The benefits of the spindex cues are clearer in long menus with many items [6] because they help per-item speedups. Also, since the spindex cues are part of the original word and natural, they do not require training. The spindex showed promising results in one dimensional menu-navigation and even in a dual task context [32].

1.4 Auditory Representation of Widgets and Structure

Auditory displays have been applied to not only text and icon items, but also entire menu structures. Brewster [33] once used earcons to implement sonically enhanced widgets including buttons, menus, scrollbars, alert boxes, windows, and drag and drop. More recent research on the addition of auditory scrollbars has also demonstrated their potential benefits for visually impaired participants [34].

Ludwig and his colleagues tried to create an audio window system through digital audio processing based on acoustics [35]. For instance, they applied self-animation, thickening, and peaking to highlight and emphasize an audio source in the foreground. In contrast, they used muffling and distancing to deemphasize less prominent parts. Additionally, muffling was suggested to help acoustically denote a metaphorical “grabbing” of an audio source, similar to “grabbing” a visual icon with a mouse in a visual system. Their work is an interesting example of attempts to translate a visual interface into an auditory interface.

To summarize, auditory display researchers have tried to use visual analogy in developing auditory user interfaces and have yielded successful outcomes as shown above. However, given the growing and varying use of audio in the user interface, more research is needed for implementing sophisticated auditory applications.

2 THE CURRENT STUDY AND HYPOTHESES

To develop refined auditory interfaces, the shift from GUI into AUI requires some knowledge of the characteristics of human auditory processing and psychoacoustics. Despite auditory researchers’ efforts, developers and designers still depend on naïve or arbitrary mappings between visual and auditory renderings when developing auditory systems. Text can be converted to synthetic speech, but this does not mean that adapting an interface for use by blind people implies simply plugging a speech synthesizer into a standard computer or terminal [3]. To illustrate, one study used a male voice to read basic text and applied a female voice to hypertext links [16] in a web application. Another study used male and female voices in a menu system to represent its hierarchy [36]. Other researchers used a beep sound for unavailable items on the menu system [18]. Whereas all of these attempted to apply differentiated mappings, none of these designs provided an intuitive mapping between functions (or structure) and acoustical characteristics.

The issue could be solved by varying sound attributes in AUIs, more like the way font or color is used in GUIs and, perhaps, even allowing users to select their favorite mappings just as they choose fonts or colors. Nevertheless, designers or engineers need a default setting value based on empirical findings. User preferences might depend on users' demographic characteristics, the type of device, or the context of use [37]. In addition, sometimes different parts or different tasks of the interface need to be rendered differently in the auditory interface. For instance, when it comes to the menu interface, considerably more visual components remained to be converted into auditory displays, such as divider, shortcut, submenu, dialog box, menu status, as well as menu items and icons.

The current study focused on the common interface issue of depicting unavailable menu items. This is a crucial aspect of menus that needs to be clearly conveyed in both visual and auditory menus. This study investigated the simple idea of using "whisper" and "attenuated voice" for unavailable items in auditory menus. Study 1 was a preliminary study in order to narrow down the basic ideas for the next quantitative studies. It was expected that both sighted and visually impaired users would prefer some types of acoustical mappings more than other alternatives. Study 2 compared the novel whisper approach to what is used in existing applications (e.g., JAWS screen reader), namely saying "unavailable" just after the menu item text. Finally, Study 3 compared the whisper to the common GUI convention of simply skipping unavailable items, in terms of learning the entire menu structure and forming a mental model for it [38].

3 STUDY 1A: SIGHTED USERS' NEEDS AND PREFERENCE FOR MENU STATUS MAPPINIG

Study 1A was a preliminary investigation with sighted undergraduates regarding the sound mapping for unavailable menu items in auditory menus in order to identify users' needs and preference and to inform a later investigation with visually impaired participants.

3.1 Methodology

Participants. Ten undergraduate students (6 female; mean age = 18.8 years) participated in this study for partial credit in psychology courses. All reported normal or corrected-to-normal vision and hearing, signed informed consent forms, and provided demographic details about age and gender.

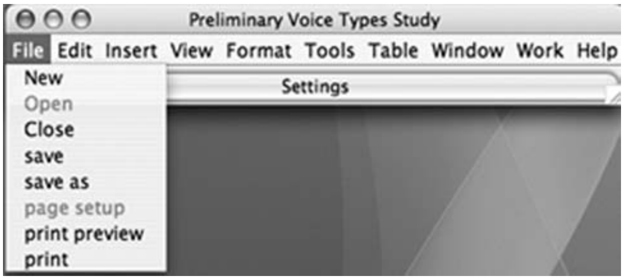


Fig. 1. A Microsoft Word-like menu structure for Studies 1a and 1b. The menu items that are not frequently used were removed from the original MS Word menu.

Apparatus. Stimuli were presented using a 17-inch iMac computer, running Mac OS X 10.5 on a 1.83 GHz Intel Core Duo processor and 1 GB of RAM. The built-in Intel High Definition Audio and speaker were used for sound rendering and listening. The computer monitor was placed on a table 40 cm in front of the seated participants.

Stimulus Menu Structure. A 2-dimensional menu structure was composed containing 84 menu items (see Fig. 1 and Table 1(a)). The menu items were based on the items in the Microsoft Word menu, although items that are not frequently used (e.g., "Send To," "Copy to Scrapbook") were not included. We had several pilots to take out those infrequent items with students who have not participated in any of our studies.

Visual Menu. A Java application was created to present the visual menu, as shown in Fig. 1. The participant was able to navigate the menu by a mouse or with cursor keys. Just as in a usual menu, available items were rendered in black text, and unavailable items were rendered in grey text, using the same font.

Auditory Menu. The auditory menu was built into the same Java program, using the APWidgets library [11]. If a participant selected a menu item, the sound cues were generated by APWidgets as follows. For available menu items, a text-to-speech (TTS) rendering of the menu item text (e.g., "Print Preview") was played. For unavailable items, a different TTS version of the text was played (e.g., a whispered version of "Print Preview").

There were eight conditions based on speech sound type for the item availability (see Table 2). The voice could be male or female; the unavailable items could be the same or different gender; and the unavailable items could be whispered or attenuated.

Auditory Stimuli Details. TTS files (.wav format) were generated for all of the menu items using the AT&T Labs TTS Demo program with the "Mike-US-English" for male voice and "Crystal-US-English" for female voice (www.research.att.com/∼ttsweb/tts/demo.php). Attenuated speech sounds were obtained 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 15 dB louder in the 4 kHz band, and the noise level was boosted by 20 dB from the original TTS.

Table 1. Menu taxonomy used in Study 1 and 2 (a): MS Word and Study 3 (b): Newer. The number in each left column of items indicates the number of syllables in the item.

