

Running head: EFFICIENCY OF SPEARCON-ENHANCED NAVIGATION

Efficiency of Spearcon-Enhanced Navigation of One Dimensional Electronic Menus

Dianne K. Palladino

Georgia Institute of Technology

A Thesis Submitted in Partial Requirement for the
Research Option in Psychology

Faculty Advisor: Dr. Bruce Walker

Second Reader: Dr. Gregory Corso

December 4, 2007

Abstract

This study simulated and compared cell phone contact book menu navigation using combinations of both auditory (text-to-speech and spearcons) and visual cues. A total of 127 undergraduates participated in a study that required using one of five conditions of alphabetically listed menu cues to find a target name. Participants using visual cues (either alone or combined with auditory cues) outperformed those using only auditory cues. Performance was not found to be significantly different among the three auditory only conditions. When combined with visual cues, spearcons improved navigational efficiency more than both text-to-speech cues and menus using no sound, and provided evidence for the ability of sound to enhance visual menus. Research results provide evidence applicable to efficient auditory menu creation.

Keywords: Auditory Menus, Spearcons, Cell Phones, Accessibility, Visually Impaired, Usability, Electronic Interfaces

Efficiency of Spearcon-Enhanced Navigation of One Dimensional Electronic Menus

Auditory displays are representations of information often using non-speech sounds (Walker & Kramer, 2004). Various types of auditory displays have been studied as either enhancements or primary display modalities for the navigation systems of small electronic devices, such as cell phones and Personal Digital Assistants (PDAs). Auditory displays are of benefit to the visually impaired because of their potential to make the latest technology more readily accessible (Nees & Walker, In Press). This is true for the permanently visually impaired, and for those with limited visibility due to environmental conditions or circumstances. For example, auditory menus on electronic devices can allow firefighters to access important information in a smoke-filled room. Since users interface with most devices through the use of various types of visual menus, it is important to determine the auditory equivalents to interface navigation structures that will allow the greatest level of efficiency, ease of learning, and dissemination of navigational information.

There are four primary auditory menu cues that have been previously suggested as feasible: regular speech, auditory icons (Gaver, 1986), earcons (Blattner, Sumikawa, & Greenberg, 1989), and most recently, spearcons (Palladino & Walker, 2007; Walker, Nance, & Lindsay, 2006). Auditory icons are representations of the sound naturally produced by or associated with a menu item, and earcons are hierarchical representations of menu items created using musical elements. Spearcons are a form of condensed speech with a unique, fingerprint-like acoustical representation. All of these auditory menu cues have their advantages and limitations, and the research that is building on display technology strives to find the optimum auditory enhancement in terms of efficiency, learning rates, and usability (Palladino & Walker, 2007; Walker & Kramer, 2004; Walker et al., 2006; Yalla & Walker, 2007).

Auditory Menus

Auditory menus allow navigation of available functionality on electronic interfaces by using sound (Yalla & Walker, 2007). Using sound to enhance menus on electronic systems, whether small electronic devices or desktop systems, widens potential uses for the devices, and increases the number of potential users. In its simplest form, an auditory menu typically consists of electronic Text-To-Speech (TTS) conversion of the words or phrases included in the menu hierarchy. Users of auditory menus typically navigate the menu using arrow keys provided on the device, and menu items are presented using sound. Sound alone or sound combined with visual menu cues can be used to assist the user with navigation. In most cases, when the user lands on the desired item, a button such as the “enter” key on the device or keyboard is used to select the item.

Auditory enhancements to a menu are sometimes prepended with cues to assist in efficient navigation. Since speech alone is relatively slow and inefficient, the goal of these cues is to provide faster recognition of the menu item in question and to create a greater navigational efficiency. It is possible for the auditory cue (or a portion of the cue) to be sufficient information for the user to determine if the current location on the menu is the desired destination or if it is necessary to navigate further. The unaltered TTS of the menu item can be (but does not necessarily need to be) included after the cue, so that if the user has any confusion about the meaning of the cue, the entire word or phrase can be heard to verify menu location. It is possible that with moderate usage of the auditory cues, the original TTS phrase will be used less frequently, and the option to remove the TTS phrases completely and utilize solely the cues to navigate the auditory menu is a potential option for users. If the auditory cues take less time

to perceive than the original TTS phrases, then once the TTS is no longer needed navigation should become more efficient for the user.

