Bachelor Thesis IK:
Learnability of sounds as representations of two dimensional menu structures

Christiaan Stierman
5767717
c.stierman@gmail.com

Abstract

Three kinds of sounds were evaluated as audio representations of two dimensional menus. The first kind, auditory icons, were sounds of real life items or materials the user might relate to what was represented, like the sound of scissors cutting as a representation for a "Cut" menu item. The second kind, spearcons, were Text To Speech (TTS) samples which were sped up, so that they can be accurate and quick. The third kind, introduced by this study, were enhanced spearcons. The enhancement was the addition of locational cues in two dimensions, vertical position through pitch and horizontal position through stereoposition. For each type learnability, accuracy and response time were measured.

Auditory icons and spearcons performed on par. Because spearcons can be easily generated however, performing on par is enough for spearcons to be preferable over the difficult to create auditory icons. The spearcons enhanced with locational information did not outperform normal spearcons, they even performed worse in accuracy. Further research is necessary to determine why the enhanced spearcons underperformed and whether or not adding locational data to spearcons is definitely undesirable.

Contents

1 Introduction **4**

1.1 Problem: Menus. 4

1.2 Solution: Sound, but which sound? 4

1.2.1 Earcons 4

1.2.2 Auditory icons 5

1.2.3 Spearcons 5

1.2.4 A comparison 6

1.3 Research Goals 6

2 Method **7**

2.1 Measures 7

2.2 Menu Structure 8

2.3 Auditory Stimuli 8

2.3.1 Auditory icons 9

2.3.2 Spearcons 9

2.3.3 Enhanced spearcons 9

2.4 Participants 9

2.5 Procedure 10

2.6 Apparatus and Equipment 10

3 Results **13**

3.1 Learnability 13

3.2 Response Time & Accuracy 14

3.3 Control Statistics 15

4 Discussion **16**

4.1 Learnability 16

4.2 Accuracy & Response Time 17

4.3 Recommendations 17

5 Conclusion **18**

5.1 Wrap-up 18

5.2 Limitations 18

5.3 Future Work 18
5.4 Acknowledgements 19

6 Appendix **20**

A Sample audio **20**

B Informed Consent Form **21**

C Questionnaire **22**

1 Introduction

1.1 Problem: Menus.

Digital menus are a well developed interaction technology. Many applications make use of simple drop down menus to bring structure to an otherwise confusing heap of functionalities. By now we are all used to looking for the right category on the top of our screen to subsequently find and click on the desired functionality in the two-dimensional list that follows. These menus are not unique to desktop interfaces. Most interactive devices with screens have two-dimensional menu structures to present the user with a set of options he can choose from. Tvs, mp3 players, gps systems, radio systems and so on are all prone to use these kinds of navigation aids.

Two-dimensional menu structures on visual displays work just fine in a stationary desktop environment. As technology becomes more ambient, mobile and more of a secondary supportive tool in other activities, visual interfaces may not always be adequate or even a possibility anymore. What if the users eyes are to be devoted to other things? What if there is no screen? An example of the latter are *Imaginary Interfaces* as described by Gustafson et al. (2010), where spacial interaction is brought to screenless devices. Could we somehow make the best out of another form of human computer interaction? A possible sensory system to support or even replace visuals in these situations would be the auditory system. Hearing is equally fast as seeing and it can process complex stimulations, as our ability to communicate through speech confirms.

1.2 Solution: Sound, but which sound?

Research on audio as a representation form for the navigation of two-dimensional menu structures has already provided some insights.

1.2.1 Earcons

One option is formed by earcons. Brewster et al. (1994) define these as *"abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of an interface"* (Brewster et al., 1994, p. 2). Since these sounds are abstract and synthetic, they have a set of easily adjustable parameters such as timbre or pitch that make them sound similar yet different. They can also be made to form unique melodies. This means you can make sounds related to each other to form groups of sounds. One can also go further and express proper multi -layer hierarchies by defining variables to indicate position in a hierarchy. An example would be to add a new note for each layer in a hierarchy, effectively creating a auditory code that can be translated into an exact position in a system. Similarly one could create earcons that express how deep and in which category the represented item is located. In previous research subjects turned out to be able to determine their location in a menu hierarchy and even determine where previously unheard sounds would belong in this hierarchy (Brewster et al., 1996).

Earcons are more effective than arbitrary sounds because of their connection to the position of the item they represent in a menu or hierarchy. However, they have no direct connection to the item they represent itself. By the framework of Peirce (1894) earcons would be symbols. The sounds do not have any link with the physical properties of what they represent. The locational cues they express are often based on artificial conventions where for instance different melodies represent different branches in a hierarchy. This can be an advantage, because it enables earcons to be able to represent anything, regardless of what it is, by its position. Since earcons have no direct connection to the item they represent, this item does not have to meet any criteria to be able to be represented. It can also be a disadvantage as determining what a sound represents only by positional cues could be more difficult than with the help of a sound that has a stronger relationship to the represented item.

1.2.2 Auditory icons

A second representation approach is that of auditory icons, *"everyday sounds meant to convey information about events in the computer by analogy with everyday events."* (Gaver, 1989, p. 1). *"Auditory icons exploit people's tendencies to listen to sources by mapping attributes of everyday sound-producing events to attributes of the model world of the computer."* (Gaver, 1989, p. 5). These sounds are real life sounds associated with the object represented in the menu. An example would be the sound of a dog barking to represent a dog. The barking sound has a direct connection to the dog that is represented. A more relevant example would be the sound of a piece of paper hitting a metal object to represent the action of putting a digital file into the digital trash can . A visual object, a trashcan, is a metaphor for a repository for bytes to be deleted. Another visual object, a piece of paper, is a metaphor for the object to be deleted. Both visual objects have a sound associated with them, the sound of metal for the trashcan and the sound of paper for the file. These metaphors can be used to represent action of moving bytes to the repository by playing the sound of a piece of paper hitting a trash can. When an object doesn't have an iconic sound associated with it, the sound tries to resemble its attributes, like size or material. A small file would be represented by the sound of a flapping piece of paper while a big file system could have a heavier more book-like sound.

Auditory icons do not include information about the position an item has in a menu or hierarchy. They rely on their link to the actual objects represented to be more effective than arbitrary sounds. The idea is that by picking a sound the user is already familiar with, it easier for the user to remember the exact meaning of a sound. This only works when the user recognizes the sound and associates it with the item represented. Auditory icons heavily depend on their relation to an object for success (Bussemakers and Haan, 2000). Making again use of the trash bin example, if the user has never heard folding paper or the metallic sound of a trash bin, or if the user is not familiar with the trash bin metaphor for deleting data, the auditory icon becomes less helpful. The nature of these relationships qualifies them as partially iconic and indexical in the framework of Peirce (1894). Auditory icons mimic certain physical properties of items associated with what they are trying to represent, or indicate. This means success for auditory icons heavily depends on general accessibility of what they represent. How does one pick an everyday sound with a clear relation to "insert new paragraph" for a menu in a word processing suite? A lot of functionalities in menus are very hard to represent in auditory icons.