MS-Word																				
	1	**File**	2	**Edit**	1	**View**	2	**Insert**	2	**Format**	1	**Tools**	1	**Table**	2	**Window**	1	**Work**	1	**Help**
	1	new	2	undo	2	normal	3	page numbers	1	font	5	spelling and grammar	2	draw table	3	zoom window	5	add to work menu	1	search
	2	open	2	redo	4	online layout	3	date and time	3	paragraph	3	thesaurus	2	insert	5	minimize window			2	word help
	1	close	1	cut	3	page layout	1	field	3	document	4	dictionary	2	delete	4	bring all to front			4	sample document
	1	save	2	copy	2	outline	2	symbol	6	bullets and numbering	2	word count	2	select	3	new window			7	use the office assistant
	2	save as	1	paste	4	notebook layout	2	comment	5	borders and shading	4	autocorrect	3	merge cells	3	arrange all			4	check for updates
	3	page setup	1	find	5	header and footer	2	footnote	2	columns			2	split cells	1	split			5	downloads and updates
	3	print preview	2	replace	2	footnotes	2	caption	1	tabs			1	Sort					7	visit the product website
	1	print	2	go to	1	zoom	4	cross-reference	4	autoformat			3	formula					5	send feedback on word
					2	full screen	4	index and tables	2	background										
							2	picture												
Newer																				
	1	**Hill**	2	**Comet**	1	**Town**	2	**Island**	1	**Lake-Pond**	1	**Stars**	1	**Nick**	2	**Mountain**	1	**Tree**	1	**Word**
	1	Bryn	2	Schaumasse	2	Cuicul	3	Didicas	1	Garth	5	Venaticorum	2	Emus	3	Taygetus	5	Bladdernut Colchis	1	Ceps
	2	Kipling	2	Kowal	4	Hippo Regius	3	Eauripik	3	Antigo	3	Algieba	2	Peachie	5	Aroania Chelmos			2	Albus
	1	Glen	1	Koff	3	Vescera	1	Als	3	Tomahawk	4	Cor Caroli	2	Big Cliff	4	Erymanthus			4	Aquaticus
	1	Penn	2	Helin	2	Lixus	2	Belyy	6	Chippewa Flowage Chief	2	Tycho	2	Jabba	3	Aegaleus			7	Mustela Strigidorsa
	2	Dunstan	1	Gunn	4	Volubilis	2	Gharbi	5	Sissabagama	4	Venatici	3	Wallabies	3	Helicon			4	Canadensis
	3	Hamilton	1	Spahr	5	Augusta Zilil	2	Charcot	2	Chetek			2	Moosta	1	Minthe			5	Ostrea Edulis
	3	Powhatan	2	Reinmuth	2	Mustis	2	Javier	1	Bass			1	Joeys					7	Actitis macularius
	1	Brent	2	Westphal	1	Valve	4	Hispaniola	4	Irogami			3	Jillaroos					5	Sterna maxima
					2	Thysdrus	4	Abemama	2	Cochram										
							2	Kola												

Table 2. Speech menu conditions in Study 1A.

Cond.	Available	Unavailable
1	Male	Male Whisper
2	Male	Female
3	Male	Female Whisper
4	Male	Male Attenuated
5	Female	Female Whisper
6	Female	Male
7	Female	Male Whisper
8	Female	Female Attenuated

Procedure. To begin, the experimenter demonstrated visually and aurally one of the speech menu types to the participant and explained the meaning of the sound mapping. Then, each participant was asked to navigate the menu with a mouse or keyboard until he or she felt comfortable with it. In every session, all of the participants were asked to navigate each condition in turn. Meanwhile, participants were encouraged to express any thoughts and feelings about the menus and sound mappings. They were allowed to ask, answer, and discuss with other participants or the experimenter. This procedure was repeated for all eight conditions (i.e., each speech menu type was presented). The order of appearance of the speech menu conditions was randomized for each session. After completing the navigation with all of the speech menu conditions, participants were asked to individually rank the top three choices as first, second, and third for the auditory representation of available and unavailable menu mappings and provide any comments about the speech mapping. All of the conversations and feedback were recorded for subsequent analysis. All of the procedures lasted around 40 minutes for each session. There were four sessions in total and each session was composed of two or three participants.

3.2 Results

As shown in Table 3, weights were allocated to each rank in order to consider the rankings more quantitatively: 3 points for the first rank, 2 points for the second rank, and 1 point for the third rank. Using this method for weighted-rank scoring, the summed points for each pair of the available and unavailable items were calculated as follows: female/female-attenuated (16) $>$ female/female-whisper (12) $>$ male/male-whisper (10) $=$ male/male-attenuated (10) $>$ male/female (5) $>$ female/male (4) $>$ female/male-whisper (3) $>$ male/female-whisper (0). Using the same weighted-rank scoring, a couple of preference analyses followed. First, to look at preferences for speech gender, the conditions were divided into two different categories (male vs. female) for available menu items. As a result of pooling the rank scores, participants preferred female (35) to male (25). Then, the conditions were re-categorized for gender consistency between available items and unavailable items (gender-consistency vs. gender-change). The results clearly showed that participants preferred gender-consistency (48) over gender-change (12) in available items. This preference tendency for the consistent gender was also reflected in the first rank pairs. For the best choice, nine out of ten participants preferred the same gender over the different gender. In particular, five participants preferred the same gender whisper pairs (three chose male/male-whisper; two chose female/female-whisper) and four participants chose the same gender attenuated pairs (three chose female/female-attenuated; one chose male/male-attenuated). Only one participant chose the female/male-whisper as her best choice.

Participants' comments were similar to the ranking results. Some participants favored whisper, noting, "It is so different and distinct. It is fun and I definitely knew which were available or not" and "I feel that the whisper is a key aspect in the voice analysis." However, others were concerned about its quality more than the functional mapping: "The quality of the whispers was not that great" and "Whisper sound is weird." Some participants supported the attenuated version: "It made logical sense and was the most effective" and "clear," but one participant pointed out that, "It needs to have a slightly larger difference in volume," though the 10 dB difference is perceptually half of the original loudness. Overall, the female voice was said to be "more clearly spoken" and "more pleasant sounding" than the male voice. Participants generally preferred hearing the same gender in available and unavailable items because the gender change between them was "too distinct," "hard to relate to," and "really confusing."

These results were sufficient to provide a sense of sighted participants' general preference. To examine whether other classes of users might have different opinions for that designing issue, the investigation continued with visually impaired adult participants.

4 STUDY 1B: VISUALLY IMPAIRED USERS' NEEDS AND PREFERENCE FOR MENU STATUS MAPPING

After completing Study 1A, we narrowed down the number of sound mapping pairs. There are two reasons for this. First, because visually impaired people should depend totally on the auditory modality in our study; eight pairs of alternatives might be hard for them to memorize and discern. Second, based on the recommendation of the Institutional Review Board for research involving vulnerable populations, we tried to reduce the total amount of research time and procedure for their convenience. As a result of Study 1A, we eliminated the gender-change conditions, and thus, four different pairs were used for Study 1B: female/female-whisper, male/male-whisper, female/female-attenuated, and male/male-attenuated.

