

# LEARNABILITY OF SOUND CUES FOR ENVIRONMENTAL FEATURES: AUDITORY ICONS, EARCONS, SPEARCONS, AND SPEECH

Tilman Dingler<sup>1</sup>, Jeffrey Lindsay<sup>2</sup>, Bruce N. Walker<sup>2</sup>

<sup>1</sup>Ludwig-Maximilians-Universität München  
Department "Institut für Informatik"  
Lehr- und Forschungseinheit Medieninformatik  
Oettingenstraße 67, 80538 München, Germany  
dingler@cip.ifi.lmu.de

<sup>2</sup>Sonification Lab  
School of Psychology  
Georgia Institute of Technology  
654 Cherry St., Atlanta, GA 30332-0170  
gte457e@mail.gatech.edu,  
bruce.walker@psych.gatech.edu

## ABSTRACT

Awareness of features in our environment is essential for many daily activities. While often awareness of such features comes from vision, this modality is sometimes unavailable or undesirable. In these instances, auditory cues can be an excellent method of representing environmental features. The study reported here investigated the learnability of well known (auditory icons, earcons, and speech) and more novel (spearcons, earcon-icon hybrids, and sized hybrids) sonification techniques for representing common environmental features. Spearcons, which are speech stimuli that have been greatly sped up, were found to be as learnable as speech, while earcons unsurprisingly were much more difficult to learn. Practical implications are discussed.

## 1. INTRODUCTION

Awareness of features and objects in the world around us is vital in many aspects of life. Their importance impacts all aspects of life, ranging from our safety and ability to travel, to helping determine our comfort and productivity levels. Landmarks are crucial to navigation, helping individuals to determine where they are and to plot a course towards a desired destination. Failure to avoid an object as a driver or pedestrian could spell disaster. We often rely on vision to make salient these aspects of our environment, but sometimes this is not preferable, or even possible. In these instances, auditory cues can be an effective alternative.

When devising an auditory display scheme for environmental features and objects, one key consideration must be how learnable the scheme is. In some situations users might not interact with a display that is difficult to learn enough to understand it well. Even in usage scenarios where extended learning time did exist, users might not wish to invest the time in doing so. In light of this, the following study was designed to investigate the relative learnability of different methods for auditory display of features surrounding a listener. Common auditory display schemes such as earcons, auditory icons and speech were examined, as well as more novel approaches such as spearcons and certain combinations of auditory icons and earcons.

### 1.1. Auditory Icons

Auditory icons [1] are brief sounds that represent objects, functions, and actions. They take advantage of

the user's prior knowledge and natural auditory associations with sound sources and causes. They are meant to be the auditory equivalent of visual icons that are broadly used in personal computing, representing objects or processes through graphical symbols. Icons basically simplify information display due to their capacity to present a lot of information in a concise and easily recognized format [2]. Due to the visual system's capability to process several dimensions such as shape, color, etc. in parallel, a variety of information can be encoded into a visual icon. The same can be said about the auditory system and the dimensions it processes (pitch, amplitude, timbre, etc.). According to Hemenway [3], icons are more easily located and processed than words, since meaning can be derived directly from the object or action they represent. Kolers et al. [4] even states how cultural and linguistic barriers can be transcended by using icons. Auditory icons can therefore be mapped to the actual object or event that is represented whether directly or indirectly. Direct relations use the sound made by the target event whereas indirect relations substitute a surrogate for the target [5]. Thus, objects are represented by the involved sound-producing events. As an example, the sound of running water or a paper towel dispenser can be used to represent restrooms. "The directness or auditory similarity between the icon and the actual object can vary considerably" [6]. As long as a sound evokes the associated sound of an object or action, it is classified as an auditory icon. Even though the utility of auditory icons in computer applications is limited, due to problems in representing abstract concepts [7], auditory icons can be considered very useful in representing real items in the environment.