The transient nature of sound causes several unique usability challenges for designers of auditory menus. Unlike the perception of visual interfaces, in which the average person can read 250-300 words per minute, average individuals can comprehend 130-150 words per minute comfortably ("Words per minute," 2007). For this reason, a primarily auditory cue can take longer for users to perceive. This difference in perceptual ability can make navigation of auditory menus slower relative to visual navigation. A second challenge is location awareness. Users need to know their current position in an auditory menu and be able to discern the fastest path to reach another position in the menu (Leplatre & Brewster, 2000). Unlike a visual menu, which can be scanned quickly to determine the current position relative to the hierarchy of the menu, an auditory menu can require a considerable amount of the user's working memory to maintain the same information. The third challenge to auditory menu design is the ability for the user to learn the auditory cues quickly. A shorter learning curve will decrease the amount of time it takes for the user to take advantage of the functionality of the phone in the shortest amount of time possible.

Evidence for the most feasible auditory menu enhancement cue type has been provided by two previous experiments. Recently, Walker, Nance, and Lindsay (2006) found that spearcons outperform auditory icons, earcons, and speech alone in time to target efficiency. An experiment completed by Palladino and Walker (2007) decided not to consider auditory icons as cues on electronic menus for functionality navigation due to the lack of natural sounds available for items typically included on these menus. This study compared rates of learning associations between earcons and spearcons and the items that they represent and found that earcons were

significantly more difficult and frustrating to learn than spearcons. Spearcons were found to be the most favorable cue type with respect to efficiency and learning rate, and the current study will collect evidence about the usability of the spearcon alone.

Auditory Icons and Earcons

Although auditory icons (Gaver, 1986) and earcons (Blattner et al., 1989) are not empirically investigated in this experiment, a brief explanation of their composition and their advantages and disadvantages is worthwhile. Both have been proposed in the past as solutions to auditory menu challenges but have disadvantages that have been at least partially overcome by the spearcon.

An auditory icon is a representation of the natural sound produced by an item (Gaver, 1986). From infancy we learn that cows “moo” and that cats “meow,” and there are a large number of items for which we have a natural automatic association between the sound and the item. For certain words, such as animals, musical instruments, and people sounds, a direct connection between the sound and the word is obvious to most people.

A problem arises when designers attempt to use auditory icons to represent actions or objects that are intangible. For example, what would be the auditory icon for “Save to Desktop” or “Options” on a typical electronic interface? There have been somewhat successful attempts to create auditory icons for some computer-related functions, as illustrated in the sound associated with Microsoft’s Recycle Bin. Although this is not a natural sound, it seems somewhat logical. Most people agree that the sound is like a crumpled up piece of paper being thrown into a metal waste paper basket. What happens when a computer user tries a Mac, however? The sound for the Trash icon on the Mac interface defaults to a completely different sound. The task of deciding the sound most appropriate to represent the same function is obviously not in

agreement. If the item represented does not make a natural sound, it is difficult to reach a consensus because the auditory icon needs to become more metaphorical (Walker & Kramer, 2004). It then is less useful due to conflicting opinions of the most appropriate auditory representation for the item. This lack of ecological validity to most electronic menu items makes an auditory icon an undesirable option for creating electronic menu enhancements.

Earcons (Blattner et al., 1989) are systematically produced representations of menu items using musical elements and can be created by varying frequency, timbre, tempos, rhythmic patterns, or combinations of any aspect of music to represent unique items on a menu. Guidelines suggested by Hereford and Winn (1994) suggest that earcons are most effective when each item represented in a group differs in as many musical elements as possible from the other members of the group. Earcons can be created to represent a hierarchy of items in a menu system by combining musical elements systematically (Brewster, Raty, & Kortekangas, 1996; Brewster, Wright, & Edwards, 1993; Leplatre & Brewster, 2000).

To create a 5-row by 5-column hierarchical menu system, a designer might consider using a different timbre of sound (piano, trumpet, flute) to represent every item in each column, and a different overlying rhythmic pattern (two quarter notes on snare drum, eighth notes on a cowbell, triplets on a wood block) to represent each row. An item on the menu grid would be represented by the simultaneous play of the two musical elements of the row and column for that particular grid position. Once the user has memorized the order of each musical element for each row and column, it can be an effective way for users to determine their position in a particular menu hierarchy, and participants in prior studies have had success in identifying and understanding this hierarchical information (Brewster et al., 1996; Leplatre & Brewster, 1998).

In 2003, Vargus and Anderson (2003) combined earcons with speech to find that the combination increased efficiency of menu navigation without additional burden on the user.