4.1 Methodology

Participants. Four visually impaired adults participated in this study and received $20 compensation. All were clients of the Center for the Visually Impaired (CVI) in Atlanta (2 female and 2 male; mean age 49.5 years, range 44–55 years). Two participants were totally blind. One participant had 1 degree of vision in the left eye with none in the right eye. The fourth had only light perception. All

Table 3. Points (ranks) for each pair in Study 1A.

Participant No.	Male-MaleWhisper	Male-Female	Male-FemaleWhisper	Male-MaleAttenuated	Female-FemaleWhisper	Female-Male	Female-MaleWhisper	Female-FemaleAttenuated
1	$3(1^{st})$	$1(3^{rd})$						$2(2^{nd})$
2	$3(1^{st})$	$1(3^{rd})$						$2(2^{nd})$
3	$3(1^{st})$				$1(3^{rd})$			$2(2^{nd})$
4	$1(3^{rd})$	$2(2^{nd})$			$3(1^{st})$			
5					$3(1^{st})$	$2(2^{nd})$		$1(3^{rd})$
6				$2(2^{nd})$		$1(3^{rd})$		$3(1^{st})$
7		$1(3^{rd})$		$2(2^{nd})$				$3(1^{st})$
8				$1(3^{rd})$	$2(2^{nd})$			$3(1^{st})$
9				$3(1^{st})$	$2(2^{nd})$	$1(3^{rd})$		
10				$2(2^{nd})$	$1(3^{rd})$		$3(1^{st})$	
Total Points	10	5	0	10	12	4	3	16

participants reported normal hearing and provided demographic details about age, gender, visual impairments, and usage of a desktop computer and a mobile phone. All of them had experience with the screen reader, JAWS, Magic, MAC VoiceOver, Windows Eyes, and System Access, and were familiar with using desktop computers (mean year of use: 18 years) and Microsoft Office. An impartial witness read the consent form to all of the participants and helped them as needed to sign the consent form.

Stimuli and Apparatus. The 2-dimensional auditory menu structure from Study 1A was used. Based on the results of Study 1A, the same gender voice was always used for both available and unavailable items within a block. This resulted in four conditions: The TTS voice could be either male or female, and the unavailable items could be either whispered or attenuated. Stimuli were presented using a 15-inch laptop running Windows XP. The four arrow keys were marked with tape to guide participants' hand(s). Participants used Sennheiser HD202 headphones to listen to the auditory stimuli.

Procedure. After obtaining informed consent, the investigator instructed participants on the goal of the study and the task. Their task was to navigate the Microsoft Word-like auditory menu (could not see the visual menu) with no specific target in mind. They were told that they could navigate the menu until they felt sufficiently familiar with the interface to notice the meaning of the auditory rendering that was composed of different acoustical characteristics for different functionality. In the instruction, however, the instructor did not explain the underlying mechanism of the sound mappings. To begin, the investigator set up one of the speech menu types for the participant and the participant navigated the menu with four arrow keys. Then, the participant was asked to guess the meaning of the sound mapping. While the participant interacted with the menu using a keyboard, he or she was encouraged to express any thoughts or feelings about the menu and sound mapping. This was repeated for all four conditions (i.e., each speech menu type was presented serially). The order of appearance of the speech menu conditions was randomized for each session. In total, we had four different sessions, in each of which we had one participant. After completing the navigation of all of the speech menu conditions, the participant answered the following five questions: (1) Which do you prefer between male voice and female voice for auditory menus? Why? (2) Which do you prefer between whisper and attenuated voice for unavailable items? Why? (3) What do you think about saying "unavailable" or "dimmed" in addition to unavailable menu items? (4) What about gender change across availability? For example, available items–male, unavailable items–female, or vice versa; does it make sense? (5) What is the best choice for unavailable items among alternatives (whispered, attenuated, saying "unavailable," saying "dimmed," and gender change)? Finally, they provided comments about the speech mapping and this study.

4.2 Results

Five Questions:

(1) Which do you prefer between male voice and female voice for auditory menus? Why?

Three out of four participants preferred the female voice over the male voice for auditory menus because the female voice was "clearer," "easier to understand," "simple," and of a "higher tone." This result agrees with that of sighted participants in Study 1A.

(2) Which do you prefer between whisper and attenuated voice for unavailable items? Why?

Two of the participants preferred whisper, noting, "it has more contrast and is a little more understandable for me" and "you have to pay attention more on lower volume (attenuated)." The other two participants liked the attenuated voice, reporting, "(attenuated voice is) clearer and easier to understand" and "whisper is hard to understand."

(3) What do you think about saying "unavailable" or "dimmed" in addition to unavailable menu items?

Two of the participants favored that strategy, saying it would be "helpful" and "beneficial." One of the participants said "it depends." She said, "it could be as good as other alternatives such as whispered or attenuated voice, depending on the situation." Also, she added, "the whisper version could be quicker for power users." The remaining participant pointed out that saying the menu state after the unavailable menu items was not good because it makes speech "too long." As mentioned, since all of four participants have used JAWS, they are accustomed to the strategy of appending "unavailable," even though they are not aware of it.

(4) What about gender change across availability? For example, available items–male, unavailable items–female, or vice versa, does it make sense?

In contrast to sighted people's responses, three out of four participants agreed with the method noting, "it is fine." One participant had even been using multiple voices for various functions in software such as Microsoft Excel. Only one participant worried about it, saying, "it's confusing."

(5) What is the best choice for unavailable items among alternatives? (whispered, attenuated, saying "unavailable," saying "dimmed," and gender change)

For this last question, the four participants provided four different answers. Each one of the participants favored each of the strategies of: saying "unavailable," whispered voice, and gender change (male for available items and female for unavailable items). The remaining participant replied that it is "situational."

Other Results:

Overall, visually impaired users preferred any acoustical contrast in an auditory menu and were excited by the attempt to apply those varied distinctions for it. They were generally satisfied with the quality of the speech implementation in this study, including the whisper sounds. Among the four participants, only one intuitively figured out the meaning of the whisper as unavailable menu items. None of them identified the meaning of the attenuated voice. Moreover, participants pointed to the problems of current screen readers, saying, "a robotic male voice is like a machine" and "the voice tone is monotone." These notions exactly correspond with previous research [15]. The results of Studies 1A and 1B were enough for us to help identify some promising alternatives of sound mapping and narrow down the set of spoken menu item approaches for the subsequent quantitative studies.

5 STUDY 2: FAMILIAR MENU NAVIGATION

Study 2 was conducted to compare these novel approaches for unavailable menu items to the current typical screen reader approach. Because new implementations such as whispered and attenuated voice for unavailable items are shorter and more intuitive than the existing method of appending additional phrases to the menu item (e.g., "unavailable"), we expected them to be favored both in objective and subjective metrics.

5.1 Methodology

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, signed informed consent forms, and provided demographic details about age and gender. None had participated in Study 1A.