1.2.3 Spearcons

A third option, a more recent development in the field, are speech-based earcons, or spearcons. Walker explains: *"Spearcons are created automatically by converting the text of a menu item (e.g., "Export File") to speech via text-to- speech (TTS), and then speeding up the resulting audio clip (without changing pitch) to the point that it is no longer comprehensible as speech."* (Walker et al., 2006, p. 64). The result is an automatically creatable sound, unique to what it represents. The TTS samples are sped up to decrease response time and decrease the speaker time required to play a spearcon, resulting in more speaker time for other audio feedback. If the speed increase is not applied the TTS messages would be so long they would start interrupting each other and user response rate would be slower as well (Palladino and Walker, 2008). Early spearcons (Walker et al., 2006) would apply the same amount of speed increase to each sound. In later research (Palladino and Walker, 2008) a logarithmic compression form was used so that longer words get compressed more than short words. This was done to avoid small one syllable words from becoming mere "clicks".

Spearcons are not dependent on the strength of a relationship to a physical object like an auditory icon but they do retain a link to what they represent. Instead they rely on language, resulting in easier learning than arbitrary sounds. Language itself is a social construct, words only have meaning because people attached them to meaning. Language is a structure of symbols by the framework of Peirce (1894). The good thing about language is that everybody spends most their life learning it. Spearcons depend on this by using

audio representations resembling speech. Because spearcons only require a link to a word for what they represent, they are hard to set up but easy to automate. Spearcons require text to speech and advanced audio processing to make, but once these are set up to work automatically spearcons can be generated on the fly without human intervention. This means spearcons do, however, depend on the quality of the text to speech generator used and the quality of the sound processing suite used.

1.2.4 A comparison

Three kinds of sounds suitable for menus have been described. How do these compare to each other? How can we decide which kind of sound is most appropriate in a two dimensional menu structure?

Earcons and auditory icons, both well established methods, have been thoroughly evaluated and compared. Comparisons between the two methods have been made in an abundance of variables. Bonebright and Nees (2007) found auditory icons to be more learnable than earcons. Bussemakers and Haan (2000) found auditory icons to be more efficient than earcons. Garzonis et al. (2008) found auditory icons to be more memorable than earcons, although they do heavily depend on their the metaphor related to the represented object for success. In a later publication Garzonis et al. (2009) found auditory icons to be superior to earcons in intuitiveness, learnability, memorability and user preference. Duarte and Carriço (2007) conclude auditory icons are more pleasurable and less intrusive than earcons. It is clear that earcons are no match for auditory icons on any of these fields. Spearcons have shown to be more efficient than earcons (Walker et al., 2006) and easier to learn than earcons by a factor five (Palladino and Walker, 2007). These previous evaluations of earcons led them to be excluded from the present experiment. Earcons only really have one advantage over the other two kinds of sounds, they allow themselves to express their position in a hierarchy really well. Not only can they easily be altered in pitch and timbre, but they can also be composed to express complex hierarchical information in a melody. Most two dimensional menus however, do not posses a deep hierarchy so this advantage is not worth the massive loss in performance.

Spearcons, being a relatively new method, have been particularly evaluated in terms of learnability and efficiency. All of this evaluation was (co)written by Bruce N. Walker, the inventor of spearcons. Walker has found spearcons to be more efficient than auditory icons (Walker et al., 2006) and conventional Text To Speech (Palladino and Walker, 2008). Spearcons outperformed auditory icons (Dingler et al., 2008) in learnability. Dingler et al. actually found spearcons to be almost as learnable as actual Text To Speech. If these results are accurate there is no reason to use auditory icons over spearcons. Apart from performing better, spearcons are also easier to make, without imposing requirements on what they represent. Assuming menus represent objects with textual reference, spearcons only need this textual reference while auditory icons require a sound people would relate to the object represented, not always an easy task and certainly a task hard to automate. Another advantage over auditory icons is their shorter length. Spearcons are compressed so they take up less time on the speakers. Auditory icons are usually longer than than spearcons, resulting in more required speaker time.

1.3 Research Goals

The first goal of the present research was to confirm the performance of spearcons compared to auditory icons. Previous research in spearcons is scarce and done by the people who created spearcons in the first place. Based only on these previous works one would feel inclined to opt for the use of spearcons. Although their findings are consistent, they consist of too little experiments to be considered conclusive. The goal here was to replicate some of this previous work and generate new data to confirm whether or not spearcons are indeed superior to auditory icons.

The second goal of the present research was to see how an enhanced form of spearcons performs compared to conventional spearcons. Palladino and Walker mention the following: *"It should be pointed out that spearcons,*

as originally formulated, do not necessarily provide the navigational information (i.e., which menu is active) that hierarchical earcons are designed to provide. However, this can be obtained by using more sophisticated spearcons that vary by, for example, gender of the speaker, or incorporate other kinds of navigational cues." (Palladino and Walker, 2007, p. 64). They leave it at that. The research for this thesis seeks to make an explicit comparison between conventional spearcons and these more sophisticated spearcons which will from now on be referred to as "enhanced spearcons".

The proposed enhancement aims to add positional information cues to spearcons. Positional information has been widely evaluated for earcons. Parameters used to express positional information in earcons include register, pitch, timbre, stereo position, melodies, intensity and length (Brewster et al., 1996; Leplâtre and Brewster, 2000). Use of intensity is not advised because it leads to annoyance for users (Brewster et al., 1995). Length is not an option because using length means the user has to wait until a sound is over before he can make up his mind, resulting in longer reaction times. Melodies are only applicable to earcons. Timbre is effective but only when the difference between timbres is clearly audible (Brewster et al., 1995). Initially this experiment was to use different voices as different timbres but technical difficulties prevented access to this functionality in the TTS suite used to make spearcons. Different registers were not supported by available TTS applications either. This left pitch and stereo position. The enhanced spearcons used stereo position to indicate where the object was located horizontally. Pitch was used to express the vertical position of the object represented. Conveniently pitch is expressed in terms of high and low. A low pitched sound would represent an object that was lower on the screen than one represented by a high pitched sound. This enhancement is aimed to give spearcons a new dimension in the framework of Peirce (1894). Rather than just relying on the symbolic nature of speech, the positional cues have more of an indexical nature to complement it.