Advantages of earcons include their usefulness in providing hierarchical menu information and their ability (unlike auditory icons) to be applied to menus containing any type of information. Earcon hierarchy can be a disadvantage, however, because the rigid nature of the menu setup makes it difficult to add or subtract an item within the hierarchy. For example, if an item is added to the fourth column, second row of the grid, it is debatable whether it would make more sense to move everything else in that column down a row and change its earcon representation or to create an entirely new row and leave that row blank in the other columns. It is not clear which (if any) of these two solutions would be the most effective. As Walker et al. (2006) have stated, the arbitrary nature of the earcon is considered both its strength and its weakness. Additionally, Palladino and Walker (2007) found that it is difficult for users to learn earcon/word associations, and this difficulty can cause frustration for the user. Auditory enhancement cues are intended to decrease user frustration and annoyance (Brewster & Crease, 1999) as well as to increase navigation efficiency, but earcons seem to fall short on these criteria (Palladino & Walker, 2007; Walker et al., 2006). For this reason, earcons are not considered in this study as possibilities for auditory cues.

Spearcons

A spearcon (Walker et al., 2006) is created by compressing a spoken phrase (created either by a TTS generator or by recorded voice) without modifying the perceived pitch of the sound. Some speech is compressed to the point that it is no longer comprehensible as a particular word or phrase. Walker et al (2006) compared the spearcon to a fingerprint because each unique word or phrase creates a unique sound when compressed that distinguishes it from

other spearcons. After a brief learning session, the associations between a spearcons and their related words or phrases are easy to recognize (Palladino & Walker, 2007).

In order to create spearcons for use as auditory menu cues, a sound file containing the speech must first be created by using TTS generation software or by simply recording a voice speaking the words or phrases. The spearcon is created from that file, and pre-pended to the original TTS file in the form of a “cue.” A small duration (250 ms) of silence is inserted between the spearcon cue and the original word or phrase. More information on spearcon creation is provided in the methods section of this document.

Spearcons are naturally briefer than the words and phrases they represent, are fast and easy to produce, and can be easily inserted into any menu structure in any position because they are direct representations and do not depend upon hierarchical positioning in a menu. Although spearcons do not provide natural hierarchical information to the user, such as those that are inherent in hierarchical earcons (Blattner et al., 1989), it would be possible to create hierarchical information for the user by implementing some sort of augmentation to the spearcons, such as adding volume cues or pitch cues to provide position information to the user. This addition may not be absolutely necessary for efficiency of navigation, however, as shown by Walker et al (2006), who found that spearcons resulted in significantly more efficient navigation than hierarchical earcons, even when using spearcons with no hierarchical information.

In a study performed by Palladino and Walker (2007), spearcons were found to be significantly easier to learn than earcons when users were trained on associations with the words and phrases they represented. Half of the participants were trained and tested on spearcon associations, and the other half of participants were tested on earcon associations. Participants found spearcon/word associations easier to learn and the learning process for earcon associations

more arduous and frustrating. With these advantages for spearcons over other enhancement types, the focus for auditory menu enhancement research has narrowed to comparing the benefits of using spearcons as pre-pended cues to TTS to using TTS alone in an auditory menu system. This comparison is the focus of the current study.

This experiment includes conditions with visual menu cues, either alone or in combination with one of the auditory representations. For an individual with normal vision, the conditions with visual cues are expected to enhance the speed to the target menu item. Visual cues, however, may not be useful to visually impaired individuals, and this experiment will focus more on the length of an auditory stimulus and its effect on the time it takes to reach a requested target item on a menu. It is of interest, however, to compare the visual and auditory stimuli to have a basis of comparison for future planned studies with visually impaired individuals.

This experiment compares navigation rates of a simulated cell phone contact book created with various combinations of visual and auditory elements. It compares auditory cues created with TTS Only, TTS with a spearcon enhancement cue, and no audio. Each auditory condition also is tested combined with a visual menu. The hypothesis of this study is that conditions with visual menus will outperform those with only auditory cues, and that spearcon enhancement prepended to the TTS will significantly outperform the other auditory conditions.

Method

Participants

A total of 127 undergraduates (55 men and 72 women, mean age = 19.74) with normal or corrected-to-normal hearing and vision volunteered to participate for extra credit in psychology courses. English was the native language of all participants. There were either 25 or 26 participants in each condition.

Design

This experiment used a between-subjects design. The first independent variable was sonification type (TTS Only, Spearcon Cue + TTS, or No Audio), and the second independent variable was visual cue (On or Off). The condition in which auditory and visual cues are simultaneously off is, obviously, not a valid condition, which leaves five appropriate experimental conditions. The dependent variable was average time to selection of target menu item.