Stimuli and Apparatus. In accordance with the results of Studies 1A and 1B, Study 2 included only female/female-attenuated and female/female-whisper renderings for unavailable items. Additionally, there was one more condition in which, rather than changing the speech type to

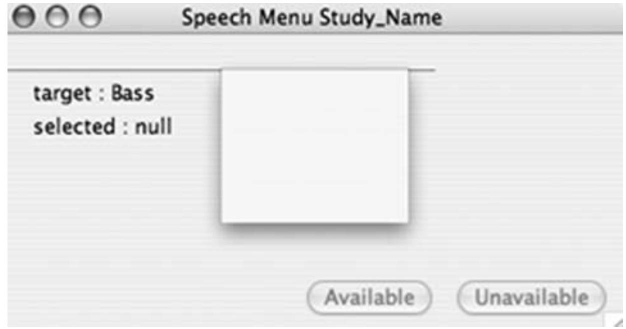


Fig. 2. The screen grab of the visual menu structure for Studies 2 and 3. In Studies 2 and 3, participants could not see the menu items. Rather, they could see the name of the target item, name of the selected item, and the empty frame of the overall submenu structure. They could also identify the target as available or unavailable on it. The familiar Microsoft Word-like menu were used in Study 2. Study 3 used acoustically similar but semantically unfamiliar items (see the Table 1 for the details).

whisper or attenuated, the system appended the word "unavailable" to the menu label. The motivation for this comes from the way in which some existing screen reader software conveys the concept of unavailable menu items. JAWS says "Print Preview. . .unavailable" if the "Print Preview" menu item is unavailable. As a sidebar, it should be pointed out that Voice Over in Mac OS X expresses the case as "Print Preview. . .dimmed." This could be considered a more problematic approach than what is used in JAWS, because it is describing the visual rendering of the menu item, rather than the functional state. This may be of limited utility to someone who cannot see and perhaps never has seen the "dimmed" visual menu.

Note that in this experiment, the menu list wrapped around at the "top" or "end" of the list. That is, if the last menu item was selected, and the down cursor key was pressed, the top menu item would become selected. Also, if a menu item was selected in a menu, and the left or right arrow key was pressed, the menu title of the adjacent menu would be selected, as is typical in visual menu implementations. Except for the new stimuli required for the unavailable condition, all of the stimuli were identical to Studies 1A and 1B. However, in Study 2 participants could not see menu items visually because we wanted to evaluate performance of menu navigation depending only on auditory modality; rather, they could see the name of the target item, name of the selected item, and the empty frame of the overall submenu structure. They could also indicate the target as available or unavailable on it (see Fig. 2). Participants listened to auditory stimuli using Sennheiser HD202 headphones, adjusted for fit and comfort. Other than the headsets, the remaining apparatus was the same as Study 1A.

Procedure. There were three within-subjects conditions based on unavailable item presentation type: female/female-whisper, female/female-attenuated, and female/female-unavailable. At the beginning of every trial, participants pressed "Option + F" to activate the auditory menu, which always started with the top left menu item.

The task of the participant was to find a randomly assigned target menu item by moving with the four arrow keys and then press the enter key as soon as the target was reached. Randomized sampling was without replacement so that it was assumed that variability in the depth of the target item in the menu was pseudo-equally distributed across targets. They then indicated whether it was available or unavailable by clicking a software button with the mouse (see Fig. 2). Menu navigation time was operationalized as the time between the first press of the arrow key to start moving and the press of the enter key. Time to indicate whether the item was available or unavailable was not recorded. Overall, two types of errors were logged: errors of target selection and errors of type-match (identify the target as available or unavailable). The trials in which participants made an error (either target selection or type-match) were not included in navigation time calculation.

There were no practice trials before the experiment. Each block contained 30 trials of different names as targets. 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. An 11-point Likert-type scale was used for the self-rated levels of perceived performance (appropriate, functionally helpful, and discernible) (e.g., "0" = not at all appropriate, "10" = very appropriate), preference (likable, fun, and annoying) [6], and clarity of the speech (articulate, clear, and comprehensible) [39] with regard to speech sounds. Finally, participants were asked to provide comments on the study.

5.2 Results

Objective Performance:

Performance metrics included time to target, errors in target selection, and errors in type-match (available vs. unavailable). Overall, objective performance was similar in the three conditions. Mean navigation time was $M = 15130$ msec, $SD = 5896$, for the female/female-whisper; $M = 15808$ msec, $SD = 5109$, for the female/female-attenuated; and $M = 15482$ msec, $SD = 5206$, for the female/female-unavailable. These results were analyzed with a repeated measures analysis of variance (ANOVA), which showed no statistically reliable difference between the speech sound types, $F(2, 44) = .085$, $p > .05$, $\eta_p^2 = .004$. Also, there was no statistically reliable difference between speech sound types in errors of target-selection, $F(2, 44) = .525$, $p > .05$, $\eta_p^2 = .023$. Mean errors were $M = 1.26$, $SD = 1.36$, for the female/female-whisper; $M = 1.13$, $SD = 1.39$, for the female/female-attenuated; and $M = .96$, $SD = 1.22$, for the female/female-unavailable condition. Similarly, there was no difference between speech sound types in errors of type-match, $F(2, 44) = 2.554$, $p > .05$, $\eta_p^2 = .104$. Mean errors were $M = 1.52$, $SD = 2.15$, for female/female-whisper; $M = 1.57$, $SD = 2.00$, for female/female-attenuated; and $M = .65$, $SD = 1.23$, for female/female-unavailable.

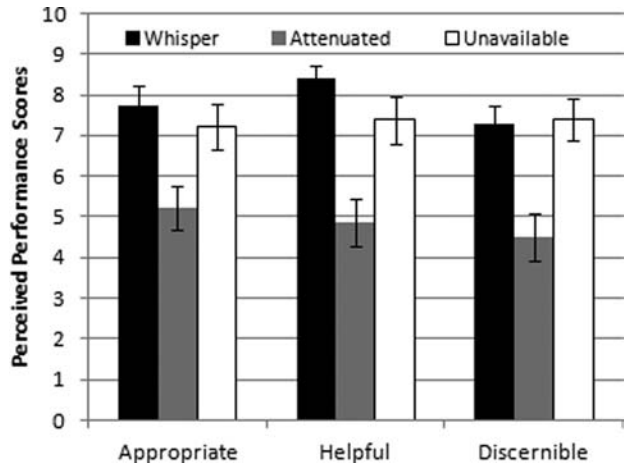


Fig. 3. Perceived performance scores for Study 2. Participants favored the female/female-whisper and the female/female-unavailable types.

Subjective Ratings:

In contrast to the objective performance, the results of the subjective ratings showed statistically significant differences between conditions as detailed below and shown in Figs. 3 to 5.

Perceived Performance. Fig. 3 shows the results of perceived performance, which is a measure of how well the participant thought each condition should perform. This figure suggests that participants favored the female/female-whisper and the female/female-unavailable types over the female/female-attenuated. Repeated measures ANOVA showed a statistically significant difference between speech sound types for "appropriate" rating values, $F(2, 44) = 8.157$, $p = .001$, $\eta_p^2 = .270$; for "functionally helpful" rating values, $F(2, 44) = 14.673$, $p < .001$, $\eta_p^2 = .400$; and for "discernible" rating values, $F(2, 44) = 13.748$, $p < .001$, $\eta_p^2 = .385$.