### 1.2. Earcons

For many items where there is no clear iconic representation, earcons can yield an effective sonification. Earcons are abstract, synthetic and mostly musical tones or sound patterns that can be used in structured combinations. They are non-verbal audio messages, which are "composed of motives, which are short, rhythmic sequences of pitches with variable intensity, timbre and register"[8]. Blattner et al. [2] have defined a system of hierarchical earcons, in which a particular structure is given to single earcons that are grouped together. Each earcon can be thought of as a node in a tree that inherits all the properties of the earcons above it. According to Brewster [8] there are a maximum of five levels to this tree since there are five

parameters to vary: rhythm, pitch, timbre, register and dynamics. Therefore earcons can be combined to produce complex audio messages. It is even possible to let an automatic system itself do the work of combining auditory properties in order to create new, but still consistent sounds. In this way a hierarchical system of earcons can be easily enhanced as a “family of sounds”. Earcons are used, for example, as a method to add context to a menu in a user interface, helping the user maintain awareness of where in the tree he or she is currently located.

Nevertheless, the relationship between the earcon and the object has a more or less metaphorical character [6]. A three-note pattern representing the deletion of a file through decreasing loudness and pitch is an example of an earcon for deleting a file, with the diminishing loudness and pitch of the sound representing the deletion. Most earcons have solely symbolic mappings between the sounds and the information they represent. The hierarchy of sounds helps during the learning process, but because they are mostly arbitrary, the associations still require considerable learning by the user.

Brewster [8] showed that earcons are better than unstructured bursts of sound in auditory displays. He developed the following guidelines for the design of earcons in order to optimize recognition rates:

- **Timbre:** Brewster concluded that musical timbres were more effective than simple tones. Multiple harmonics support perception, avoid masking and should basically be used in such a way that different timbres are easy to tell apart.

- **Pitch:** It is hard for the listener to distinguish two earcons that differ solely in pitch. Due to this, pitch should never be used as a single indicator of distinction in earcons. It is better to combine complex intra-earcon pitch structures with, for example rhythmic variations. Suggested ranges for pitch would be: Max.: 5kHz and Min.: 125Hz-150Hz.

- **Register:** Brewster also recommends not using register on its own to differentiate earcons. As with pitch, it is better to combine changes in register with other sound dimensions or at least use large differences of two or three octaves in order to achieve good rates of recognition.

- **Rhythm:** According to Patterson [9] sounds using similar rhythms are very likely to be confused. Because of this it is important to make the rhythms of separate earcons as different as possible. Furthermore, studies have shown that different numbers of notes in each rhythm highly supported the differentiation among earcons.

- **Intensity:** Generally, a designer needs to be careful when dealing with sound amplitudes. Since the perception of loudness differs from person to person, the listener should always be in control of the overall sound level of an auditory display. That is why it is important to keep all earcons within a close range of intensity, so that if the user changes the system’s volume, no sounds get lost. Suggested ranges of intensity are: Max.: 20dB above threshold and Min.: 10dB above threshold [9].

- **Combinations:** Between two earcons that are played sequentially, it is recommended that some kind of gap be inserted in between, so the listener can tell when one earcon ends and the next begins. Brewster suggests a delay of 0.1 seconds.

### 1.3. Speech

The most obvious presentation of an object is conveyed by simple speech. However, having a voice announcing the object can create various problems. One issue is that because of its relatively small bandwidth, a spoken word or phrase is harder to spatially localize than a sound using a high bandwidth of frequencies [10]. Another reason is that processing speech requires a lot of mental resources. It is thereby extremely hard for example, to hold a conversation with another person while receiving speech cues. However, in comparison to auditory icons and earcons creating speech sounds is fairly simple since there is software that does the job of turning text into speech (TTS: Text To Speech).

### 1.4. Spearcons

Looking for ways of improving performance and usability of menu-based interfaces, Walker et al. [7] have developed spearcons as a further speech-based type of auditory representation. Spearcons use spoken phrases sped up until they may no longer be recognized as speech. Building on the simplicity of creating speech cues mentioned earlier, spearcons can be created automatically by using basic text-to-speech software and an algorithm to speed up the phrase. Each spearcon is unique due to the specific underlying speech phrase, which allows them to be both distinct but at the same time allows similar phrases to form families of related sounds, much like earcons. Palladino and Walker [11] found that learning rates for an auditory menu scheme were faster when using spearcons compared to earcons. Because the mapping between spearcons and the object they represent is non-arbitrary, less learning is required.

