Threshold of Spearcon Recognition for Auditory Menus

Edin Sabic and Jing Chen
New Mexico State University, Las Cruces, New Mexico

The present study examined aspects of a type of auditory feedback that aims at making interactions with menus more efficient. Specifically, manipulations were made to the auditory counterparts of common computer commands and unrelated words to examine identification performance. By investigating the point at which sped-up versions of text-to-speech, *spearcons*, can be identified accurately as words, the boundaries at which the efficiency of spearcon use begins to decline can be better defined. Additionally, by examining the effect of category membership on speed and accuracy of spearcon identification, whether an overarching category can facilitate spearcon use can be examined. Results in two experiments demonstrated that spearcon identification began to decline drastically after linear compression leading to 40% of the original audio length. Reaction time data also demonstrated that spearcon efficiency began to decline after the same level of linear compression. Efficiency scores combining reaction time and accuracy also supported the bottom limit of spearcons at 40% of original audio length.

INTRODUCTION

Auditory feedback is an important part of computer menus, which are integral to many of the tasks users accomplish on modern electronic devices. The increasing complexity of computer menus on modern interfaces creates new problems for both designers and users alike. Users may interact with an inefficient interface that causes them to spend too much time on a task, get lost in navigation, or develop a negative opinion of the interface or device as a whole. With this in mind, it is in the interest of both users and designers to make interactions with menus as efficient and conducive to task-demands as possible.

While there is a substantial amount of literature available pertaining to visual aspects of menus, there is a relative lack of literature available on auditory menus (Yalla & Walker, 2008). Auditory menus have the potential to increase the accessibility of various interfaces to a wider range of users, while also providing a significant supplement to solely visual feedback. Types of auditory feedback take on many forms, including one of the most popular: TTS (Text-To-Speech). TTS is produced by a program that reads text on the device using a synthesized voice. While TTS provides auditory feedback that is phonologically similar to the text, there are numerous other varieties of auditory feedback that do not produce recognizable speech sounds.

The early research in auditory feedback focused on two popular methods: *auditory icons* and *earcons*. According to Gaver (1986), auditory icons are sounds that directly represent the icon they accompany. That is, they are not arbitrary. An example of an auditory icon could be the shutter sound for a camera application on a phone. In general, these sounds have a distinct resemblance to the command or program they are representing (Gaver 1986; Pirhonen, Murphy, McAllister, & Yu, 2006; Palladino & Walker, 2007). While it serves no function in opening or executing a file or program, an auditory icon can provide feedback as to what function the user has selected. This auditory feedback can serve as both a cross-check for some individuals, and a primary method of identifying what was selected for other users. Auditory icons can be effectively used to represent the sounds of many objects, but are less useful for accompanying abstract, or numerous, objects and functions. Given that computer menus and other interfaces often include many abstract objects or functions, the usage of auditory icons usually does not enhance auditory menus (Walker, Nance, & Lindsay, 2006).

Earcons, on the other hand, are non-verbal sequences of musical notes or brief melodies (Walker, et al., 2006). Unlike auditory icons, earcons are somewhat arbitrary in that the sounds are not representative of the command or function that they represent. One experiment using earcons posited that earcons are easier to learn than unstructured sound stimuli accompanying commands (Brewster et al., 1992). Due to the abstract musical nature of earcons, this type of auditory feedback is able to represent the objects and commands of many menu functions in a way that auditory icons, due to their need to represent a function in a congruent and meaningful way, are not.

Spearcons were introduced by Walker et al. (2006) as an alternative to both auditory icons and earcons. Spearcons are created by first converting the command, text, or label of a menu item into a TTS sound file using speech software. The tempo of the audio file is then increased while keeping pitch stable so that it becomes or approaches a non-speech sound. This type of compression can either be done linearly, or by varying the amount of compression relative to word length. The latter is a more specialized variant of audio compression that leads to larger compression of longer words and smaller compression of shorter words. Prior research has shown that the use of spearcons in auditory menus can increase speed and accuracy (Walker et al., 2006). Specifically, due to their shortened nature, spearcons can convey the same or similar information in a shorter timeframe. The user can therefore accomplish their tasks in a shorter timeframe, potentially leading to higher overall efficiency and satisfaction with the device.