For the multiple comparisons among the speech sound types, planned paired-samples t-tests were conducted. On the "appropriate" scale, participants rated the female/female-whisper ($M = 7.74$, $SD = 2.24$), $t(22) = 3.441$, $p < .05$, and the female/female-unavailable ($M = 7.22$, $SD = 2.65$), $t(22) = -3.602$, $p < .05$, significantly higher than the female/female-attenuated ($M = 5.22$, $SD = 2.58$). Also, on the "functionally helpful" scale, participants rated the female/female-whisper ($M = 8.39$, $SD = 1.56$), $t(22) = 5.273$, $p < .001$, and the female/female-unavailable ($M = 7.39$, $SD = 2.79$), $t(22) = -3.710$, $p = .001$, significantly higher than the female/female-attenuated ($M = 4.87$, $SD = 2.83$). In the same fashion, on the "discernible" scale, participants rated the female/female-whisper ($M = 7.30$, $SD = 2.12$), $t(22) = 4.320$, $p < .001$, and the female/female-unavailable ($M = 7.39$, $SD = 2.52$), $t(25) = -4.386$, $p < .001$, significantly higher than the female/female-attenuated ($M = 4.52$, $SD = 2.84$).

Subjective Preference. Fig. 4 shows the results of subjective preference scores. This figure suggests that participants favored the female/female-whisper over other conditions. Repeated measures ANOVA for "fun" rating values

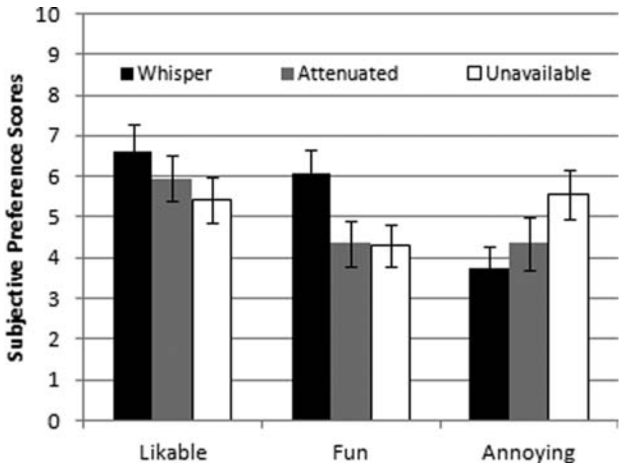


Fig. 4. Subjective preference scores for Study 2. Participants favored the female/female-whisper over other conditions.

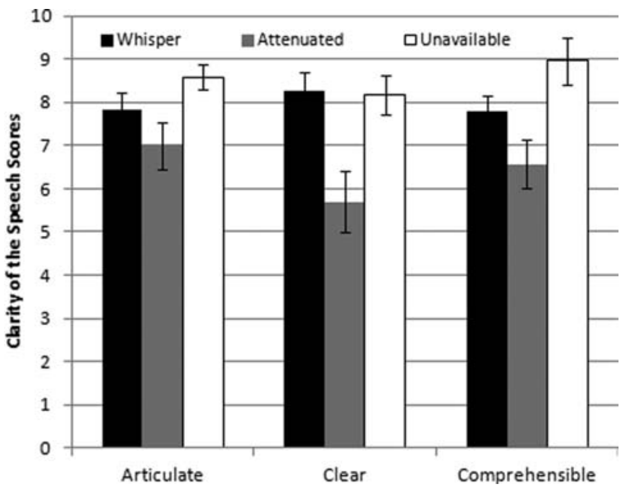


Fig. 5. Clarity of the speech scores for Study 2. Participants also favored the female/female-whisper and the female/female-unavailable.

showed a statistically significant difference between speech sound types, $F(2, 44) = 3.546$, $p < .05$, $\eta_p^2 = .139$. Paired samples t-tests showed significant results on the "fun" scale. The female/female-whisper ($M = 6.09$, $SD = 2.70$) showed higher scores than the female/female-attenuated ($M = 4.35$, $SD = 2.60$), $t(22) = 2.220$, $p < .05$, and the female/female-unavailable ($M = 4.30$, $SD = 2.53$), $t(22) = 2.138$, $p < .05$. However, there was no statistically significant difference between speech sound types for "likable" rating values, $F(2, 44) = 1.021$, $p > .05$, $\eta_p^2 = .044$. Even though "annoying" rating values showed only marginal difference, $F(2, 44) = 2.489$, $p = .095$, $\eta_p^2 = .102$, users tended to rate the female/female-unavailable condition very high on the "annoying" scale.

Clarity of the Speech. Fig. 5 shows the ratings of speech clarity. This figure shows that participants also favored the whisper and the unavailable conditions over the attenuated. Repeated measures ANOVA showed a statistically significant difference between speech sound types for "articulate" rating values, $F(2, 44) = 4.569$, $p < .05$, $\eta_p^2 = .172$; for "clear" rating values, $F(1.499, 32.969) = 8.182$, $p = .001$, $\eta_p^2 = .271$; and for "comprehensible" rating values, $F(2, 44) = 7.660$, $p = .001$, $\eta_p^2 = .258$.

For the multiple comparisons among the speech sound types, we conducted paired-samples t-tests again. On the "articulate" scale, participants rated the female/female-unavailable ($M = 8.57$, $SD = 1.41$), $t(22) = -3.116$, $p < .05$, significantly higher than the female/female-attenuated ($M = 7.00$, $SD = 2.65$). Also, on the "clear" scale, participants rated the female/female-whisper ($M = 8.26$, $SD = 2.12$), $t(22) = 2.837$, $p < .05$, and the female/female-unavailable ($M = 8.17$, $SD = 2.12$), $t(22) = -4.205$, $p < .001$, significantly higher than the female/female-attenuated ($M = 5.70$, $SD = 3.38$). In the same fashion, on the "comprehensible" scale, participants rated the female/female-unavailable ($M = 8.96$, $SD = 2.65$), $t(25) = -3.551$, $p < .05$, significantly higher than the female/female-attenuated ($M = 6.57$, $SD = 2.71$). Also, the female/female-whisper ($M = 7.78$, $SD = 1.83$), $t(22) = 2.089$, $p < .05$, was significantly higher than the female/female-attenuated.

In summary, as shown in Figs. 3 to 5, on the perceived performance and clarity ratings, the female/female-whisper and the female/female-unavailable were on identical levels and both of them led to higher scores than the female/female-attenuated. However, on the preference scale, the female/female-whisper was generally rated higher than the other two conditions.