All of these various sonification methods have advantages and disadvantages and are relatively common in auditory displays. When designing such displays one important consideration is how learnable the constituent sounds are, as this will affect the overall learnability of the entire interface. The following study was conducted to investigate this issue.

### 1.5. Hybrid Sounds

It is possible to combine different sound types (e.g., earcons and auditory icons) in various ways to generate hybrid sounds. Such sounds might allow the strengths of each of their parent sounds to compensate for the drawbacks that the other parent sound might possess.

## 2. METHODS

### 2.1. Participants

Thirty-nine undergraduate students who reported normal or corrected-to-normal hearing and vision participated for partial credit in a psychology course. There were 25 male and 14 female students, who ranged in age from 18 to 23 (mean = 20, st dev = 1.3).

Feature	Category	Direct Sound	Size
Public Building	Building	No	large/huge
Pedestrian Light	Intersection Aids	No	medium
Crosswalk		No	large
Curb Cut (up and down)		No	medium
Street Light/Sign	Obstacles	No	medium
Fire Hydrant		No	small
Parking Meter		No	small
Road Work		Yes	large
Tree	Plants	No	medium
Bush		No	small
Bench	Usable Objects	No	medium
Public Phone		Yes	medium
Emergency Phone		Yes	medium
Garbage Can		No	medium
Stairs (up and down)		No	medium
Bus Stop		No	medium
Fountain	Landmarks	Yes	large
Landmark		No	medium/large

Table 1. Environmental features used in the study, as well as their classification by category, sound production, and size.