While the use of spearcons has been demonstrated to be beneficial to auditory menu navigation (Walker et al., 2013), there has been little analysis on spearcons themselves. Specifically, although certain amounts of compression for spearcons has been disclosed in previous research, there is still some ambiguity concerning how fast a spearcon should be, and how recognition of spearcons is affected by variables such as categorical membership. At what point a spearcon becomes unrecognizable as a word could be beneficial for assessing the benefits of spearcon use in various scenarios. To determine this point, we utilized the recognition threshold method used in psychophysics research. Research using the absolute threshold has been conducted to analyze vision, hearing, and odor thresholds. A recognition threshold distinguishes the bottom limit at which a stimulus is still recognizable. For the purposes of our experiments, an identification rate of 75% was set as the threshold.

To evaluate and assess the aforementioned questions, two experiments were conducted. Experiment 1 explored participants' accuracy rates of recognizing 30 words which were linearly compressed from 100% audio length all the way down to 20% of the original audio length in 10% increments. Additionally, this experiment compared accuracy rates of categorical, in this case words reflecting common computer commands such as copy, paste, etc., versus non-categorical words to assess whether categorical membership could facilitate spearcon recognition. Experiment 2 was similar to Experiment 1, except that we assessed efficiency of various speeds of spearcons by logging reaction time (RT) from the stimulus onset to when the participant recognized the word. This change allowed us to create an overall efficiency measure across speed levels. Accuracy data was collected in addition to RT to again examine whether there were differences in accuracy depending on categorical membership.

EXPERIMENT 1

The primary purpose of Experiment 1 was to investigate the threshold at which spearcons can be consistently recognized as words. In addition, two different sets of words used for spearcon creation were either categorically or non-categorically assembled to assess whether categorical membership influenced recognition and accuracy. Category membership was predicted to potentially increase perceptibility because higher levels of organization of the words may allow the participant to better remember previously encountered spearcons. The aims of Experiment 1 were exploratory, with the goal being to determine an approximate threshold for further analysis in later experiments. This finding would assess the generalized lower limit of average compression used by Walker et al. (2006) and others for spearcon stimuli.

Methods

Participants. Ten undergraduate students (8 females) at New Mexico State University participated in the experiment for research credits in various psychology classes. Participants were an average 19.4 years old, *SD* = 1.17.

Apparatus and equipment. A Dell OptiPlex 7020 with Microsoft Windows 7 Professional OS was used to run the experiment and collect data. E-Prime software was used to present audio stimuli. Participants wore a pair of Koss UR23iK headphones during the experiment. Computer audio volume was kept constant at 50% of full volume across all participants.

Auditory stimuli. Fifteen TTS categorical words and 15 non-categorical words (see Appendix) were produced using the TTS software NaturalReader 14.0, which reads entered text in a neutral voice. Categorical words were common computer commands found in menu lists. Computer commands were chosen for the categorical words because many devices used on a daily basis incorporate at least some of these words, and, therefore, the results may generalize better to real-world situations. A free open source digital audio editor, Audacity, was used to record the TTS words produced by NaturalReader via the WASAPI playback option, which captures audio played without the usage of a microphone. Audacity was also used to then manipulate the tempo of the TTS audio files while keeping the pitch constant. Each one of the 30 original TTS audio files was systematically compressed to correspond with 10% decreases in overall sound length. This resulted in nine versions of each word, spanning from the tempo of typical conversational speech at 100%, all the way down to 20% of the original sound length. Through this process, 270 audio files were created. Audio files were saved in .mp3 format for usage in E-Prime. All auditory stimuli were presented randomly to each participant.