6 STUDY 3: NOVEL MENU NAVIGATION

Study 3 was conducted to compare the whisper version of speech menus with a menu system that would be plausibly created when software developers just follow the typical defaults for programming menus. As a specific example, the default menu behavior in a GUI (in particular on the MS Windows platform, which is most commonly used by people with visual impairments) is often to simply skip over unavailable items when navigating through the menu with the cursor keys. A visually impaired user who hears the auditory menu produced by a screen reader would never know about those unavailable, grayed-out, visually skipped items. 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 set for participants to learn [see 6, 31 for the use of a block design in auditory menu navigation experiments]. See Table 1(b) for the new menu items [see 40 for the use of a new menu items].

6.1 Methodology

Participants, Stimuli, and Apparatus. Twenty-six undergraduate students (10 female; mean age = 20 years) participated in this study for partial credit in psychology courses. They reported normal or corrected-to-normal vision and hearing, signed informed consent forms, and pro-

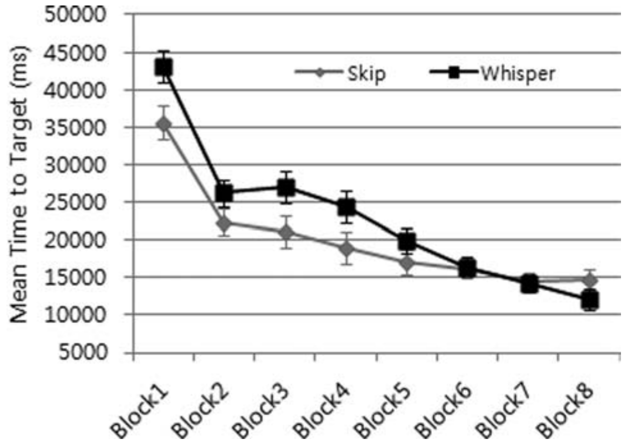


Fig. 6. Overall mean time to target (ms) for Study 3. Lower times indicate better performance. Error bars show standard error of the mean.

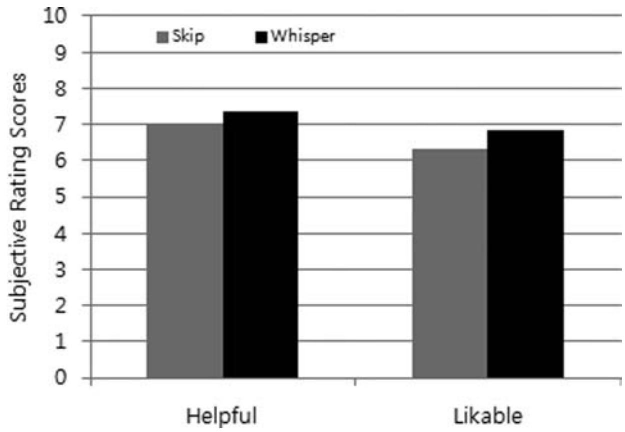


Fig. 7. Subjective rating scores for the helpfulness in understanding of the overall menu structure and likability for Study 3.

vided demographic details about age and gender. None had participated in Study 1 or Study 2.

Given that learning rates were the focus of this study, it was likely that any familiarity with the menu items could contaminate the results. Thus, Study 3 included a new 2-dimensional menu structure. The new menu had an identical layout to the Microsoft Word-like menu used in Study 2, however, its menu items were 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, Edit, and Insert, the new menu included unfamiliar titles such as Hills, Stars, and Islands (see Table 1(b)). The menu design maintained as much acoustic (or at least syllabic) similarity as possible, without any semantic similarity. For example, Insert became Islands, which both have two syllables, and Edit became Comet. As in Study 2, participants could not see the menu items (Fig. 2) as the sounds of the auditory menu were played. The apparatus was the same as in Study 2.

Procedure. As in Study 2, the task of the participant was to reach the target item in the auditory menu as fast as possible but without sacrificing accuracy. The target was randomly chosen among available items only and was visually displayed on the left side of the application. Each condition contained 8 blocks and each block included 15 trials. In every block, 30% of menu items were randomly designated as unavailable. Study 3 used a between-subjects design in order to look at learning effects. Thus, there were two between-subjects conditions based on speech type: skipping unavailable items and whispering unavailable items. All participants experienced the same procedure for each block, regardless of the assigned speech conditions. After completing all eight blocks, participants filled out a short questionnaire. An 11-point Likert-type scale also was used for the self-rated levels of "functional helpfulness" in order to understand the entire menu structure and "likability" with regard to using speech type for unavailable items.

6.2 Results

The results are depicted in Figs. 6 and 7. In particular, as shown in Fig. 6, the mean time to target (i.e., "search time," in ms) of the skip condition was lower than that of the whisper condition in early blocks. However, in Block 6, the whisper condition reached the same level and after that, search time of the whisper condition became lower than that of the skip condition. These results were analyzed with a 2 (Speech type) × 8 (Block) repeated measures ANOVA, which revealed a statistically significant difference between blocks, $F(7,\ 168) = 60.554$, $p < .001$, $\eta_p^2 = .72$. Overall, there was no difference between speech types, $F(1,\ 24) = 3.827$, $p = .062$, $\eta_p^2 = .14$, whereas 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, $t(24) = .961$, $p = .346$. For the subjective rating data as shown in Fig. 7, neither "helpfulness" scores, $t(24) = -.493$, $p = .627$, nor "likability" scores, $t(24) = -.613$, $p = .546$ produced any statistically reliable differences.

7 GENERAL DISCUSSION

This paper introduced whispered and attenuated TTS sounds as an alternative design for the unavailable items in auditory menus. The use of these new implementations for unavailable items was compared to the typical current implementation (appending "unavailable") and to the common default policy for visual menu implementation (skipping unavailable items when manipulating arrow keys). Overall, results were subjectively (in Study 2) and objectively (in Study 3) in favor of the "whisper" approach although sound quality needs to be improved. The result that the whisper version showed lower response time compared to the skipping version after a moderate amount of practice might imply that the whisper version is more effective than the skipping version from the long-term perspective.

In more detail, Study 1 examined how basic properties of the speech sound are preferred by sighted and visually impaired users. Both groups favored the female voice over the male voice. This confirms previous research in which a female voice showed more positive results than a male voice [39, 41, 42]. While visually impaired people accepted changing the voice gender to signal an unavailable item, sighted participants clearly preferred the same gender for both available and unavailable items in a menu. This was because the gender metaphor was not understood as being related to the availability of menu items. Admittedly, since Study 1 employed only ten sighted participants and four visually impaired participants, the preference for voice property could still depend on contexts such as gender of target users, goal, and domain of services [37]. That is why the subsequent studies employed more quantitative methods to investigate those issues.

Study 2 showed that the use of whisper for unavailable items was subjectively preferred by participants. On the perceived performance and the clarity of the speech, whisper and saying "unavailable" gained similar high scores, but on the likability scale, whisper was clearly preferred over saying "unavailable." Moreover, the annoyance score of saying "unavailable" was higher than for the other methods. This is important, given that JAWS employs the "unavailable" method, which was shown to be less favorable here. It is also important to point out that Voice Over in Mac OS X follows the same path as JAWS, but instead of speaking "unavailable," OS X says "dimmed." In addition to likely being less-well rated (it is in the same family as "unavailable"), the "dimmed" term is a spoken description of the visual interface. This is less useful to the visually impaired listeners of an auditory menu because they might have no idea what "dimmed" refers to. Similar possible phrases like "grayed-out" are based on the visual rendering choices and are not appropriate either. The auditory rendering of a menu should convey the menu item state or function and not just describe its visual rendering.