Feature	Auditory Icon	Earcon	Speech	Spearcon	Earcon-Icon Hybrid	Sized Hybrid
Public Building	<a href="#">[ai_building.wav]</a>	<a href="#">[e_building.wav]</a>	<a href="#">[s_building.wav]</a>	<a href="#">[spr_building.wav]</a>	<a href="#">[eah_building.wav]</a>	<a href="#">[sz_building.wav]</a>
Pedestrian Light	<a href="#">[ai_pedestrian_light.wav]</a>	<a href="#">[e_pedestrian_light.wav]</a>	<a href="#">[s_pedestrian_light.wav]</a>	<a href="#">[spr_pedestrian_light.wav]</a>	<a href="#">[eah_pedestrian_light.wav]</a>	<a href="#">[sz_pedestrian_light.wav]</a>
Crosswalk	<a href="#">[ai_crosswalk.wav]</a>	<a href="#">[e_crosswalk.wav]</a>	<a href="#">[s_crosswalk.wav]</a>	<a href="#">[spr_crosswalk.wav]</a>	<a href="#">[eah_crosswalk.wav]</a>	<a href="#">[sz_crosswalk.wav]</a>
Curb Cut Up	<a href="#">[ai_curb_cut_up.wav]</a>	<a href="#">[e_curb_cut_up.wav]</a>	<a href="#">[s_curb_cut_up.wav]</a>	<a href="#">[spr_curb_cut_up.wav]</a>	<a href="#">[eah_curb_cut_up.wav]</a>	<a href="#">[sz_curb_cut_up.wav]</a>
Curb Cut Down	<a href="#">[ai_curb_cut_down.wav]</a>	<a href="#">[e_curb_cut_down.wav]</a>	<a href="#">[s_curb_cut_down.wav]</a>	<a href="#">[spr_curb_cut_down.wav]</a>	<a href="#">[eah_curb_cut_down.wav]</a>	<a href="#">[sz_curb_cut_down.wav]</a>
Street Light/Sign	<a href="#">[ai_street_light.wav]</a>	<a href="#">[e_street_light.wav]</a>	<a href="#">[s_street_light.wav]</a>	<a href="#">[spr_street_light.wav]</a>	<a href="#">[eah_street_light.wav]</a>	<a href="#">[sz_street_light.wav]</a>
Fire Hydrant	<a href="#">[ai_fire_hydrant.wav]</a>	<a href="#">[e_fire_hydrant.wav]</a>	<a href="#">[s_fire_hydrant.wav]</a>	<a href="#">[spr_fire_hydrant.wav]</a>	<a href="#">[eah_fire_hydrant.wav]</a>	<a href="#">[sz_fire_hydrant.wav]</a>
Parking Meter	<a href="#">[ai_parking_meter.wav]</a>	<a href="#">[e_parking_meter.wav]</a>	<a href="#">[s_parking_meter.wav]</a>	<a href="#">[spr_parking_meter.wav]</a>	<a href="#">[eah_parking_meter.wav]</a>	<a href="#">[sz_parking_meter.wav]</a>
Road Work	<a href="#">[ai_road_works.wav]</a>	<a href="#">[e_road_works.wav]</a>	<a href="#">[s_road_works.wav]</a>	<a href="#">[spr_road_works.wav]</a>	<a href="#">[eah_road_works.wav]</a>	<a href="#">[sz_road_works.wav]</a>
Tree	<a href="#">[ai_tree.wav]</a>	<a href="#">[e_tree.wav]</a>	<a href="#">[s_tree.wav]</a>	<a href="#">[spr_tree.wav]</a>	<a href="#">[eah_tree.wav]</a>	<a href="#">[sz_tree.wav]</a>
Bush	<a href="#">[ai_bush.wav]</a>	<a href="#">[e_bush.wav]</a>	<a href="#">[s_bush.wav]</a>	<a href="#">[spr_bush.wav]</a>	<a href="#">[eah_bush.wav]</a>	<a href="#">[sz_bush.wav]</a>
Bench	<a href="#">[ai_bench.wav]</a>	<a href="#">[e_bench.wav]</a>	<a href="#">[s_bench.wav]</a>	<a href="#">[spr_bench.wav]</a>	<a href="#">[eah_bench.wav]</a>	<a href="#">[sz_bench.wav]</a>
Public Phone	<a href="#">[ai_public_phone.wav]</a>	<a href="#">[e_public_phone.wav]</a>	<a href="#">[s_public_phone.wav]</a>	<a href="#">[spr_public_phone.wav]</a>	<a href="#">[eah_public_phone.wav]</a>	<a href="#">[sz_public_phone.wav]</a>
Emergency Phone	<a href="#">[ai_emergency_phone.wav]</a>	<a href="#">[e_emergency_phone.wav]</a>	<a href="#">[s_emergency_phone.wav]</a>	<a href="#">[spr_emergency_phone.wav]</a>	<a href="#">[eah_emergency_phone.wav]</a>	<a href="#">[sz_emergency_phone.wav]</a>
Garbage Can	<a href="#">[ai_garbage_can.wav]</a>	<a href="#">[e_garbage_can.wav]</a>	<a href="#">[s_garbage_can.wav]</a>	<a href="#">[spr_garbage_can.wav]</a>	<a href="#">[eah_garbage_can.wav]</a>	<a href="#">[sz_garbage_can.wav]</a>
Stairs Up	<a href="#">[ai_stairs_up.wav]</a>	<a href="#">[e_stairs_up.wav]</a>	<a href="#">[s_stairs_up.wav]</a>	<a href="#">[spr_stairs_up.wav]</a>	<a href="#">[eah_stairs_up.wav]</a>	<a href="#">[sz_stairs_up.wav]</a>
Stairs Down	<a href="#">[ai_stairs_down.wav]</a>	<a href="#">[e_stairs_down.wav]</a>	<a href="#">[s_stairs_down.wav]</a>	<a href="#">[spr_stairs_down.wav]</a>	<a href="#">[eah_stairs_down.wav]</a>	<a href="#">[sz_stairs_down.wav]</a>
Bus Stop	<a href="#">[ai_bus_stop.wav]</a>	<a href="#">[e_bus_stop.wav]</a>	<a href="#">[s_bus_stop.wav]</a>	<a href="#">[spr_bus_stop.wav]</a>	<a href="#">[eah_bus_stop.wav]</a>	<a href="#">[sz_bus_stop.wav]</a>
Fountain	<a href="#">[ai_fountain.wav]</a>	<a href="#">[e_fountain.wav]</a>	<a href="#">[s_fountain.wav]</a>	<a href="#">[spr_fountain.wav]</a>	<a href="#">[eah_fountain.wav]</a>	<a href="#">[sz_fountain.wav]</a>
Landmark	<a href="#">[ai_wreck.wav]</a>	<a href="#">[e_wreck.wav]</a>	<a href="#">[s_wreck.wav]</a>	<a href="#">[spr_wreck.wav]</a>	<a href="#">[eah_wreck.wav]</a>	<a href="#">[sz_wreck.wav]</a>

Table 2. Links to all of the sounds used in the experiment for each sound type and feature.