Procedure. The task was to identify what word was presented aurally through headphones. On each trial, a word was played through the headphones and participants responded by pressing the space bar when they recognized the word, and then typing what they heard. They were instructed to take their best guess if they were unsure about which word they heard. Additionally, participants were able to see their typing on the screen. Participants were allowed to delete characters by pressing the backspace key, and they were instructed to press the enter key when they finished typing the word. After pressing the enter key, the next trial began. Participants were instructed to respond as quickly and accurately as possible.

Each participant performed two trial blocks, the categorized-words (i.e., command) block and the uncategorized-words block. The order of the blocks was counterbalanced among participants, and there was a short break between the blocks. Each trial block included 270 trials, with each word at each particular speed being repeated once. There was a practice session at the beginning of the experiment. During the practice, participants were instructed on how to respond to auditory stimuli, and what to do if they did not understand the sound presented. The practice session served two purposes: (1) To instruct the participant how to respond to the stimuli, and (2) to assess whether the 50% of full volume was adequate. Further, spearcons presented during the practice session spanned all levels of linear compression used in the experiment, and did not belong to a category. All auditory stimuli used in the practice were not used during the experimental trials, and participants were not familiar with the stimuli prior to starting the experiment.

Results

A repeated-measure analysis of variance (ANOVA) was conducted with speed and categorical type as within-subjects factors on the recognition rate. As expected, the main effect of speed on recognition was significant (see Figure 1), $F(8, 72) = 109.77$, $p < .001$, $\eta_p^2 = .92$. No significant effect was found for category, $F(1, 9) = 1.94$, $p = .197$, $\eta_p^2 = .18$, nor was the interaction between the two factors significant, $F(8, 72) = .94$, $p = .489$, $\eta_p^2 = .10$. While the categorical words had higher recognition rate numerically compared to the non-categorical group (*M*s = .87 vs. .85 for categorical and non-categorical words, respectively), this difference was not statistically significant. Figure 1 illustrates the decline of accuracy as a function of spearcon speed. Accuracy begins to consistently decline from around 80% of the original sound length onwards. The most significant decline in accuracy begins at 30% original audio length. Means of both groups at 20% original audio length were barely above 50%.

The recognition threshold crosses the categorized group and the uncategorized group at similar, although slightly different, points. Both cross the recognition threshold at around 30% original sound length, although the category group did maintain almost 80% accuracy (.79) at the 30% original sound length level. Results also showed that some spearcons were much less recognizable at levels of 20% linear compression: For example, the accuracy for “close” at 20% original audio length was .00 whereas that for “options” at 20% original audio length was .85.

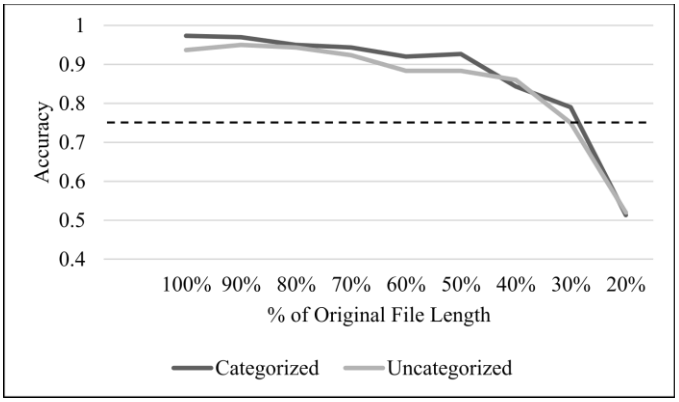

Figure 1. Mean accuracy rate versus spearcon speed. The black dotted line represents the 75% recognition threshold.