Even if the addition of a word like "unavailable" is avoided, there are plenty of other ways to auditorily render the concept of a menu item being unavailable. Some will be better than others. Using just the naïve analogy to visual interfaces, adopting lower loudness might be considered as a straightforward way to convey an unavailable item, given that gray text is commonly used to represent unavailable items, in cases where available items are shown in black text. Moreover, since shorter is better in auditory menus, applying lower volume for unavailable items might seem to be more intuitive and better than adding "unavailable" or "dimmed" to the item label. However, that is not the only way to convey the concept of being unavailable. An examination of everyday listening provides a plausible application of whispered speech as another way to denote an unavailable item. It is only through empirical evaluations that one can determine which of these two methods (attenuating and whispering) leads to better preference and performance.

Study 3 showed the difference between a typical implementation of menus (skipping) and an alternative design approach with respect to learning a new auditory menu. In some operating systems and software platforms, the default for the unavailable menu items when navigating with cursor keys is to skip them. If a developer directly applies this interface convention to auditory menus, it would be problematic because visually impaired users would not hear the "missing" menu items, unlike the sighted user who can scan over the grayed-out menu item and can learn where they are for future use. According to Norman [43], when users create and represent the system model, a "cognitive layout" may be used (regardless of whether it is described as scripts, metaphors, or production rules). The problem is that too often menus hide the organization and structure of the tree rather than explicitly using it to the benefit of the interface [43]. The skipping of unavailable items can initially obtain efficiency in auditory menus, but in the long run it seems to prevent users from forming the desirable cognitive layout for the entire menu structure. Moreover, usability depends not only on navigation time but also on information transmission between the system and user. Once users hear the order of the menu items, it can affect their cognitive layout. If users hear a different order of menu items later because some are now spoken (or silent), it might diminish trust and familiarity with the system. It is interesting to note that the spoken menus in VoiceOver on Mac OS X, criticized earlier for using the visual word "dimmed" to identify unavailable menu items to visually impaired users, does a good job when it comes to the issue of skipping items. When VoiceOver is turned off (the usual case), unavailable items are skipped, but when VoiceOver is turned on, unavailable items are not skipped. This is a nicely adaptive feature of that particular auditory menu system.

The difficulty of learning a menu with a "skipping" auditory menu is similar to what can happen with "adaptive" and also "collapsing" menus in a GUI. Adaptive menus are sorted based on usage frequency and recent access. Because of the uncontrollable and unexpected changes in the interface, the adaptive menu has been found to be significantly slower than a static menu [40]. Theories of human problem solving suggest that the understanding and representation of the problem domain aids in the solution. To this end, good user interface design should convey a sense of meaning and engage schemata that lend themselves to solutions of the tasks being performed [43]. It is not surprising, then, that the results with auditory menus presented here show that letting users obtain a consistent mental structure outperforms the adaptive (or skipping) menu.

8 CONCLUSION

This project attempted to enhance the speech-based auditory menu with a simple idea, using "whisper" to render unavailable menu items. We focused on the assertion that designers should go beyond a naïve translation from text into speech when creating auditory systems. However, as mentioned earlier, the speech-based menus research has to be accompanied with adequate non-speech sounds, just as graphical user interfaces adopt ample non-text components. Through a multi-faceted design effort, designers should be able to create subjective satisfaction as well as usability and universal accessibility of the devices and ultimately provide essential information to various user populations efficiently and pleasantly.

9 REFERENCES

[1] G. Kramer, "An Introduction to Auditory Display," in G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*, MA, pp. 1–77 (Addison-Wesley, 1994).

[2] M. A. Nees and B. N. Walker, "Auditory Interfaces and Sonification," in Stephanidis Constantine (Ed.), *The Universal Access Handbook*, New York, pp. 507–521 (CRC Press Taylor & Francis, 2009).

[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," presented at the *NordiCHI* 2008, Lund, Sweden (2008).

[5] D. Palladino and B. N. Walker, "Navigation Efficiency of Two Dimensional Auditory Menus Using Spearcon Enhancements," presented at the *Annual Meeting of the Human Factors and Ergonomics Society*, New York, USA (2008).

[6] M. Jeon and B. N. Walker, "Spindex (Speech Index) Improves Auditory Menu Acceptance and Navigation Performance, *ACM Transactions on Accessible Computing*, vol. 3, No. 3, 10:1–26 (2009).

[7] W. W. Gaver," Auditory Icons: Using Sound in Computer Interfaces," Auditory Icons: Using Sound in Computer Interfaces," *Human-Computer Interaction*, vol. 2, pp. 167–177 (1986).

[8] E. Mynatt and W. Edwards, "Mapping GUIs to Auditory Interfaces," presented at the *5th Annual ACM Symposium on User Interface Software and Technology, Monteray*, CA, USA (ACM, 1992).

[9] T. V. Raman, *Auditory User Interfaces: Toward the Speaking Computer*, Boston (Kluwer Academic Publishers, 1997).

[10] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch, "Earpod: Eyes-Free Menu Selection Using Touch Input and Reactive Audio Feedback," presented at the *SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA (ACM, 2007).

[11] B. D. Davison and B. N. Walker, "AudioPlusWidgets: Bringing Sound to Software Widgets and Interface Components," presented at the *14th International Conference on Auditory Display*, Paris, France (2008).

[12] E. P. Glinert and B. W. York, "Computers and People with Disabilities," *ACM Transactions on Accessible Computing*, vol. 1, pp. 1–7 (2008).

[13] P. Yalla and B. N. Walker, "Advanced Auditory Menus," *Georgia Institute of Technology GVU Center Technical Report* (2007).

[14] F. James, "Lessons from Developing Audio HTML Interfaces," presented at the *Annual ACM Conference on Assistive Technologies, Marina del Rey*, CA, USA, pp. 27–34 (ACM, 1998).

[15] S. Morley, H. Petrie, and P. McNally, "Auditory Navigation in Hyperspace Design and Evaluation of a Non-Visual Hypermedia System for Blind Users," presented at the *Annual ACM Conference on Assistive Technologies*, Marina del Rey, CA, USA (ACM, 1998).

[16] C. Asakawa and T. Itoh, "User Interface of a Home Page Reader," presented at the *Annual ACM Conference on Assistive Technologies*, Marina del Rey, CA, USA (ACM, 1998).

[17] I. J. Pitt and A. D. N. Edwards, "Improving the Usability of Speech-Based Interfaces for Blind Users," presented at the *Annual ACM Conference on Assistive Technologies*, Vancouver, BC, Canada (ACM, 1996).