It is worth noting that out of these parameters only stereo position is an option for possible enhanced auditory icons. Changing the pitch or register of an auditory icon simply results in a different sound, which means it loses the link with what it represents. Changing timbre is also impossible, as it is hard to determine what the timbre is for a sound that is not musical or vocal. Timbre is the quality of a sound that distinguishes it from other sounds. In earcons different musical instruments are used to produce different timbres. Voices can sing the same note yet sound differently for TTS to give spearcons different timbres. What is a different timbre for the sound of a trash bin hit by a crumpled piece of paper though? It is hard to imagine one. The result is that spearcons have four different parameters one could use to enhance them to express locational data, of which stereo position and pitch will be tested in the experiment. Meanwhile, out of the parameters found in earcon litarature, only stereo position is applicable to auditory icons. This means a similar experiment with auditory icons enhanced to express two dimensional locational data is impossible, unless a new parameter is found to express vertical position in the way pitch does for the proposed enhanced spearcons.

2 Method

2.1 Measures

How was the performance of the audio representations measured?

The primary measure of performance for the auditory icons, spearcons and enhanced spearcons as representations of items in two dimensional menus was learnability. Arguably most sounds could function as functional representations of a two dimensional menu. If a person hears a certain sound in association with an event often enough, even a completely synthetic unintuitive sound will obtain a "natural meaning"(Ulfvengren, 2007). Good sound interaction design aims to reach this stage as quickly as possible by making sounds more learnable. Even if all kinds of sounds can become natural to a user, the effort this process takes can differ immensely. By ensuring the learning process for a sound association is as short as possible, an application becomes functional at an earlier stage. Ideally sounds would require no learning at all to function. If sounds

Animals:	People sounds:	Objects:	Nature:	Instruments:
Bird	Snoring	Car	Ocean	Piano
Horse	Sneeze	Typewriter	Thunder	Flute
Dog	Clapping	Camera	Rain	Trumpet
Cow	Laughing	Phone	Wind	Marimba
Elephant	Cough	Siren	Fire	Violin

Figure 1: Used menu structure

are hard to learn and users use an application before the learning process is complete, this hurts performance in all aspects. They will misunderstand, make errors and lose time trying to remember what a sound represents. Therefore the main focus of the present experiment was evaluating the three kinds of sounds in learnability as representations for objects in two dimensional menus.

Secondary measures of performance for the auditory icons, spearcons and enhanced spearcons as representations of items in two dimensional menus were response time and accuracy. To test the primary measure, learnability, subjects were required to learn the meaning of a set of sounds in a menu. Once this process was complete, in theory, subjects for all three kinds of sounds were on the same level of proficiency with the sound they had been learning. Regardless of how much learning it took, they had passed the same task with their respective sound type as those with other kinds of sounds. At this stage it was interesting to measure response time and accuracy. Spearcons were generally shorter than auditory icons. Did this mean people interpreted and responded to them more quickly? The enhanced spearcons were the only one out of the three sound types with locational information. Did this mean people found them more quickly in a menu? Accuracy was good to measure because, no matter how fast people responded, if people were significantly inaccurate the sound representation was dysfunctional.

2.2 Menu Structure

The menu structure used in the present experiment, see Figure 1, originates from the article that introduced spearcons (Walker et al., 2006). They state they wanted to make sure their menu was unbiased against auditory icons. As a result they excluded traditional computer-themed menus as auditory icons can not be reliably created for items such as “select table” or “open new window”. They also state that the menu should not relate to previously learned menu structures. If a menu from a common application was used, it would be biased towards users already familiar with this structure. On the other hand the objects themselves should be common for the auditory icons to work. If the participants were not familiar with an item they would not be able to relate to the auditory icons either. The result was a menu structure of common items that are uncommon in a menu structure, divided into five clear categories. Convinced by their arguments and in order to make the present experiment and Walker his experiment more comparable, the present experiment used the same structure as the experiment conducted by Walker et al. (2006).

2.3 Auditory Stimuli

Sadly B.N. Walker refrained from answering a request to use his sounds. As a result all the sounds used were created specifically for the present experiment. Audio stimuli samples can be found in Appendix A.

2.3.1 Auditory icons

The auditory icons are a collection of royalty free sound files from online collections, found through SoundJax, a sound sample search engine. They all had a direct connection to the item they represent. All the samples sounded like recordings of sounds made by the item they represent. The length of the samples differed per item and was usually between 0.5 and 2 seconds. Where possible the samples were cropped. For instance, the original siren sound went on for five seconds but it was cut off after two wails, resulting in a one second sound.

2.3.2 Spearcons

The origin of the spearcons were TTS files generated by Festival (`http://www.cstr.ed.ac.uk/projects/festival/`), a free to use software package made to advance speech synthesis research. The default male voice was used with the standard "text2wav" script. This script simply used Festvox to create a TTS .wav file based on the text you feed it with. To make spearcons out of these files Audacity (`http://audacity.sourceforge.net/`), free audio recording and editing software, was used. The "change tempo" functionality was used to speed the sounds up without effecting pitch. Shorter sounds were compressed less than longer sounds to avoid them becoming mere clicks. Most sounds were around 0.25 seconds long, all sounds were shorter than 0.35 seconds. Because the software used to play the sounds for the experiment turned out to inexplicably heighten the pitch of the sounds it played, all spearcons had their pitch lowered by 25% using the Audacity "change pitch" functionality to compensate. In the experiment the pitch of the spearcons sounded natural for a male voice.

2.3.3 Enhanced spearcons

The enhanced spearcon files were based on the spearcon files. They were similarly edited in pitch, except that their height in the menu determined the percentage they were edited in. Brewster et al. (1995) advise staying between 125Hz and 5 kHz in pitch for synthesized earcons. The pitch for spearcons is harder to determine since speech does not stay on one pitch during a word. The text to speech the spearcons were based on is generated using a complex combination of tones that form something resembling speech. Audacity frequency spectrum analysis for the unmodified spearcons already included a small (<10%) percentage of tones below 125 Hz. Brewster et al. (1995) also state that pitch should also be supported by the timbre. A high pitch does not match a contra-bass. To make sure enhanced spearcon pitch still matched its timbre, speech, spearcons with varying enhanced pitches were evaluated to see if they still sounded somewhat natural to speech. After testing 45, 30 and 15% modifications in pitch, 15% was deemed to be the only modification that still sounded natural. The difference was in steps of 7.5%, meaning the highest spearcons were 15% higher than the middle ones, with a 7.5% higher tier between them. The lowest spearcons were 15% lower than the middle spearcons with a 7.5% lower tier in between them. Stereo position was also edited using Audacity. Sound spectrum analysis for the lowered spearcons seemed to indicate the % of waves below 125 Hz did not change much form the original spearcons.