Discussion

Data from Experiment 1 demonstrated that even speeds of spearcons representing 30% original audio length were still above the 75% recognition threshold. Interestingly, the category group was able to maintain close to 80% accuracy (.79) even at linear compression that led to 30% of the original audio length. Across all spearcon speeds, the accuracy of the category group ($M = .870$, $SD = .032$), was greater than the no category group ($M = .847$, $SD = .066$), although this difference was not found to be significant. Overall, these results speak to the ability of individuals to understand spearcons at linear compression down to 30% of the original audio length. Although the differences between the category and no-category group were not statistically significant, the means seemed to be different enough to merit examination in a larger sample size in Experiment 2.

While limited in scope, Experiment 1 demonstrated that individuals can respond efficiently to up to 15 commands with high compressed rates in a given task. Individual differences were also found in spearcon recognition, with some individuals having much more success at higher levels of linear compression than others. This experiment supports the ability of individuals to understand spearcons, even at speeds increased to over two times the rate of spoken dialogue. This finding suggests that there is an amount of redundancy in the speed at which most TTS is played at, and further supports the added ability of some TTS programs, such as *NaturalReader*, to allow users to increase the speed of text-to-speech software so that they may complete tasks and receive information more efficiently.

Results from Experiment 1 were important in that they outlined thresholds for spearcon recognition at specific levels of increased tempo. The difference between means of category groups led us to keep this manipulation in Experiment 2, especially given the relatively small sample size in Experiment 1. However, Experiment 2 was designed to not only assess the threshold of spearcon recognition, but also to evaluate how efficient each level of spearcon speed was.

EXPERIMENT 2

Experiments 1 and 2 were identical except that in Experiment 2 RT measures were captured, corresponding to when the individual recognized the audio stimuli. By collecting RT measures, an inverse efficiency score (IES) measure was calculated by dividing RT over accuracy (Townsend & Ashby, 1978). Experiment 2 aimed to examine the efficiency of spearcons, even after linear compression resulting in 20% of the original audio length. Based on the accuracy data in Experiment 1, we hypothesized that efficiency of spearcons would continuously increase up to 30% original audio length. While the accuracy data in Experiment 1 supported the use of spearcons at linear compression leading to 30% original audio length, Experiment 2 tested whether this was an acceptable level in terms of RT and IES.

Methods

Participants. Twenty-six undergraduate students (22 females) at New Mexico State University participated in the experiment for research credits in various psychology classes. Participants were an average 20.4 years old, $SD = 2.91$.

Apparatus and equipment. Similar apparatus and equipment were used in the previous study. Namely, the same Dell OptiPlex 7020 with Microsoft Windows 7 Professional OS was used to collect responses and headphones were used for presentation of auditory stimuli.

Auditory Stimuli. The exact same auditory stimuli were used from Experiment 1.

Procedure. Participants had a similar task to the one in Experiment 1. Participants were again instructed to press the space bar when they recognized the word they heard. The time from the onset of the auditory stimulus until a space bar response was recorded as RT. The following steps were similar; namely, participants were instructed to type the word that they heard and press enter to proceed to the next trial. Participants were instructed that they could press the space bar at any time, even before the audio file had completely finished.

Results

A similar ANOVA was conducted as in Experiment 1. There was a significant main effect of speed on accuracy, $F(8, 200) = 237.94$, $p < .001$, $\eta_p^2 = .91$, and on RT, $F(8, 200) = 10.42$, $p < .001$, $\eta_p^2 = .48$, and on the efficiency measure (IES), $F(8, 200) = 36.49$, $p < .001$, $\eta_p^2 = .29$. Interestingly, there was a significant interaction between speed and category for accuracy but not other dependent variable measures: For RT, $F < 1$; for accuracy, $F(8, 200) = 2.54$, $p = .012$, $\eta_p^2 = .09$; for the efficiency measure, $F < 1$. For all three dependent variable measures, efficiency, accuracy, and RT, there was no significant main effect for category, *F*s < 1. Importantly, an analysis of RT data comparing spearcons at 40% original audio length and words at original (100%) audio length demonstrated that the 40%-spearcons were more quickly recognized, $F(1, 25) = 12.90$, $p = .001$, $\eta_p^2 = .34$. Further, analysis of IES data demonstrated that spearcons at 40% original audio length were not significantly different from the words at original audio length, $F(1, 25) = 3.36$, $p = .079$, $\eta_p^2 = .12$. See Table 1 for mean RT data in each condition. Although not significant, the mean data suggest that people took longer to recognize category words compared to no category words.