## 2.2. Apparatus and Equipment

A Dell Optiplex GX620 computer was used in this experiment. It was running the Windows XP operating system and has a Pentium 4 3.2 GHz processor and 1 GB of RAM. An external Creative Soundblaster Extigy sound card was used for sound production and participants listened using Sennheiser HD202 circumaural headphones. The software used in this experiment was created for that purpose, using a Flash-

based front end for the experiment interface and a Java-based server applet for data logging.

## 2.3. Stimuli

Eighteen common environmental features were selected from the area outside a campus building. They were drawn specifically from this area to be used in a later part of the experiment involving an auditory virtual reality of this area that is not reported here. However, the features chosen are common in many urban

environments, and not (with one exception) unique to the location they were drawn from. Each feature was then classified into a high level category, a size category, and by whether it directly produces a sound or not (see Table 1 for a list of the features as well as their classifications). Two of the features, stairs and curb cuts have both an up and a down version, for a total of twenty features.

There were six high level categories, which were chosen based on the perspective of a visually-impaired pedestrian; building/area, intersection helpers, obstacles, plants, usable objects, and landmarks. Buildings indicated large structures that an individual could enter. Intersection helpers are features that are useful when attempting to cross the street at an intersection. Features that would not be used and need to be avoided by a visually-impaired pedestrian were classified as obstacles. All vegetation was classified as plants. Features in the environment a visually-impaired pedestrian might need to interact with were designated as useful objects. The landmarks category was comprised of distinctive features that could aid in navigation. The 'landmark' feature in this category referred to a unique historical site on campus. The category classifications of direct sound and size are self-evident. Six sounds were then constructed for each feature, one for each sonification design to be tested: auditory icons, earcons, speech, spearcons, earcon-icon hybrids and sized hybrids. The sounds ranged in duration from approximately .25 s to 4s.

### 2.3.1. Auditory Icons

In building the auditory icons, the initial focus was the object and its natural sound. Since most of the identified objects, such as streetlights or crosswalks, did not emit any kind of natural sounds, an indirect auditory representation was needed. As an example, a tree is represented by the sound of the wind going through the leaves mixed with the sound of bending wood. (All of the sounds used in this experiment can be listened to by clicking the links in Table 2.) In some cases there were no natural sounds that could be used as a representation (e.g., a crosswalk or a street light). In these cases musical instruments or the sound of the materials these objects were made of were used. The sounds were gathered from a comprehensive sound effects library. In most cases various sound files were mixed together to achieve the desired icon. Hints for category allocation are not included into the auditory icon sounds. Thus, each sound stands for a specific object and comprises neither a category teaser nor a size allocation. An auditory icon is simply the most natural representation of an object we could create. They are mostly short, straightforward and without additional object information.

### 2.3.2. Earcons

As mentioned previously, earcons are musical patterns that can be decomposed into five dimensions: rhythm, pitch, timbre, register and dynamics. Due to their ability to build hierarchies, the design of the earcons included the object categorizations. Thus, each earcon starts with an opening sound that represents the category the sound belongs to. We used distinctive instruments for each object category:

- Buildings: Whirly keyboard
- Intersection helpers: Dings and dongs, mallets
- Obstacles: Grand piano

- Plants: Drums and percussion sounds
- Practical objects: Flute
- Landmarks: Organ

After the category sound, the actual object sound begins. Each object was represented by a unique melody or rhythm. Since the chosen instruments and melodies were more or less arbitrary, we tried to choose the instruments to be an appropriate representation of the according category. For example, plants were assigned naturalistic percussion sounds like wood blocks. Natural mappings were also considered when designing the single melodies, examples of this are the two feature sounds for stairs. The melody displays the direction of the stairs in terms of an increasing or decreasing melody. Apple's 'GarageBand' software [12] was used to compose the teasers as well as the melody sounds.