The left category only transmitted on the left speaker, the second category was configured at 25% of the left to right spectrum in the Audacity configurations. The middle category was in the middle, the fourth at 75% and the fifth category only played from the right speaker.

2.4 Participants

This study was conducted with 32 participants distributed over three groups. Two experiments were interrupted, which left three groups of ten participants. The interruptions were caused by excessive noise halfway

through the tests, forcing the experiments to be stopped. The nature of the experiment did not allow later continuation as the learning process would be too different for the interrupted participants. Each Group included three females and seven males. The youngest participant was 18 years old and the oldest 59. The age means of the three groups were comparable, namely 29.1, 28.4 and 31.6. Participants were also asked to describe how good they considered their own hearing on a scale from one to five. The results did not significantly deviate per group. The overall mean for this question was 4.0 with a standard deviation of 0.97.

2.5 Procedure

The experiment started with the signing of an informed consent form (Appendix B).

Once consent had been given, participants were told that they first had to learn the sound representations of 25 menu items. It was explained each item would be lit green, top to bottom per category, and the participant would hear the sound representing the item twice. Participants were also told that once they had heard all the sounds, the sounds would be played in random order and they would be required to click the item the sound represented. If they did not achieve a 100% accuracy they would have to repeat this entire process, adding a learning cycle, the primary way of measuring learnability of a sound type. Once the participants understood all this, they could adjust the volume of the sound to their liking. Then the test started. This experiment tried to replicate parts of experiments performed by Palladino and Walker (2007). Their experiment also used amount of learning cycles before 100% accuracy is achieved as a measure for learnability.

The test was conducted using software built in pygame, a set of python modules made to form a platform for open-source game creation in the programming language Python. The experiment was inspired by the game of memory and pygame supports everything needed for this experiment: sounds, interactivity, visuals and measurements. The user interface was made to be clear and simplistic. The menu looked like a table and feedback to the user was in big color coded letters. The window filled most of the laptop screen. Figures 2, 3, 4 and 5 show the application in different stages of the experiment.

Once the participants had achieved 100% accuracy in the learning phase they were required to fill out a small questionnaire. This questionnaire measured how difficult the participants perceived the learning of the sounds to be, how hard they perceived picking the right items to be, whether they had trouble hearing the sounds, and how they estimated their own hearing ability. For each of these questions the participants could make use of a scale from one to five. Finally gender and age information were also obtained. The questionnaire can be found in Appendix C.

The final part of the experiment was meant to measure response speed and accuracy. The participants were instructed to listen to the sounds again and select the related menu item for each sound. It was made clear that this time accuracy and speed were essential. An adapted version of learning.py was used. This program omitted the learning phase and went straight to the testing phase. During this phase it measured response time per item, an overall average response time, errors and % accuracy. These numbers were all stored in text files in the results folders. Once participants conducted this final test they were thanked for their participation and the test was over.

2.6 Apparatus and Equipment

All experiments were conducted on a 13 inch macbook, late 2008 model. A set of Shure SHR840 headphones were used, which stop noise fairly well. The participants were seated behind a desk in a fairly quiet environment, but not a laboratory.

Sounds

Animals:	People Sounds:	Objects:	Nature:	Instruments:
Bird	Snoring	Car	Ocean	Piano
Horse	Sneeze	Typewriter	Thunder	Flute
Dog	Clapping	Camera	Rain	Trumpet
Cow	Laughing	Phone	Wind	Marimba
Elephant	Cough	Siren	Fire	Violin

You are hearing the green item: Laughing

Figure 2: Participant to learn the item "laughing"

Sounds

Animals:	People Sounds:	Objects:	Nature:	Instruments:
Bird	Snoring	Car	Ocean	Piano
Horse	Sneeze	Typewriter	Thunder	Flute
Dog	Clapping	Camera	Rain	Trumpet
Cow	Laughing	Phone	Wind	Marimba
Elephant	Cough	Siren	Fire	Violin

Which sound is this?