Materials

Participants were tested with a computer program written with Macromedia Director MX and Lingo on a Windows XP platform listening through Sennheiser HD 202 headphones. They were given an opportunity at the beginning of the experiment to adjust volume for personal comfort.

A random name generator (<http://www.xtra-rant.com/gennames/>) created the 50 names used for the contact book stimuli. Auditory TTS was generated for all of the names using the AT&T Labs, Inc. Text-To-Speech Demo program.

Spearcons were created for the TTS conversion of each name by running them through a MATLAB algorithm (See Appendix) that compressed each name logarithmically while maintaining original sound frequency. Logarithmic compression is currently considered the preferred compression technique for creating spearcons because it compresses longer phrases more than shorter phrases. Shorter words (particularly those that are monosyllabic) tend to sound more like “clicks” if they are compressed too much and become indistinguishable. Since they are very short to begin with, the advantage of compression of very short words is much less than for a longer phrase. Phrases of several words or syllables can be compressed at a much

higher ratio since they contain a higher level of language context. Higher compression makes the spearcons shorter and more efficient without losing the context needed to identify them as unique.

Stimuli for the Spearcon Cue + TTS condition were created by using Audacity software to prepend the cue to the TTS with a 250 ms post-cue interval between them. Visual stimuli consisted of a list of names displayed to the participant in 30-point text. Names were displayed in alphabetical order by first name in a “window” five at a time, and the list scrolled downward or upward based upon the key presses of the participant. For both the auditory and visual components, if the participant reached the bottom of the list, the list did not wrap around. Although this design does not simulate the exact functionality of the screen on a cell phone or PDA contact book menu, this feature is necessary to control for distance to the target name on the list. As the focus changed to each menu item, auditory and visual menu cues were presented simultaneously when the condition required both modes of display.

Procedure

A simulated cell phone contact book menu was presented that contained items constructed with auditory, visual, or both representations. The contact book consisted of 50 names (first and last) in alphabetical order by first name. The up and down arrow keys were used to navigate the menu, and the enter key was used to select the appropriate item. Participants were assigned to one of five conditions. Two conditions provided only auditory cues for each menu item: one with TTS cues and one with spearcons pre-pended to the TTS. The other three conditions all combined visual cues with sound: one with no auditory cues, one with TTS, and one with spearcon cues pre-pended to the TTS. Two lists of 25 of the 50 names (first name and last name) were alternated five times for a total of 10 blocks of 25 trials. All

participants experienced the same procedure for each block, regardless of the assigned menu display condition. The order of appearance of the list halves was counterbalanced among subjects.

Participants first saw a brief instruction screen that taught them about menu navigation and that the required task was to find the requested name on the menu as quickly as possible without sacrificing accuracy. The participant was then presented with a name (e.g. “Allegra Seidner”) on the top of the screen that indicated the target name. When the first up or down key was pressed, the timer started. Participants navigated through the menu system to find the assigned target name and hit the “enter” key to indicate selection of the requested target. Hitting the enter key recorded the end time. The time to target response was calculated by subtracting the start time from the end time. Each participant immediately was shown the next target name, and the procedure was repeated for all 25 names in the block. Participants were then shown a screen that indicated that the next block of 25 trials was about to start. Each of the nine subsequent blocks proceeded in the exact same way. After the 10th block, participants filled out a brief demographics questionnaire regarding age, gender, ethnicity, musical training information, and a free-format opportunity was also provided to comment on their experience with the experiment and any strategies they may have used to complete the task.

Results

An alpha level of .05 was used for all statistical analysis. After disqualifying 1.53 % of trials due to incorrect item selection (37 in Visuals Off/Spearcons condition, 16 in Visuals Off/TTS condition, 112 in Visuals On/No Sound condition, 96 in Visuals On/TTS condition, and 36 in Visuals On/Spearcons condition), a total of 31272 trial records remained with which to perform the data analysis. A one-way ANOVA was performed on the data to check for

significant differences among the different experimental conditions. As predicted, overall performance on all conditions including visual cues were significantly faster than those including only auditory cues $F(1, 31270) = 4963.665, p < 0.001$. This result is illustrated in Figure 1, which plots all mean times to target for each condition in each block of the experiment.

