EXPERIMENTAL COMPARISON OF COMPLEX AND SIMPLE SOUNDS IN MENU AND HIERARCHY SONIFICATION

Juha Marila

Nokia Research Center,
Visual Communications Laboratory
PO Box 407, FIN-00045 Nokia Group, Finland
juha.marila@nokia.com

ABSTRACT

Sounds can provide user with either new (functional) or redundant information about events and actions in electronic device user interfaces. In this study, we developed one simple and another complex sound scheme for menu browsing in a mobile phone. The complex sounds were hypothesized to enhance users' performance, compared with silence, but they did not. Instead, the simple scheme enhanced users performance. The results are discussed in relation to why and how informativeness and semantic character of sounds would affect user performance.

1. INTRODUCTION

Sonifying menu hierarchies has been realized with different methods in previous decade. The idea is not very familiar to general public and hasn't yet made a break-through in consumer IT. The usefulness of menu sonification has been demonstrated by e.g. Brewster & al., in avoiding and correcting selection errors [1, 2]. The implementation of earcon sounds has had positive effect in form of shortened performance times and decreased errors, compared with silence. Also, users have been satisfied with the implementation.

The earcon method of sonification has been tested in desktop and telephone line environment, but has not taken off with mobile devices. Possible reasons for this are investigated in Helle & al. study on earcon implementation (called *Menu Tones*) to mobile phones [3].

When sounds are used together with visual displays, they are often redundant. This means that the information they convey is already available on the visual display. This redundancy has been studied by e.g. Lemmens & al. [4] and Bussemakers & de Haan [5]. The aforementioned authors tested the difference between differently functioning sounds: abstract sounds (earcons) and categorically congruent (i.e. “natural”) sounds (auditory icons). The difference between the sound types is thus semantic, more than related to sound characteristics or attributes.

In aforementioned studies, it was noticed that abstract sounds increase response times whereas categorically congruent (natural) sounds decrease response times, compared to silence, in categorization tasks on visual objects [5]. These tasks have been basically recognition reaction time tasks, where subject is to recognize and respond to different types of (visual) stimuli. The auditory stimuli in these tests have been viewed as either distracting or facilitating this visual recognition and response process.

The Lemmens & al.[4] definition of redundant sounds contains, that the sounds have no additional information to the visual feedback. Since they offer no new information, they can only facilitate or distract a computer user's performance upon the visual domain tasks. *Functional* sounds, in contrast, would offer the user some information to act upon that would not appear in visual domain. But, redundancy and functionalness in real usage settings are not so clear-cut as in laboratory experiments. If there is only one visual object for user to attend to, abstract redundant sounds will increase response time. But, what if the visual display has two or three streams of information to concentrate at? In such situation, user's attention is divided between the several visual streams – only a certain percentage can be allocated to each visual stream. The probability of missing some information thus increases. Here sounds, although initially redundant to display information, would still utilize their 'attention-grabbing' property and provide functional information to the user. Thus, an initially redundant sound could speed up reaction time, if the visual feedback would otherwise be noticed only after a long delay; or prevent an erroneous action, if the visual feedback otherwise could be missed altogether.

In this sense, redundancy vs. functionalness becomes a probability issue. Sounds’ informativeness is tied to perceiver's attention: sound accompanied by a visual event has low probability of being functional if user's attention is fully on the visual event, and high probability of being functional if user's attention is only infrequently on the visual event. The idea is visualized in figure 1 below, which describes the *p* of redundancy vs. functionalness of auditory feedback in a situation where congruent auditory and visual feedback occur.

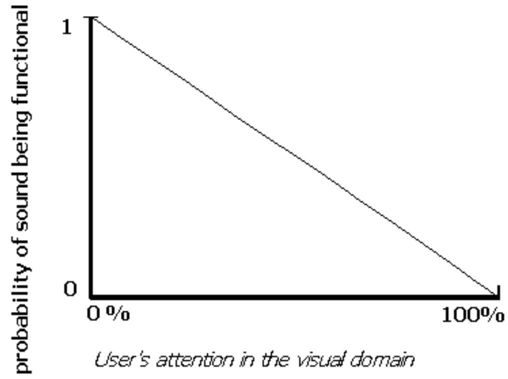

Figure 1. *Probability of functionalness for auditory feedback, based on user's attention.*

The graph is assumed linear, because we do not know about its possible non-linearity or situational variance of the functionalness. It could be drawn differently, based on empirical or theoretical investigations.

So, redundancy can be seen as the system property of sounds when a system is observed in isolation from tasks performed in time. But, when a human user, who in his/her actions and perception is subservient to linear time, is taken into the system, the informativeness attribute of sounds (functional vs. redundant) becomes dependent on human behavior.

1.1. Sonifying hierarchical menu in simplified ways

Mobile electronic devices, especially phones, have a small display area. This should fit well into visual focus of users, and thus diminish the need for auditory redundant information. Looking at the results from Helle & al. on complex earcon sonification for mobile phone menus, called *Menu Tones*[3], it is clear that long, complicated earcons would not be suitable in mobile phone usage. The reasons behind this are supposedly the social distraction and modest perceived utility. The problem that the Menu Tones sonification tried to address was that users tend to get lost while browsing hierarchical menus with limited visual feedback. After 3 weeks usage, however, only four users (out of 17) saw this useful [ibid., p. 260]. For this problem, we devised a simpler scheme that followed roughly the same principles of user guidance as the original Menu Tone earcons did.

The complex tones were dubbed *MenuTonesII* (*MTII)*. Their basic idea was similar to Brewster's auditory scrollbar test [2], with similar method for sonifying the steps, direction and changes of the menu. The basic method was to raise the pitch of the keypad tone by a semi-note when scrolling forward and lower it when scrolling backward. When user selected a sub-menu, the same scheme started over, providing the user with information about the direction and progress of scrolling, but in addition the tone from the previous menu level was also played in advance to the scrolling tone. This basically reminded the user all the time about the menu level s/he was at.