Figure 3: Participant to click the item represented by the played sound

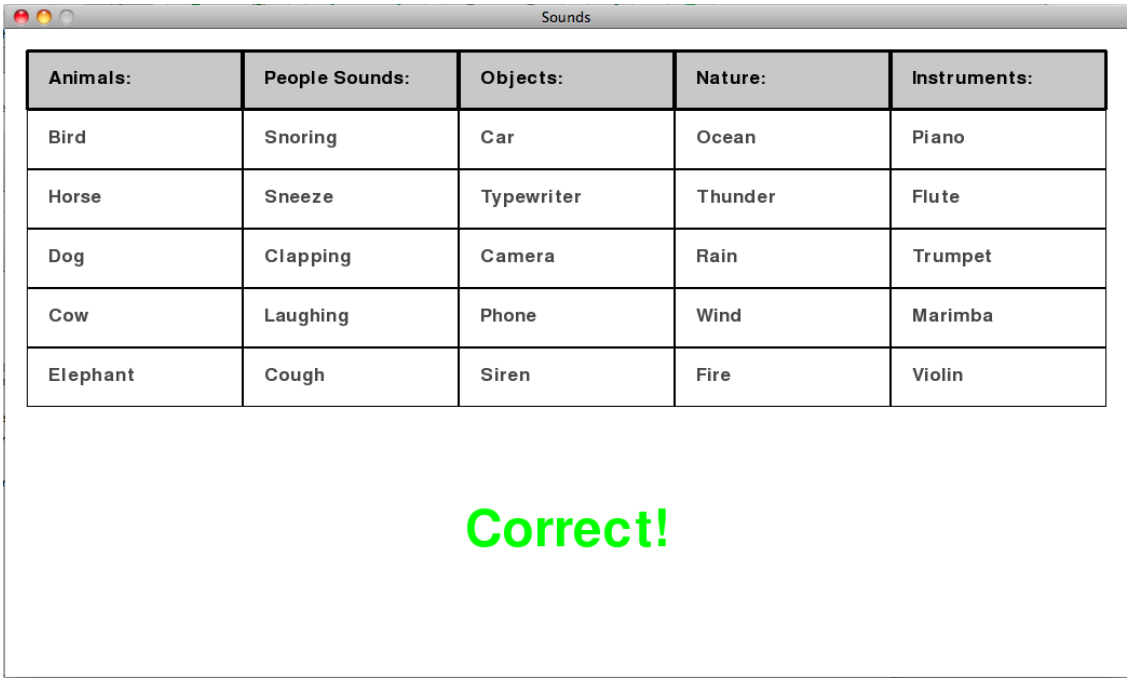

Figure 4: Participant just clicked the right item

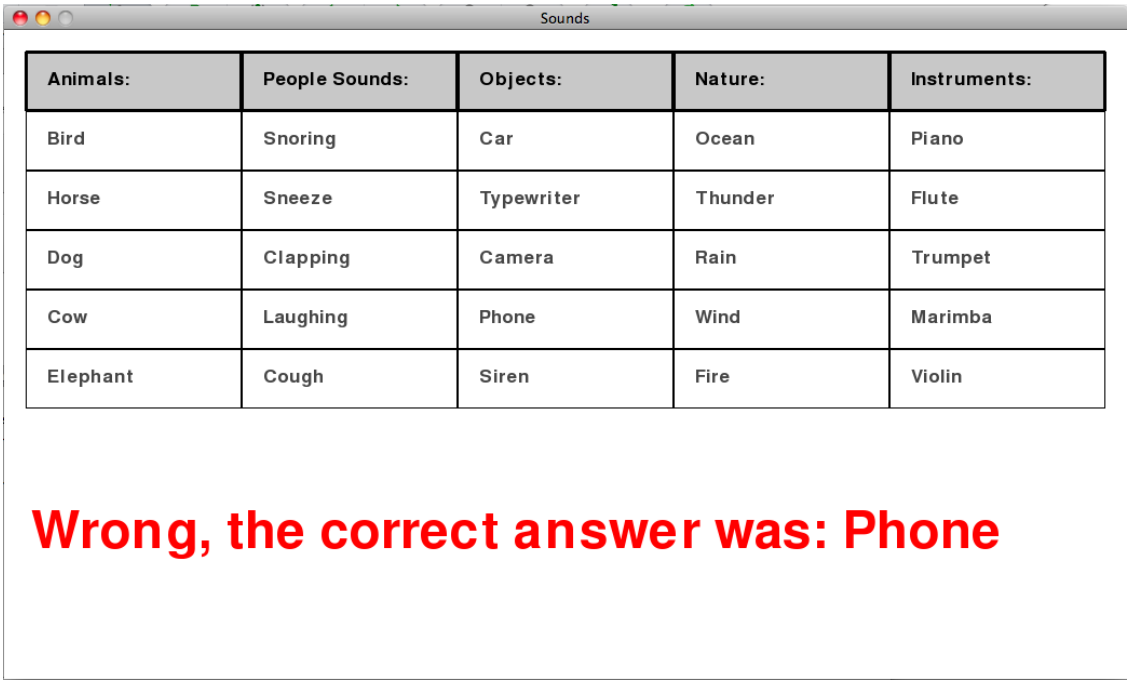

Figure 5: Participant clicked the wrong item

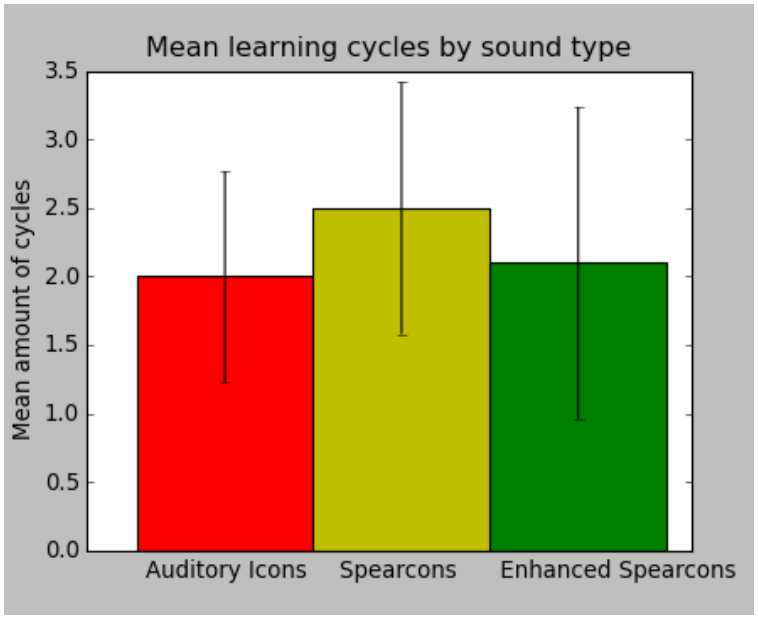

Figure 6: Learning cycles

3 Results

3.1 Learnability

Learnability was measured in two ways. The primary measure was the amount of learning cycles required to achieve a 100% accuracy. The results are displayed in Figure 6.

The mean amount of required cycles was about equal between the auditory icons and enhanced spearcons, while the normal spearcons took slightly longer to be learned. A one-way between subjects ANOVA was conducted to compare the effect of the audio type on the amount of required learning cycles for auditory icons, spearcons and enhanced spearcons. No significant effect was found: $F(2,27) = 0.69$, $P = 0.510$.

The secondary measure consisted of two questionnaire questions.

Q1:Did you feel like learning the sounds was difficult?

Q2:Did you feel like picking the right items was hard?

The questions could be answered on a scale from one to five where one equaled "not at all"and five equaled "very much so"as shown in Appendix C. While Q1 is evidently aimed at measuring learnability, Q2 could be interpreted in two ways. The intended interpretation was a whether or not it was hard to recall the item one had to press, measuring if the learning was succesful. Another interpretation was whether or not it was technically hard to select the item participants intended to click. Because of this regrettable ambiguity in the the question formulation, it is hard to determine which interpretation participants followed. This makes the results for Q2 less relevant. The results are visible in Figure 7. Once again, a lower score means that the item was more learnable.

The mean for auditory icons was lower than that of spearcons and enhanced spearcons. They also had a lower standard deviation than the other two groups. A one-way between subjects ANOVA was conducted to compare the effect of audio type on the answer on Q1 for auditory icons, spearcons and enhanced spearcons.

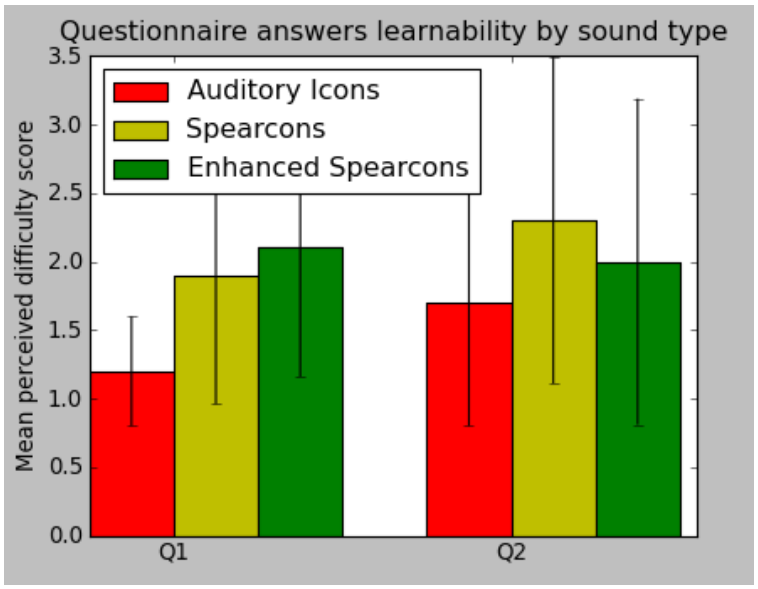

Figure 7: Results Q1 & Q2

No significant effect was found: F(2,27) = 3.108, P = 0.061. The criterion for significance in this study is P<0.05. The same analysis was conducted for Q2, where no significant effect was found: F(2,27) = 0.711, P = 0.500.