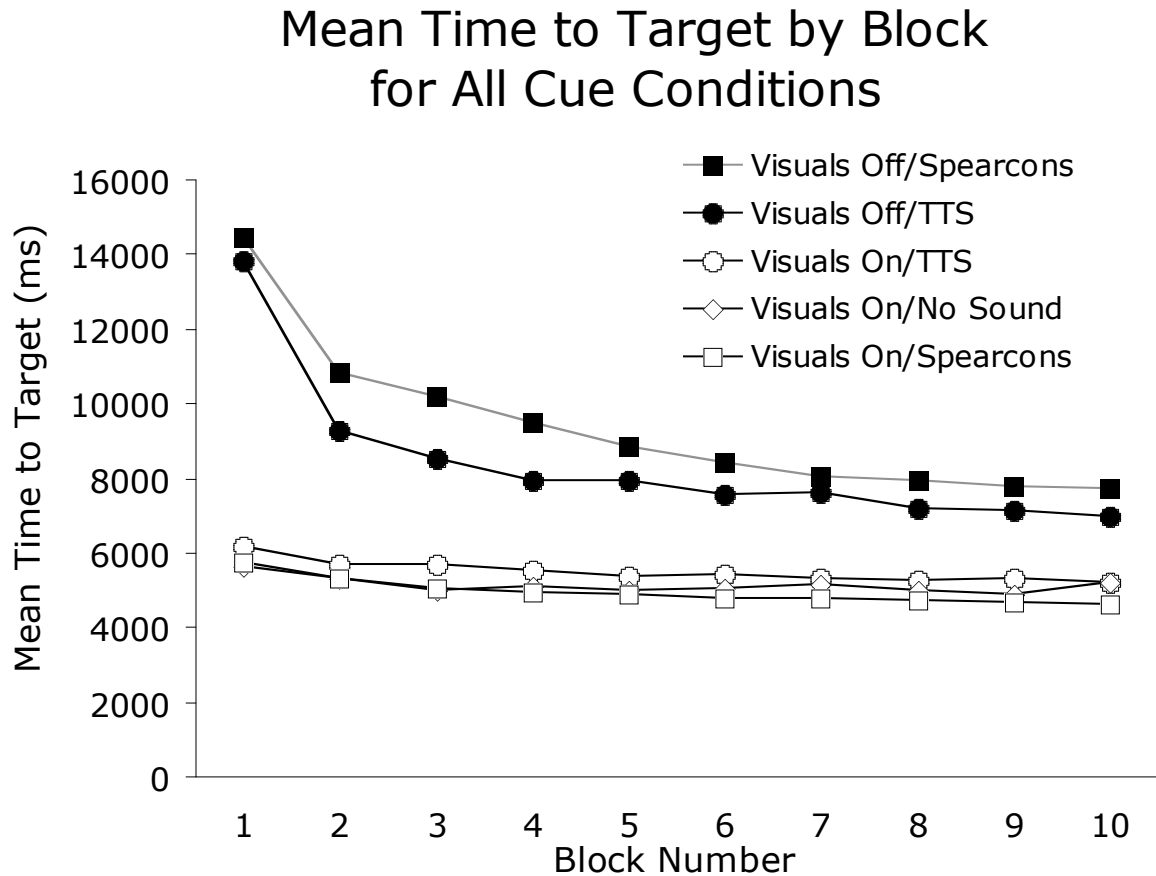


Figure 1. Mean time to target in milliseconds for all conditions over all blocks. Learning effects were found for all conditions, and were most significant for the two conditions that did not use visual cues. The TTS condition outperformed the spearcon condition in auditory-only conditions, but spearcon conditions outperformed both of the conditions using visual cues consistently, although not significantly. The Visuals On/Spearcons condition outperformed the condition that did not use auditory cues, and provided evidence that auditory cues may enhance the performance of menu navigation if used in conjunction with visual information.

The plotlines for the auditory-only conditions show consistently longer mean times to target throughout the blocks than the conditions that contained both visual and auditory cues. A Tukey honestly significant difference analysis of Block 10 data for each condition found no significant difference between any of the three conditions including visual cues at the $p < 0.01$ level. Auditory-only cues compared to any of the other four conditions, including the other auditory-only condition, were found to lead to significantly different performance, all providing a Tukey value less than .01. It is clear from the graph, however, that even though the differences between the conditions using auditory-only and auditory and visual cues in Block 10 are significant, there is much less of a difference between the auditory-only and visual conditions than existed in the first block of the experiment.

Figure 2 illustrates the mean time to target for the five categories in the first and tenth blocks. There was a significant difference in the means collapsed over all conditions between the first ($M = 9253$, $SD = 8890$) and tenth ($M = 5979$, $SD = 3944$) blocks $F(1, 6273) = 355.635$, $p < 0.001$, indicating learning effects across blocks.