Table 1. Mean(*SD*) for RT and Accuracy by Category and Speed

% of original length	Category Words		No Category RT	
	RT	Accuracy	RT	Accuracy
20%	771(233)	.47(.14)	740(218)	.48(.15)
30%	697(195)	.69(.12)	667(183)	.73(.14)
40%	652(172)	.79(.11)	650(168)	.83(.13)
50%	667(188)	.86(.12)	651(180)	.87(.11)
60%	669(176)	.88(.09)	646(172)	.85(.11)
70%	688(191)	.88(.09)	656(201)	.87(.11)
80%	671(141)	.89(.10)	641(152)	.90(.12)
90%	704(182)	.91(.10)	656(159)	.90(.11)
100%	694(171)	.90(.08)	693(183)	.90(.11)

Discussion

The purpose of Experiment 2 was to further determine at what point spearcons become unrecognizable as words, and whether categorical membership could influence accuracy, response time, or an efficiency measure combining the two. Analyses did show differences in these measures based on category in the form of a significant interaction term, although only the main effect of speed was found to be significant. Results of Experiment 2 also demonstrated that spearcon recognition begins to drastically decrease after linear compression leading to 40% of the original file length. Although accuracy of the two fastest spearcons, 20% and 30% original audio length, was still above 50%, efficiency scores did not support the usage of spearcons at these levels of linear compression. See Figure 2 for a graph depicting efficiency scores and accuracy rates.

Importantly, data from Experiment 2 supported that there were differences in-between groups across levels of spearcon speed. Participants in Experiment 2, similar to Experiment 1, did not have any practice with the auditory stimuli prior to testing. Experiment 2 helps to outline standards of spearcon linear compression to maintain word recognition without training. This word recognition may have some effect on maintaining the non-arbitrary nature of spearcons. Given that what makes spearcon use theoretically superior to alternatives like earcons and auditory icons is that they are non-arbitrary and can map onto abstract objects or commands, this distinction is important.

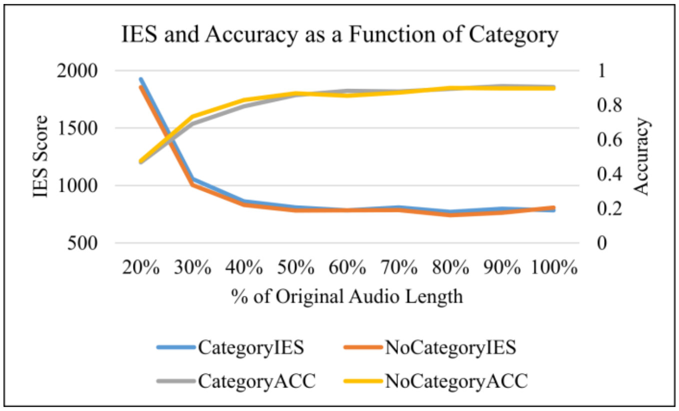

Figure 2. Inverse efficiency score (IES) and accuracy measures as a function of speed and category status.

GENERAL DISCUSSION

An advantage of spearcons compared to other forms of auditory feedback is that they are both non-arbitrary and able to correspond to abstract objects and commands. Spearcons are phonologically related to the objects and commands they represent, and are able to represent abstract concepts due to this characteristic. Additionally, spearcons can convey the same information faster, leading to higher efficiency and potentially higher user satisfaction. It follows then that determining the point at which spearcons lose their semantic information is important. As the results in Experiments 1 and 2 showed, efficiency decreased drastically for the two fastest spearcon levels (linear compression resulting in 30% and 20% original audio length). In conjunction with this, spearcon accuracy decreased drastically for the two fastest spearcon levels.