Because the ANOVA result for Q1 almost returned a significant difference, independent-samples t-tests were conducted between the three types for the scores of Q1. There was a significant difference in the scores for auditory icons (M=1.2, SD=0.422) and enhanced spearcons (M=2.1, SD=0.994) conditions; t (18)=-2.635, p = 0.017 (<0.05). There was no significant difference in the scores for auditory icons (M=1.2, SD=0.422) and spearcons (M=1.9, SD=0.994) conditions; t (18)=-2.049, p = 0.055. No significant difference was found between spearcons (M=1.9, SD=0.994) and enhanced spearcons (M=2.1, SD=0.994) condictions; t (18)=0.45, p = 0.658. This means auditory icons were perceived to be easier to learn than enhanced spearcons by participants themselves. Normal spearcons fell just short of statistical significance but the means still indicate spearcons were perceived to be harder to learn than auditory icons by participants themselves as well.

It is worth noting that the meager amount of participants, 30 distributed over three groups, may have played a part in the statistic analysis. For Q1 more participants might have resulted in a significant difference between spearcons and auditory icons next to the difference between enhanced spearcons and auditory icons. For the other two measures, with P values around 0.5, it seems unlikely more participants would quickly create a significant difference between sound types.

3.2 Response Time & Accuracy

After the extensive learnability tests were conducted, the assumption was that participants for all sound types would be on the same level of familiarity. At this moment a response time and accuracy test was conducted. The results for response time, measured from the moment the sound started playing, can be seen in Figure 8.

A one-way between subjects ANOVA was conducted to compare the effect of audio type on response time for auditory icons, spearcons and enhanced spearcons. No significant effect was found: F(2,27) = 0.382, P

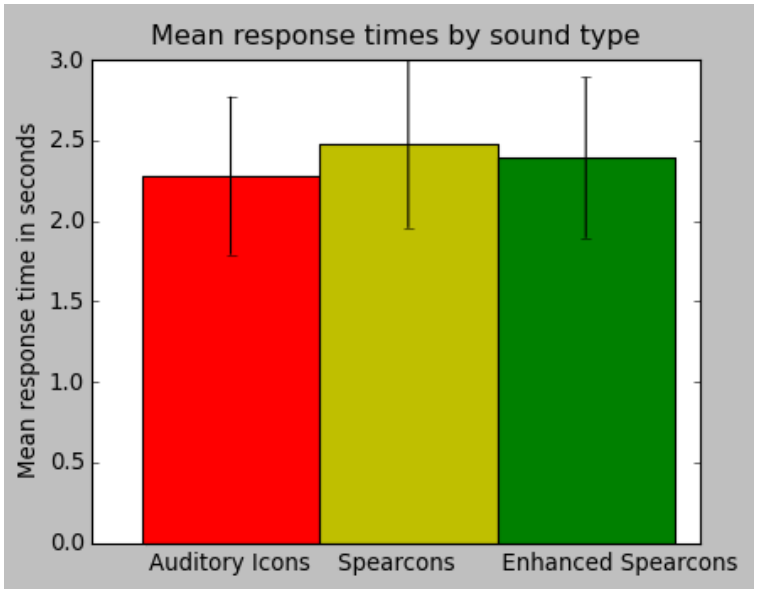

Figure 8: Response time

= 0.686. This is no surprise after looking at the bar charts, the response times for each time were indeed about equal.

Figure 9 shows the measured accuracy values. A one-way between subjects ANOVA, conducted to compare the effect of audio type on accuracy for auditory icons, spearcons and enhanced spearcons showed a significant effect: F(2,27) = 9.825, P = 0.001. The results for auditory icons are not normally distributed though. The data consists of one count of 84% accuracy, one count of 92% and eight counts of 100% accuracy. To account for this distribution a Kruskal-Wallis H test was conducted instead. This test yielded similar results: H(2)=13.306, P=0.001. This means one or more of the three types had significantly different accuracy results from the other(s).

To find out where the difference originated, independent-samples t-tests were conducted between the three types. There was a significant difference in the scores for enhanced spearcons (M=89.6, SD=4.695) and spearcons (M=92, SD=3.293) conditions; t (18)=-4.191, p = <0.001. There was also a significant difference in the scores for enhanced spearcons (M=89.6, SD=4.695) and auditory icons (M=5.4, SD=5.4) conditions; t (18)=-3.536, p = 0.002. No significant difference was found for spearcons (M=92, SD=3.293) and auditory icons (M=5.4, SD=5.4) conditions; t (18)=-0.2, p = 0.844. Since the auditory icon values were still not normally distributed however, Mann-Whitney U tests were also conducted where auditory icons were involved. These tests confirmed enhanced spearcons values significantly differed from those measured for auditory icons; U(18)=12, p = 0.003 while there was no significant difference between spearcons and auditory icon measurements; U(18)=39, p = 0.367.

3.3 Control Statistics

Participants were asked to describe how good they considered their own hearing to be on a scale from one to five. A one-way between subjects ANOVA was conducted to make sure the three groups were equipped with equal hearing abilities. No significant effect was found: F(2,27) = 0.097, P = 0.908. The medians ranged from 3.9 to 4.1, indicating the groups answered this question quite similarly indeed.