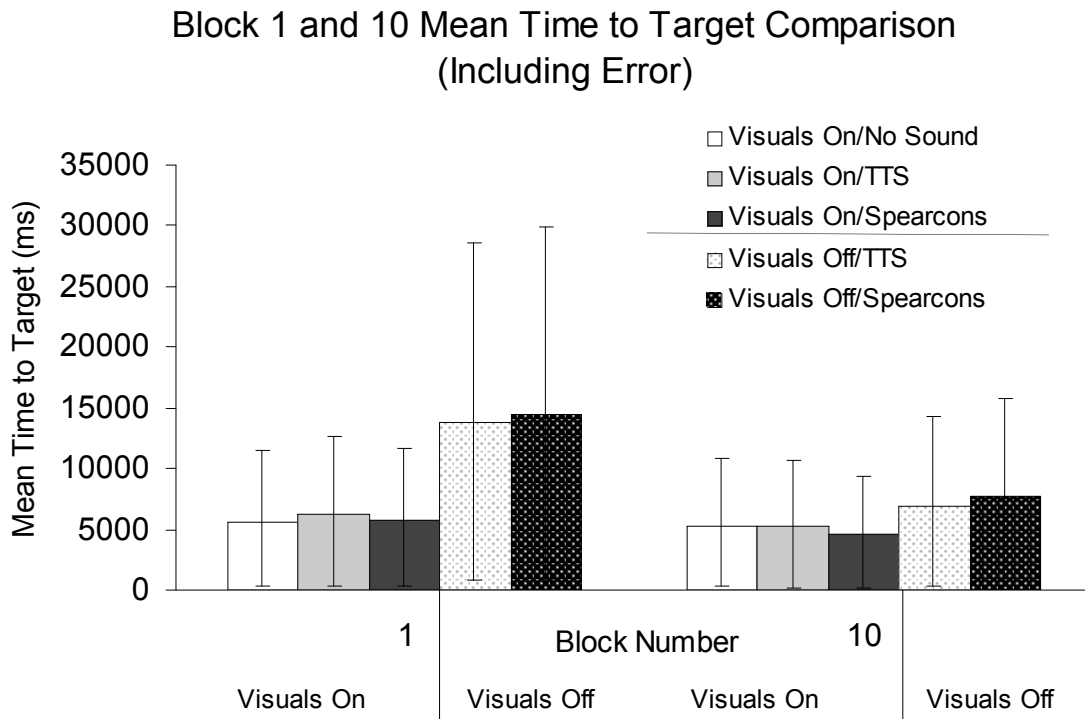


Figure 2. Mean time to target in milliseconds for all conditions in blocks one and ten. The difference in performance between auditory only conditions and those including visual cues decreases by the last block of the experiment. This is evidence that solely auditory menu cues have the ability to approach the efficiency of menus with visual elements. Error bars show 95% Confidence Intervals.

Table 1 summarizes mean and standard deviation information comparing visual and auditory conditions and their performance improvements between the beginning and end of the trials. Comparison of the change in performance among the auditory cues between the first and tenth block revealed a main effect of sonification type $F(2, 6269) = 86.113, p < 0.001$ with an interaction of sonification type and block number $F(2, 6269) = 35.761, p < 0.001$ indicating a more significant improvement from Block 1 to Block 10 for the Spearcon conditions than for the No sound condition. Post-hoc analysis indicated that spearcons and TTS did not show significantly different performance improvements.

Comparing the conditions with visual cues to those without visual cues revealed a main effect of visual cue $F(1, 6271) = 1128.36, p < 0.001$ between the first and tenth block with the non-visual conditions facilitating a larger improvement in performance by the end of the experiment, as indicated with a significant interaction between visual cue condition and block $F(1, 6271) = 355.75, p < 0.001$.

Table 1
Means, Standard Deviations, and Change of Time (ms) to Target Name for Blocks 1 and 10 Collapsed Over Visual and Sound Conditions

Condition	Block 1			Block 10			Δ (ms)
	Mean	SD	N	Mean	SD	N	
Visual On	5828	3535	1845	5014	3635	1842	814
Visual Off	14147	11552	1291	7350	3961	1297	6797
Spearcon	10279	9796	1239	6245	3826	1241	4034
TTS	10089	9408	1258	6111	3436	1264	3978
No Sound	5618	3326	639	5194	4909	634	424

Discussion

Although the results confirm that the conditions including visual cues lead to faster performance overall when compared to conditions with only auditory cues, expectations that the spearcon cues would outperform TTS cues significantly were not corroborated. Instead, performance in conditions using spearcons were found to be consistently, but not significantly, faster to navigate when auditory cues were combined with visual cues. This finding provides some evidence that spearcons combined with visual cues may enhance performance when navigating visual menus. More research should be conducted to determine if this effect is significant.