[18] J. Thatcher, "Screen Reader/2 Access to OS/2 and the Graphical User Interface," presented at the *First Annual ACM Conference on Assistive Technologies*, Marina Del Rey, California, USA, pp. 39–46 (ACM, 1994).

[19] A. Kehoe and I. Pitt, "Designing Help Topics for Use with Text-to-speech," presented at the *24th Annual ACM International Conference on Design of Communication, Myrtle Beach*, SC, USA, pp. 157–163 (ACM, 2006).

[20] W. W. Gaver, "The SonicFinder, a Prototype Interface that Uses Auditory Icons," *Human-Computer Interaction*, vol. 4, pp. 67–94 (1989).

[21] E. Mynatt, "Transforming Graphical Interfaces into Auditory Interfaces for Blind Users," *Human-Computer Interaction*, vol. 12, pp. 7–45 (1997).

[22] E. Mynatt and G. Weber, "Nonvisual Presentation of Graphical User Interfaces: Contrasting Two Approaches," presented at the *SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, pp. 166–172 (ACM, 1994).

[23] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction*, vol. 4, pp. 11–44 (1989).

[24] S. A. Brewster, "Using Non-speech Sound to Overcome Information Overload," *Displays*, vol. 17, pp. 179–189 (1997).

[25] S. A. Brewster and P. G. Cryer, "Maximising Screen-Space on Mobile Computing Devices," presented at the *SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA, USA, pp. 224–225 (ACM, 1999).

[26] G. Leplatre and S. A. Brewster, "Designing Non-Speech Sounds to Support Navigation in Mobile Phone Menus," presented at the *6th International Conference on Auditory Display*, Atlanta, GA, USA, pp. 190–199 (2000).

[27] B. N. Walker, A. Nance, and J. Lindsay, "Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus," presented at the *International Conference on Auditory Display*, London, UK, pp. 95–98 (2006).

[28] D. Palladino and B. N. Walker, "Learning Rates for Auditory Menus Enhanced with Spearcons Versus Earcons," presented at the *International Conference on Auditory Display*, Montreal, Canada, pp. 274–279 (2007).

[29] T. Dingler, J. Lindsay, and B. N. Walker, "Learnability of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons, and Speech," presented at the *14th International Conference on Auditory Display*, Paris, France (2008).

[30] D. Palladino and B. N. Walker, "Efficiency of Spearcon-Enhanced Navigation of One Dimensional

Electronic Menus," presented at the *International Conference on Auditory Display*, Paris, France (2008).

[31] B. N. Walker and A. Kogan, "Spearcons Enhance Performance and Preference for Auditory Menus on a Mobile Phone," presented at the *5th International Conference on Universal Access in Human-Computer Interaction at HCI International* 2009, San Diego, CA, USA (2009).

[32] M. Jeon, B. K. Davison, M. A. Nees, J. Wilson, and B. N. Walker, "Enhanced Auditory Menu Cues Improve Dual Task Performance and Are Preferred with In-Vehicle Technologies," presented at the *First International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Essen, Germany, pp. 91–98 (2009).

[33] S. A. Brewster, "The Design of Sonically-Enhanced Widgets," *Interacting with Computers*, vol. 11, pp. 211–235 (1998).

[34] P. Yalla and B. N. Walker, "Advanced Auditory Menus: Design and Evaluation of Auditory Scroll Bars," presented at the *Annual ACM Conference on Assistive Technology*, Halifax, Nova Scotia, Canada (ACM, 2008).

[35] L. Ludwig, N. Pincever, and M. Cohen, "Extending the Notion of a Window System to Audio," *Computer*, vol. 23, pp. 66–72 (1990).

[36] P. Shajahan and P. Irani, "Manipulating Synthetic Voice Parameters for Navigation in Hierarchical Structures," presented at the *11th International Conference on Auditory Display*, Limerick, Ireland, pp. 49–54 (2005).

[37] M. Jeon, S. Gupta, and B. N. Walker, "Advanced Auditory Menus II: Speech Application for Auditory Interface," *Sonification Lab Technical Report*, Georgia Institute of Technology (2009).

[38] M. Helander and T. K. Landauer, *Handbook of Human-Computer Interaction*, New York, USA (Elsevier Science Inc, 1997).

[39] L. Gong and J. Lai, "Shall We Mix Synthetic Speech and Human Speech? Impact on Users' Performance, Perception, and Attitude," presented at the *SIGCHI Conference on Human Factors in Computing Systems*, Seattle, WA, USA, pp. 158–165 (ACM, 2001).

[40] L. Findlater and J. McGrenere, "A Comparison of Static, Adaptive, and Adaptable Menus," presented at the *SIGCHI Conference on Human Factors in Computing Systems* (ACM, 2004).

[41] B. Thorne, C. Kramearae, and B. Henley, *Language, Gender and Society*, MA, USA (Newbury House, 1983).

[42] M. K. Margulies, "Male-Female Differences in Speaker Intelligibility: Normal Versus Hearing Impaired Listeners," presented at the *Speech Communication Papers, 97th Meeting of the Acoustical Society of America*, New York (1979).

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

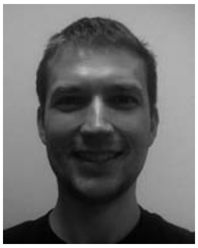
[44] S. A. Brewster, "Using Non-Speech Sounds to Provide Navigation Cues," *ACM Transactions on Computer-Human Interaction*, vol. 5 issue 3, pp. 224–259 (1998).

[45] S. A. Brewster, V. P. Raty, andA. Kortekangas, "Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy," presented at the *HCI'96* (1996).

THE AUTHORS

M. Jeon

S. Gupta



B. K. Davison

B. N. Walker

Myounghoon "Philart" Jeon is an assistant professor in the Department of Cognitive and Learning Sciences at Michigan Tech. His research areas encompass auditory displays, affective computing, assistive technology, and automotive interface design. His research has yielded more than 60 publications across various journals and conference proceedings. He received his Ph.D. from Georgia Tech in 2012. His dissertation focused on the design of in-vehicle emotion regulation interfaces using auditory displays. Before attending Georgia Tech, Philart worked at LG Electronics Design Center and was responsible for all of their sound designs.

Siddharth Gupta completed his MS in human-computer interaction at Georgia Tech, during which he assisted on research projects in the Sonification Lab. He is now working in user experience and user interaction design.

Ben Davison is a human centered computing Ph.D. student in the School of Interactive Computing at Georgia Tech. His interests include studying the usage, development, and deployment of auditory interfaces in everyday computing situations, particularly in K-12 classrooms.

Bruce N. Walker holds appointments in psychology and computing at Georgia Tech. His Sonification Lab studies non-traditional interfaces for mobile devices, auditory displays, and assistive technologies. Dr. Walker earned a Ph.D. in psychology from Rice University and has authored over 100 peer reviewed journal articles and conference papers. He has also worked for NASA, governments, private companies, and the military. Originally from Canada, Bruce now makes his home in Atlanta with his wife and two young children.