Participants were also asked whether they had trouble hearing the sounds on a scale from one to five. Because

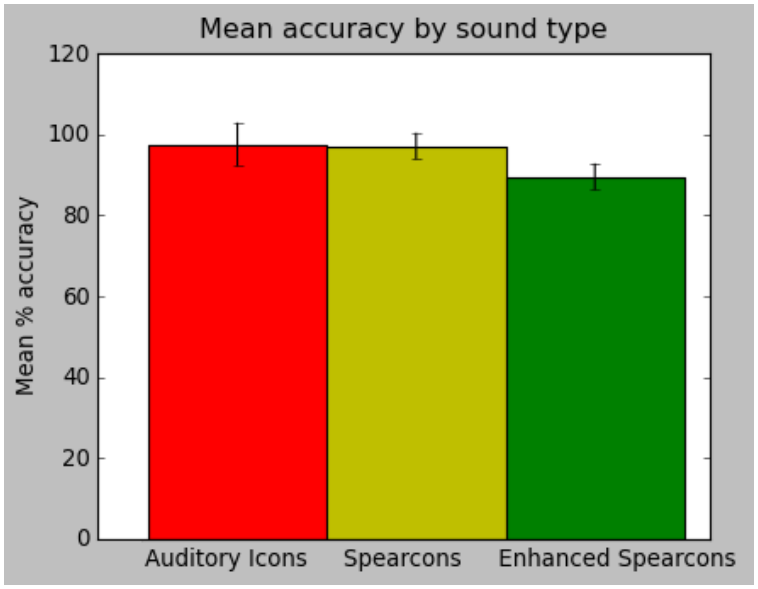

Figure 9: Accuracy

both enhanced spearcon and auditory icon values were not normally distributed, a Kruskal-Wallis H test was conducted rather than an ANOVA test. This test showed no significant deviation between sound types: H(2)=3.138, P=0.208. For what its worth, the mean for auditory icons was 1.3 while both spearcon groups had a mean of 1.7. This might be a slight indication spearcons are harder to hear or distinguish from each other than auditory icons, although no statistically significant difference was found.

4 Discussion

4.1 Learnability

To recap, no significant difference was found for learnability performance between any of the sounds, although participants perceived the learning of auditory icons to be slightly easier. This is a solid indication that spearcons perform about as well as auditory icons with respect to learnability as representations of digital menus. The only previous comparison between auditory icons and spearcons in learnability concluded that spearcons require significantly less learning than auditory icons (Dingler et al., 2008). Results from the present experiment contradict this by not finding a significant difference at all. In fact, the objective measurement of the learning cycles as presented in Figure 6 shows that spearcons took more time than any of the other two to be learned. If any of the two sound representations outperformed the other it would have been auditory icons, as they scored a slightly better mean in all three tests, see Figures 6 & 7. The Q1 part in Figure 7 also shows that participants perceived the learning of auditory icons as easier than any of the spearcon types, though only the difference between auditory icons and spearcons was significant. It becomes clear that the user experience can significantly differ from the real user behavior. Thus, experiential aspects, such as primary and secondary emotions, novelty, or quality of sound, need further investigation to clearly determine how the different sound representations really function. This might mean that, although the actual learning process did not take substantially longer, auditory icons might still be preferable to spearcons with the user experience in mind.

Enhanced spearcons yielded no advantages over spearcons in any learnability measurement. They scored

slightly better than normal spearcons in learnability cycles, but they also had a higher standard deviation. This might indicate that the enhancements helped some while they hindered others. Regardless, without a significant improvement in the results the enhancements dont seem to improve spearcon learnability.

4.2 Accuracy & Response Time

Walker et al. (2006) found spearcons to outperform auditory icons significantly in both mean response time and accuracy. Results from the current experiment do not confirm their findings, see Figures 8 & 9. No significant difference between auditory icons and spearcons was found for either measurement. This still means that spearcons perform well and might be desirable over auditory icons for other reasons than accuracy and response time. It is surprising spearcons do not have a lower response time than auditory icons though, since the played sounds are a lot shorter for spearcons. This means participants either identified auditory icons well before playback was completed, or that spearcons required participants to think for a longer period before they could make up their minds.

Unexpectedly, enhanced spearcons performed worse than normal spearcons. Response time did not deviate but accuracy was significantly lower than that of both normal spearcons and auditory icons. This means the enhancements did not only fail to improve the spearcons, but they actually hurt accuracy.

4.3 Recommendations

How does it all add up? Can one of these sound types be recommended as an audio representation for two-dimensional menu structures over the other two?

Spearcons mostly performed on par with auditory icons. The only reason to use auditory icons that can be derived from the present experiment are the results for perceived difficulty of learning the sounds, as auditory icons almost scored significantly better than spearcons in Q1. This one statistic pales in comparison to the mentioned advantages of spearcons. Spearcons can be automatically generated, while auditory icons need to be hand picked to match the item they represent. Spearcons are also applicable to any item, they do not depend on a metaphor to an object or action like auditory icons. Spearcons also take less time to play. Even though the present experiment denies this leads to a quicker response time as well, it still leaves more time for other, perhaps more important, audio stimuli to be played from the same channel. For all of these practical reasons and the fact that spearcons are equally learnable, as fast and as accurately detectable, it can be stated that spearcons might be a better representation for two-dimensional menus than auditory icons.

This leaves the enhanced spearcons. Sadly the addition of locational data through stereo position and pitch did not yield any gains, it merely caused reduced accuracy in the present experiment. How could this be? When the enhanced spearcon experiments started yielding surprisingly negative results halfway through the experiments, participants were asked if they even noticed the differences in stereo position and pitch and the link these differences had to location. Some noticed just stereo position, others only noticed pitch, others did not notice any differences. None of the five participants asked noticed all of the enhancements and understood their link to location. On the other hand some participants did express trouble hearing the sounds with an artificially lower pitch, perhaps causing the reduced accuracy. This leads to the conclusion that enhancements made to spearcons in the present research were not beneficial and they are not to be recommended in their present form. Normal spearcons were found to be the better choice.

5 Conclusion

5.1 Wrap-up

Three kinds of sounds were evaluated as representations of two dimensional menu structures. They were auditory icons, spearcons and spearcons enhanced to express locational data.

First and foremost this research showed that spearcons are indeed a good alternative to auditory icons. It also rejected the notion that spearcons are superior to auditory icons in learnability, accuracy and response time, as is suggested by spearcon evaluation authored by the inventors of spearcons(Walker et al., 2006; Dingler et al., 2008). The present experiment discerned that spearcon performance in these fields merely equals that of auditory icons. Because spearcons are a lot more practical in their creation than auditory icons however, it is concluded that spearcons are indeed recommendable over auditory icons as representations of two dimensional menu structures.

In an attempt to improve spearcons an enhanced set of spearcons was developed. These enhanced spearcons expressed their position in two dimensions through enhanced pitch and stereo position. Evaluation of this set of sounds determined they were equal in learnability and response time, but slightly less accurate than normal spearcons. This meant the enhancements made to spearcons in the present experiment only harmed performance and could not be recommended untill future research might prove otherwise.