### 2.3.3. Earcon-Icon Hybrids

Due to the fact that earcons are more or less arbitrary their learnability often suffers. On the other hand, each auditory icon is distinct and bears no categorical resemblance to other related icons. In order to use the strengths of each to overcome the weaknesses of the other, earcon-icon hybrids were developed by combining the opening sound of each object category from the earcon and the auditory icon of a specific object. Thus, each feature consists of an opening sound according to the category it belongs to and a unique icon sound.

### 2.3.4. Sized Hybrids

In order to give an impression of the size of an object a sound layer containing size information was added to the earcon-icon hybrid sounds. A size classification with four steps was introduced: Small, medium, large and huge. For each size category a unique melody was composed differing in pitch and duration. The sound representing huge objects is low pitched and long for example, whereas a short and high pitched two note melody is used for small features. Because the category teaser and the object sound are sequentially arranged we considered adding the size sound at the end of the icon sound. However, it was decided have the size sound play in parallel to the actual auditory icon in order to keep the sounds shorter. The size sounds were designated using frequencies such that they do not interfere with the actual object sound. The resulting sounds were checked to ensure that no masking effects took place.

### 2.3.5. Speech

To create a spoken phrase for each environmental feature, we used text-to-speech software, specifically the male voice 'Mike - US English' of the web application from AT&T [13], to create the entire set of speech-based feature sounds.

### 2.3.6. Spearcons

To create the spearcons the speech stimuli were compressed using a logarithmic algorithm coded in MatLab, as described by Palladino and Walker[11].

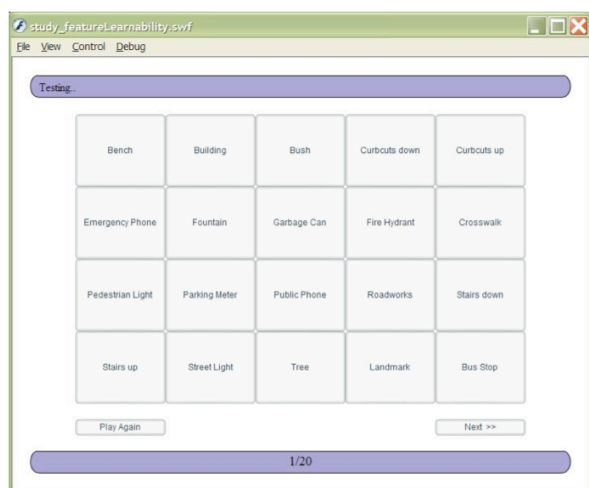


Figure 1. The grid that participants used to select an answer during the testing phase. Clicking the ‘Play Again’ button in the lower left corner allowed them to hear a sound as many times as they liked. The ‘Next’ button in the lower right corner indicated their answer choice was final.

#### 2.4. Procedure

Participants’ informed consent was obtained, their age was recorded and they were randomly assigned to one of the six sound conditions. Participants were given instructions and then began the experiment.

In the training phase of the experiment, participants were shown a single target word (e.g. “Bench”) and the sound associated with that environmental feature was played once. Participants would then advance the program to see the next feature and hear its associated sound. After being trained on all 20 stimuli, the testing phase would begin. Participants were presented with a grid containing all of the features presented in the training phase (see Figure 1). A sound from the training phase was then played, and participants were asked to select the environmental feature associated with that sound from the grid by clicking on it with the mouse. Participants were given the option to listen to a sound as often as they liked before making a selection by clicking a ‘Play Again’ button. Once they had made their final selection, they clicked on the “Next” button and the next sound was played. At the end of the testing phase, after having been presented with all 20 stimuli, participants were shown their performance (e.g. 12/20). If a participant had not answered all 20 items correctly, then the training phase was started again, after which another testing phase occurred. This process was repeated until a participant had successfully identified all 20 features correctly in a single testing phase. All answers given by participants were recorded by the software, which also noted the aggregate percentage correct of a given participant across all testing phases as well as how many training cycles were required to reach perfect performance.