Data from Experiment 1 supported that spearcons could be recognized at linear compression leading to 40% of the original audio length (Walker et al., 2006), or even at 30% of original audio length. Although there was a decline in accuracy at the level of linear compression to 30% original audio length, the results of Experiment 1 supported that individuals were still able to recognize spearcons above the 75% recognition rate. The threshold at which spearcons can be recognized as words can yield benefits to research in auditory feedback. There may be idiosyncratic differences between spearcons that can be recognized as words, and those that cannot. While Experiment 1 data revealed that spearcons at 30% linear compression were above the 75% recognition rate, Experiment 2 demonstrated that the use of these spearcons was inefficient.

The present study has supported that individuals are able to recognize spearcons without practice up to linear compression resulting in 40% original audio length. After this point, however, efficiency and accuracy begin to decline. This analysis helps support the usage of spearcons at a lower limit around 40% of the original audio length. Although there was no main effect of category found, the interaction term was significant for accuracy. Future research may attempt to analyze why these two different sets of words were not influenced by linear compression at the same rate. Future studies may also look into how recognition rates would differ if users knew of the word list ahead of time. This research might find that users who knew the word lists ahead of time would be less susceptible to spearcon speed when it comes to accuracy. Additionally, the point at which accuracy crosses the 75% recognition might be affected. Overall, these two conditions might predict the performance of a user using an unfamiliar spearcon menu, compared to a user who is familiar with a spearcon menu.

In conclusion, the ability of individuals to understand spearcons at 40% original audio length, without prior practice, has been supported. In addition, the data from Experiments 1 and 2 suggest that usage of spearcons at 20% and 30% of original audio length is unwise. At these levels, accuracy and efficiency, assessed via inverse efficiency scores, begin to significantly decrease. Although a significant main effect of categorical membership was not found, the significant interaction effect alludes to the fact that different groups of spearcons across menus may not perform similarly. As a result, in future research and design practice, it is important to consider the relation among the words within a spearcon menu when deciding for a specific tempo for the spearcons. Spearcons can enhance auditory menus by allowing for the same object or command to be communicated in a faster way than text-to-speech, so long as compression does not reach levels of 30% or less of the original audio length.

REFERENCES

Brewster, S. (1992). *Providing a model for the use of sound in user interfaces*. University of York, Department of Computer Science.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-computer interaction*, *2*, 167-177.

Townsend, J. T., & Ashby, F. G. (1978). *Methods of modeling capacity in simple processing systems*. In J. Castellan & F. Restle (Eds.), Cognitive Theory. Vol. 3. (pp. 200-239). Hillsdale, N.J.: Erlbaum.

Palladino, D. K., & Walker, B. N. (2008, September). Navigation efficiency of two dimensional auditory menus using spearcon enhancements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1262-1266). SAGE Publications.

Pirhonen, A., Murphy, E., McAllister, G., & Yu, W. (2006, June). Non-speech sounds as elements of a use scenario: a semiotic perspective. In *Proceedings of the 12th International Conference on Auditory Display* (pp. 134-140). London, UK.

Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. . In *Proceedings of the 12th International Conference on Auditory Display* (pp. 63-68). London, UK.

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors*, *55*, 157-182.

APPENDIX

Table 2. Words used in both Category and No Category Group

Category Words	No Category Words
Close	Cattle
Copy	City
Delete	Doctor
Export	Inbox
Info	Ocean
Open	Order
Options	Pepper
Paste	Picket
Preview	Pillar
Print	Rare
Redo	Room
Reload	Sky
Search	Sound
Select	Spider
Setup	Trumpet