5.2 Limitations

The biggest limitation to this research was a small amount of resources. Being conducted by one bachelor student without any funding, it was a struggle to find enough participants willing and able to conduct a 15-20 minute experiment in silent surroundings. Another limitation was the used set of spearcons. If the spearcons used by Palladino and Walker (2007) were available the results of Walkers studies of earcons and the present experiment would be a lot more comparable. A third limitation was the crudeness of the way learnability was measured in the present experiment. In favor of making results more comparable to previous research, learnability was primarily measured in learning cycles to 100% accuracy. This means one participant could make 25 errors in his first cycle and ace the second cycle while another participant could take a lot more than two cycles by failing at only one sound per cycle, even though he learned the other sounds right away. In this scenario the current experiment would deem the latter participant his learning process the hardest, even though he made a lot less errors per sound than the former. A less crude measurement, one that would have taken the amount of errors per cycle into account, may have provided more explanations to the disappointing results for enhanced spearcons.

5.3 Future Work

Even though the present experiment did not yield favorable results for adding locational information to spearcons, this innovation should not be forsaken without more evaluation. Would people actually learn a lot faster if they were briefly explained that there was a link between stereo position, pitch and location? Would using different timbres for the voices rather than pitch or stereo position yield better results? A future experiment could be conducted similarly to the present experiment, but with finer measurements for learnability. A future study could also measure errors per sound per learning cycle, rather than just measuring if the accuracy is 100% or not. This way different degrees of audio, enhancements to express locational data, on different kinds of spearcons could be evaluated seperately in learnability. Hopefully this would lead to a kind of spearcon with locational cues that outperforms spearcons without them.

5.4 Acknowledgements

This thesis was made possible thanks to:

- Dr. Frank Nack, guiding teacher throughout this project.
- The University of Amsterdam, facilitating my education.
- Audacity, pygame, Festival, Statistics Open For All, matplotlib and MiKTeX, free software packages used.

References

Bonebright, T. and Nees, M. (2007). Memory for auditory icons and earcons with localization cues. In *Proc. ICAD*, pages 419–422.

Brewster, S., Raty, V., and Kortekangas, A. (1996). Earcons as a method of providing navigational cues in a menu hierarchy. *In Proceedings of HCI'96 (Imperial College, London, UK), Springer,*, pages 169–183.

Brewster, S., Wright, P., and Edwards, A. (1994). A detailed investigation into the effectiveness of earcons. In *SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-*, volume 18, pages 471–471. Citeseer.

Brewster, S., Wright, P., and Edwards, A. (1995). Experimentally derived guidelines for the creation of earcons. In *Adjunct Proceedings of HCI*, volume 95, pages 155–159. Citeseer.

Bussemakers, M. P. and Haan, A. D. (2000). When it Sounds like a Duck and it Looks like a Dog . . . Auditory icons vs . Earcons in Multimedia Environments. *Proceedings of the International Conference on Auditory Display (ICAD)*, pages 184—-189.

Dingler, T., Lindsay, J., Walker, B. N., and St, C. (2008). LEARNABILTIY OF SOUND CUES FOR ENVIRONMENTAL FEATURES : AUDITORY ICONS , EARCONS , SPEARCONS , AND SPEECH. *Proceedings of the 14th International Conference on Auditory Display, Paris, France, June 24-27, 2008*, pages 1–6.

Duarte, C. and Carriço, L. (2007). on Digital Talking Books 2 Audio Presentation Techniques in Rich Digital Talking Books. *C. Stephanidis (Ed.): Universal Access in HCI, Part III, HCII 2007, LNCS 4556*, pages 259–268.

Garzonis, S., Bevan, C., and Neill, E. O. (2008). Mobile Service Audio Notifications : intuitive semantics and noises. *Proceeding OZCHI '08 Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, pages 156–163.

Garzonis, S., Jones, S., Jay, T., and Neill, E. O. (2009). Auditory Icon and Earcon Mobile Service Notifications : Intuitiveness , Learnability , Memorability and Preference. *Proceeding CHI '09 Proceedings of the 27th international conference on Human factors in computing systems*, pages 1513–1522.

Gaver, W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction*, 4(1):67–94.

Gustafson, S., Bierwirth, D., and Baudisch, P. (2010). Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 3–12. ACM.

Leplâtre, G. and Brewster, S. (2000). Designing non-speech sounds to support navigation in mobile phone menus. In *Proceedings of the International Conference on Auditory Display (ICAD)*, number April, pages 190–199. Citeseer.

Palladino, D. and Walker, B. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons. In *Proceedings of the 13th International Conference on Auditory Display*, pages 274–279.

Palladino, D. and Walker, B. (2008). Navigation efficiency of two dimensional auditory menus using spearcon enhancements. In *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 52nd ANNUAL MEETING—2008*, volume 52, pages 1262–1266. Human Factors and Ergonomics Society.

Peirce, C. (1894). What is a sign? *MS 404*, pages 1–9.

Ulfvengren, P. (2007). Design of natural warning sounds in humanmachine systems. *Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada, June 26-29, 2007*, pages 146–153.

Walker, B. N., Nance, A., and Lindsay, J. (2006). SPEARCONS : SPEECH-BASED EARCONS IMPROVE NAVIGATION PERFORMANCE IN AUDITORY MENUS. In *Proceedings of the 12th International Conference on Auditory Display, London, UK, June 20-23, 2006*, pages 63–68.

6 Appendix

A Sample audio

These are sample audio files for the Bird and the Violin item for each of the three tound types. Bird and Violin were chosen because they were in opposite corners of the menu. This way the enhanced spearcons fully showcase the used pitch and stereoposition ranges.

Auditory icons:

Bird and Violin .

Spearcons:

Bird and Violin .

Enhanced spearcons:

Bird and Violin .

B Informed Consent Form

Experiment form for bachelors thesis Christiaan Stierman.

Informed consent form:

I state that I am over 18 years of age and I wish to participate in the evaluation study being conducted by Christiaan Stierman for his thesis at the University of Amsterdam (UvA).

The purpose of this study is to evaluate types of sound as representations of two dimensional menu structures.

The procedure will involve monitored use of software designed to evaluate sounds as representations of two dimensional menu structures and filling out a small questionnaire.

All information collected in the study is confidential, and my name will not be identified at any time.

I understand that I am free to ask questions or to withdraw from participation at any time without penalty.

Signature participant: Date:

_____ _____

C Questionnaire

Questionnaire:

How many learning phases did you go through before achieving a 100% accuracy?

On a five point Likert scale:	Not at all				Very much so
Did you feel like learning the sounds was difficult?	1	2	3	4	5
Did you feel like picking the right items was hard?	1	2	3	4	5
Did you have trouble hearing the sounds?	1	2	3	4	5
How good would you describe your own hearing?	1	2	3	4	5

Other questions:

How many years of age are you?

What is your gender? *male / female*