There are potential reasons why, in the case of the auditory-only conditions, spearcons were not found to lead to shorter navigation times. Since the spearcons were presented as cues

preended to the TTS phrase, some participants may have felt compelled to listen through the spearcon and the silent interval to hear the TTS phrase, rather than concentrating on the spearcon itself. This would have certainly increased time to target. It would be interesting to run the study again without the convenience of the TTS phrase inclusion. Perhaps including a training session before starting the experiment on the associations between the words and the sounds would decrease the impulse to wait for the TTS as well. These considerations should be tested in future studies.

Auditory-only cue performance showed a strong learning curve, and after 10 blocks, the auditory-only condition performance had improved to a point that remained significantly different from the auditory and visual cues combined. Figure 2 shows a compelling picture that reveals the level of performance to be much more level for all five conditions than in the first block of the experiment. There are few who would expect performance on a strictly auditory menu to be better than one including visual cues for a person without a visual impairment. The fact that performance improved to such a degree for individuals accustomed to a visual world lends interest to a replication of this study with visually impaired individuals, who are used to navigating an auditory world. This replication would provide a more well-rounded picture of navigational performance in different contexts.

Additional future research will include studying auditory enhancements, particularly spearcons usage on irregularly shaped menus and submenus, replication of this study on actual cell phone devices, and replication and focus groups with visually impaired and blind users. The latter two plans are critical to the success of the project because visually impaired users will be the prime beneficiaries of audio menu technologies.

Utilizing auditory menus and enhancements in small electronic devices is clearly feasible, and the electronics industry appears ready to take on the challenge of incorporating accessible technology into their interfaces, particularly for cell phone menus. With strong empirical science backing up the feasibility of the spearcon, it is hoped that it will not be long before those with temporary and permanent visual disabilities will more easily be able to enjoy the productivity of electronic devices to the same extent as individuals with normal vision. From the viewpoint of both the manufacturers and the potential users, this research is expected to lead to positive advancements in accessible technology.

Appendix

MATLAB code for algorithm used for Spearcon creation.

compress.m

```
fs = 16e3;
nbits = 16;
l = 0.1;
c = 2;
win = 200;

y = dir('./original/*.wav');

for i = 1:length(y)

    [x,fs,nbits] = wavread(sprintf('./original/%s',y(i).name));
    r = length(x)/fs/l;
    ratio = c*log10(r)+1
    xc = solafs(x',ratio,win);
    wavwrite(xc,fs,nbits,sprintf('./test/%s',y(i).name));
```

end

solafs.m

```
function Y = solafs(X, F, W, Wov, Kmax, Wsim, xdecim, kdecim)
% Y = solafs(X, F, W, Wov, Kmax, Wsim, xdec, kdec) Do SOLAFS timescale mod'n
% Y is X scaled to run F x faster. X is added-in in windows
% W pts long, overlapping by Wov points with the previous output.
% The similarity is calculated over the last Wsim points of output.
% Maximum similarity skew is Kmax pts.
% Each xcorr calculation is decimated by xdecim (8)
% The skew axis sampling is decimated by kdecim (2)
% Defaults (for 22k) W = 200, Wov = W/2, Kmax = 2*W, Wsim=Wov.
% Based on "The SOLAFS time-scale modification algorithm",
% Don Hejna & Bruce Musicus, BBN, July 1991.
% 1997may16 dpwe@icsi.berkeley.edu $Header:
/homes/dpwe/matlab/dpwebox/RCS/solafs.m,v 1.3 2006/04/09 20:10:20 dpwe Exp $
% 2006-04-08: fix to predicted step size, thanks to Andreas Tsiartas

if (nargin < 3)    W    = 200; end
if (nargin < 4)    Wov  = W/2; end
if (nargin < 5)    Kmax = 2 * W; end
if (nargin < 6)    Wsim = Wov; end
```

```

if (nargin < 7)    xdecim = 8; end
if (nargin < 8)    kdecim = 2; end

Ss = W - Wov;

if(size(X,1) ~= 1) error('X must be a single-row vector'); end;

xpts = size(X,2);
ypts = round(xpts / F);
Y = zeros(1, ypts);

% Cross-fade win is Wov pts long - it grows
xfwin = (1:Wov)/(Wov+1);

% Index to add to ypos to get the overlap region
ovix = (1-Wov):0;
% Index for non-overlapping bit
newix = 1:(W-Wov);
% Index for similarity chunks
% decimate the cross-correlation
simix = (1:xdecim:Wsim) - Wsim;

% prepad X for extraction
padX = [zeros(1, Wsim), X, zeros(1,Kmax+W-Wov)];

% Startup - just copy first bit
Y(1:Wsim) = X(1:Wsim);

xabs = 0;
lastxpos = 0;
lastypos = 0;
km = 0;
for ypos = Wsim:Ss:(ypts-W);
    % Ideal X position
    xpos = F * ypos;
    % disp(['xpos=',num2str(xpos),' ypos=',num2str(ypos)]);
    % Overlap prediction - assume all of overlap from last copy
    kmpred = km + ((xpos - lastxpos) - (ypos - lastypos));
    lastxpos = xpos;
    lastypos = ypos;
    if (kmpred <= Kmax && kmpred >= 0)
        km = kmpred; % no need to search
    else
        % Calculate the skew, km
        % .. by first figuring the cross-correlation
        ysim = Y(ypos + simix);