### 3. RESULTS

The independent variable of sound type was analyzed with respect to the dependent variables of 1) the number

of training cycles required to reach 100% accuracy and 2) the aggregate percentage accuracy of a participant across all testing cycles. A multiple analysis of variance (MANOVA) found a significant effect of sound type,  $F(10, 64) = 9.66, p < .001, Wilk's Lambda = .159$ . Subsequent univariate tests showed a significant effect of sound type for both the number of training cycles,  $F(5, 33) = 10.77, p < .001$ , and aggregate percent accuracy,  $F(5, 33) = 20.15, p < .001$ .

In terms of the number of training cycles necessary to achieve 100% accuracy, the spearcons and speech sound types clearly required the smallest number of cycles (mean = 1.14, st. dev. = .378 and mean = 1.14, st. dev. = .378 respectively) which can be seen in Figure 2. Pairwise comparisons confirmed both sound types to require significantly fewer trials compared to all other sound types. Earcons required the largest number of cycles (mean = 8.50, st. dev. = 4.087). Pairwise comparisons determined this to be significantly more than all other sound types except for earcon-icon hybrids. The inclusion of a “size” attribute to the sounds led to no statistically significant different performance between earcon-icon hybrids and sized hybrids.

The aggregate percentage accuracy also showed spearcons and speech to be identical to each other (mean = 99.64%, st. dev. = .945 and mean = 99.64%, st. dev. = .945 respectively). Pairwise comparisons revealed both spearcons and speech to have a significantly higher aggregate accuracy compared to the other sound types. Earcons on the other hand had a significantly worse aggregate accuracy than any other sound type except for sized hybrids as indicated by pairwise comparisons. No statistically significant difference was found between the earcon-icon hybrids and the sized hybrids. All pairwise comparisons used a Bonferroni adjustment to control for Type-I error.

### 4. DISCUSSION

The principle finding of this study is that spearcons are as easy to learn as speech. Performance with both dependent measures was identical, with almost no errors across any trials and very few participants taking more than one cycle to identify all the feature sounds correctly. This seems to indicate that spearcons, like speech, require virtually no learning to comprehend. In addition, spearcons are faster than speech, and hence do not occupy as much of the display time as speech. Spearcons are also not actually speech, which allows the speech channel to be left unimpeded while they are being used. Taking into account all of these advantages, it is clear that spearcons are a distinct and useful sonification methodology.

Another interesting finding is in the results of the two novel sound types, earcon-icon hybrids and sized hybrids. In terms of both dependent measures, combining earcons and auditory icons would lead to better performance than earcons alone. This increased learning performance is likely due to the familiarity that the auditory icons lend to the sounds. However, both earcon-icon hybrids and sized hybrids showed worse learning performance than auditory icons alone. This is possibly due to the fact that these two new sound types are much more complex than the auditory icons, and therefore possibly more difficult to learn. While these two sound types do allow the hierarchical structuring of auditory icons, the overshadowing performance of

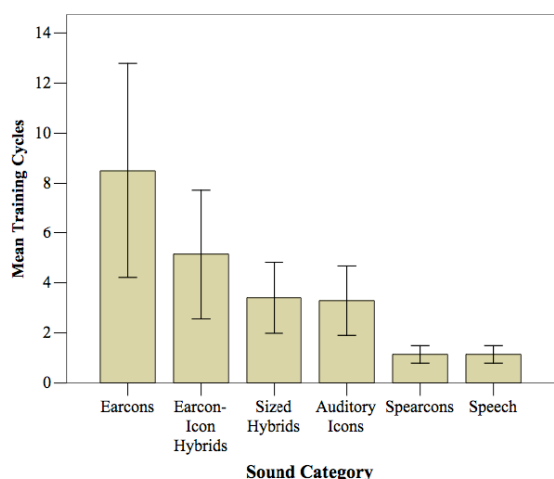


Figure 2. Mean number of training cycles needed to reach 100% accuracy in a testing phase. The error bars indicate the 95% confidence intervals of the means.

spearcons and speech make those far more appealing options when interface learnability is a concern.