```

```

% Clear the Rxy array
rxy = zeros(1, Kmax+1);
rxx = zeros(1, Kmax+1);
Kmin = 0;

for k = Kmin:kdecim:Kmax
    xsim = padX(floor(Wsim + xpos + k + simix));
    rxx(k+1) = norm(xsim);
    rxy(k+1) = (ysim * xsim');
end
% Zero the pts where rxx was zero
Rxy = (rxx ~= 0).*rxy./(rxx+(rxx==0));
% Local max gives skew
km = min(find(Rxy == max(Rxy))-1);
end
xabs = xpos+km;

% Cross-fade some points
Y(ypos+ovix) = ((1-xfwin).*Y(ypos+ovix)) + (xfwin.*padX(floor(Wsim+xabs+ovix)));
% Add in remaining points
Y(ypos+newix) = padX(floor(Wsim+xabs+newix));
end

```

References

- Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11-44.
- Brewster, S., & Crease, M. G. (1999). Correcting menu usability problems with sound. *Behaviour & Information Technology*, 18(3), 165-177.
- Brewster, S., Raty, V., & Kortekangas, A. (1996). *Earcons as a method of providing navigational cues in a menu hierarchy*. Paper presented at the HCI'96 Conference, Imperial College, London, UK.
- Brewster, S., Wright, P. C., & Edwards, A. D. N. (1993). *An evaluation of earcons for use in auditory human-computer interfaces*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Amsterdam.
- Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2(2), 167-177.
- Hereford, J., & Winn, W. (1994). Non-speech sound in human-computer interaction: A review and design guidelines. *Journal of Educational Computing Research*, 11(3), 211-233.
- Leplatre, G., & Brewster, S. (1998). *An Investigation of Using Music to Provide Navigation Cues*. Paper presented at the International Conference for Auditory Display, Glasgow, UK.
- Leplatre, G., & Brewster, S. (2000). *Designing non-speech sounds to support navigation in mobile phone menus*. Paper presented at the International Conference for Auditory Display, Atlanta, GA.

- Nees, M. A., & Walker, B. N. (In Press). Auditory interfaces and sonification. In C. Stephanidis (Ed.), *The Universal Access Handbook* (pp. TBD). Mahwah, NJ: Lawrence Erlbaum Associates.
- Palladino, D. K., & Walker, B. N. (2007). *Learning rates for auditory menus enhanced with spearcons versus earcons*. Paper presented at the International Conference on Auditory Display, Montreal, Canada.
- Vargas, M. L. M., & Anderson, S. (2003). *Combining speech and earcons to assist menu navigation*. Paper presented at the International Conference for Auditory Display, Boston, MA.
- Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning masking. In J. G. Neuhoff (Ed.), *Ecological Psychoacoustics* (pp. 150-175). New York: Academic Press.
- Walker, B. N., Nance, A., & Lindsay, J. (2006). *Spearcons: Speech-based earcons improve navigation performance in auditory menus*. Paper presented at the International Conference on Auditory Display, London, U.K.
- Words per minute. (2007, December 1, 2007). Retrieved December 2, 2007, from http://en.wikipedia.org/w/index.php?title=Words_per_minute&oldid=174992400
- Yalla, P., & Walker, B. N. (2007). *Advanced Auditory Menus* (No. GIT-GVU-07-12.): Georgia Institute of Technology GVU Center.

Author Note

I thank Dr. Bruce Walker for his valuable mentorship on every aspect of this project. I also thank Dr. Gregory Corso and Dr. Karen Adams for advice on the written quality of this thesis. I acknowledge Mark Godfrey for programming the algorithm used for spearcon creation. Finally, thank you to Naomi Warnick, Sara Cantu, and Melissa Hardman for their help with running participants.