One aspect of the earcons, auditory icons (and the hybrid sounds that derive from them) that can potentially be important in an auditory interface is localizability. Speech is relatively difficult to localize [14] compared to some of the more broad spectrum sounds that comprise auditory icons and, possibly, earcons. While not always an important factor, interfaces that rely on spatialized sound as a fundamental aspect of the display [e.g., 15] must certainly take this into account.

## 5. CONCLUSIONS

In conclusion, spearcons have once again proven to be comparable to speech, this time in regard to learnability. At the same time, they are different enough to leave the speech channel open and are briefer and therefore occupy less display time. This reinforces their potential as an excellent sonification methodology. Also, while fusing auditory icons and earcons does allow for a combination of some of their strengths, it also dilutes the learnability of the auditory icons, which is one of their principle advantages. Additional work is underway studying the utility of each of these sound types in a navigation task. Comparison of navigation performance as well as learnability will be important in determining the best sound display solutions for representing environmental features.

## 6. REFERENCES

- [1] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human-Computer Interaction*, vol. 2, pp. 167-177, 1986.
- [2] 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.
- [3] K. Hemenway, "Psychological issues in the use of icons in command menus," presented at CHI'82 Conference on Human Factors in Computer Systems, New York, 1982.

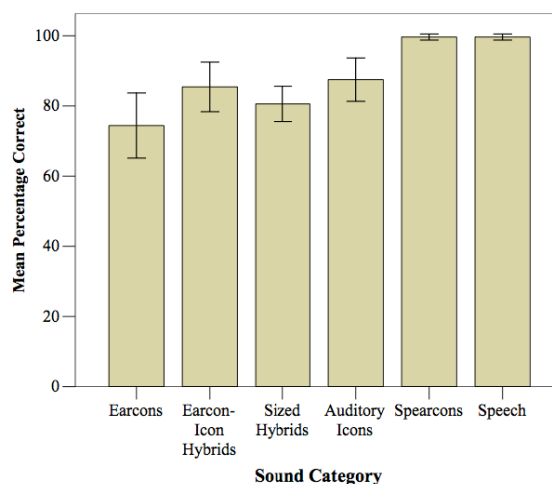


Figure 3. Mean percentage accuracy of participants with each sound category across all trials. The error bars represent the 95% confidence intervals of the means.

- [4] P. Kolers, "Some formal characteristics of pictograms," *American Scientist*, vol. 57, pp. 348-363, 1969.
- [5] P. Keller and C. Stevens, "Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons," *Journal of Experimental Psychology: Applied*, vol. 10, pp. 3-12, 2004.
- [6] B. N. Walker and G. Kramer, "Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making," in *Ecological psychoacoustics*, J. Neuhoff, Ed. New York: Academic Press, 2004, pp. 150-175.
- [7] B. N. Walker, A. Nance, and J. Lindsay, "Spearcons: Speech-based earcons improve navigation performance in auditory menus," presented at International Conference on Auditory Display, London, England, 2006.
- [8] S. Brewster, P. C. Wright, and A. D. N. Edwards, "A detailed investigation into the effectiveness of earcons," presented at First International Conference on Auditory Display, Santa Fe, New Mexico, 1992.
- [9] R. D. Patterson, "Guidelines for auditory warning systems on Civil Aircraft," Civil Aviation Authority, London 1982.
- [10] B. N. Walker and J. Lindsay, "Effect of beacon sounds on navigation performance in a virtual reality environment," presented at Ninth International Conference on Auditory Display ICAD2003, Boston, MA, 2003.
- [11] D. Palladino and B. N. Walker, "Learning rates for auditory menus enhanced with spearcons versus earcons," presented at International Conference on Auditory Display, Montreal, Canada, 2007.
- [12] Apple, "GarageBand," 2007.
- [13] I. AT&T Labs, "AT&T Natural Voices," 2007.
- [14] T. V. Tran, T. Letowski, and K. S. Abouchakra, "Evaluation of acoustic beacon characteristics for navigation tasks," *Ergonomics*, vol. 43, pp. 807-827, 2000.
- [15] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert, "SWAN: System for Wearable Audio Navigation," presented at International Symposium on Wearable Computers, Boston, MA, 2007.