This method would enable us to indicate the whole depth of current menu hierarchy, by playing a sound for each level. However, we decided to avoid such lengthy sequences and limit the sounds to just two. So the tones did not increase when progressing deeper into the menu hierarchy – they always just consisted of the previous level tone + current item tone. The resulting sequences were supposed to be short and acceptable enough. The tones in MTII were 25 ms long, with 25 ms of silence between them, totaling 75 ms for two-tone sounds. Starting frequency was A4 (442 Hz).

Table 1 presents an example of the notes played in a random menu scrolling and selection task.

Going to set the **cell info display** ***property (under*** **Settings** ***menu)***	
Menu item	Tone sequence
Messages	A
Call Register	B
Profiles	H
Settings (*selected*)	C
- Alarm	C A
- Clock	C B
- Auto Date&Time	C H
- Call Settings	C C
- Phone Settings (*sel.*)	C C#
- - Language	C# A
- - Cell info Display (*sel.*)	C# B
- - - On	B A
- - - Off (*sel.*)	B B

Table 1. *Example of a menu item path and accompanying sonification with MTII.*

As an alternative to this, we also put into test a simple keypad audio enhancement, named *EndIndicator*. It produced a short (4 ms) tone of 1211 Hz when scrolling a menu back- or forward, and another 4 ms tone (1964 Hz) when the last item in the menu list was reached. Menus were looped, so in effect the EndInd tone warned the user that s/he had scrolled through the list already and thus probably not found the item s/he was looking for in there.

Reason for using different pitch in EndInd and MTII was, that the MTII pitch was too low to be audible with 4 ms duration, and EndInd pitch was too high for the highest MTII tones.

Both menu sonification schemes were designed for Nokia mobile phone menus. For test purposes, we modified the menu structure to have a roughly equal amount of items in the sub-menus used for the test.

For comparison, we also included the current default keypad tone in the test, but as a shortened version (originally 100 ms, here 25 ms, 900 Hz)

All the sounds were tried against silence, i.e. no sounds as feedback.

2. THE EXPERIMENT

Three sound schemes were thus compared with each other and silence – a within-subjects experiment design, thus. As test variables, the performance-related **scrolling time** and **key press frequency** were selected. Additionally, user's subjective preference and experienced workload were also measured with questionnaires.

Sounds were implemented in a PC simulation of a mobile phone. The simulation was operated with a phone that had its keypad connected to the PC. Visual feedback was provided on the PC display. Sounds were played to the user via a small mobile device speaker that was attached to the phone. The PC simulation software logged all key presses, menu items and the time between each key press in milliseconds.

Subjects were displayed a picture of the phone display at a certain menu item, after which they were supposed to find this item using the phone keypad and the simulation. Subjects were instructed to find the targets as quickly as they were able to conveniently do. If the subject after trying announced that s/he couldn't find the target, facilitator told him/her the correct path.

The test conditions were:

0) a short introduction and rehearsal of the equipment, with silence
1) 5 targets with silence
2) 10 targets with new (or old) sound
3) 10 targets with old (or new) sound
4) 5 targets with silence

The experiment was a mixed design: half of the subjects had MTII as the new sound, other half had EndInd. The old sound was the simulated current keypad tone used in Nokia phones. Order of presentation was rotated as well: half of the subjects received the new sound first; other half received the old sound first. Silent condition was used both first and last because learning effect could be expected to affect performance during the experiment. By having the silent condition before and after the sounds treatment, we could estimate the magnitude of the learning effect.

8 test subjects participated in the test. 50% of subjects were female. All subjects were of ages between 25 to 35 yrs. All had previous usage experience with several Nokia phone models.

2.1. Expected effects from sounds

The tones (or their components) can be classified based on their informativeness (redundant/functional) and semantics (abstract/categorical). Following table of classifications (Table 2) can be thus produced.

	Abstract information	
Tone	Redundant	Functional
MTII	- direction of scrolling - position in current menu level	- Current position in the menu hierarchy - crossing the end of list
EndInd	None	None
Default	None	None
	Categorical information	
Tone	Redundant	Functional
MTII	None	None
EndInd	Advancement of scrolling	Crossing the loop boundary
Default	Advancement of scrolling	None

Table 2. *Informational and semantic properties of the tones in test.*

As can be seen, EndInd and Default tones have been classified here as categorical. This does not imply that they are natural. But, what would be "natural" sound for a text-based menu item appearance? The fact that the sound does not generally change according to display information (as MTII does) but instead accompanies each key press makes these sounds closer to natural and non-abstract.

The visual feedback in the tasks consisted of the current menu item text, which was the primary target of attention, and a scrollbar on the right column of the display, which indicates moving down or up and displays the number of the current item in list. So, information on current list position and direction of movement were available on visual modality. Scrollbar thus also implies the list end crossing, but since subjects' attention was supposed to be on the display center (showing the item text), auditory information on crossing should have a good chance of being functional rather than redundant.

As for the general effect of sound vs. silence, we expected sounds to increase response times, since they are extra information for the user to be processed.

2.2. Test variables

Time between successive key presses of the 'Down' key and the average number of the key presses were used as the quantitative measures of performance.

Time indicated the performance flow: if the current phone display was not the target, then subject scrolled to the next display item. Most often the scrolling direction was down – approximately 60% of the key presses, as post hoc analysis revealed - hence we selected the 'Down' key press as the observed action. The smaller the time between successive key presses, the smaller the distraction caused by the sounds.

Likewise, the amount of 'Down' key presses indicated the subject's awareness of focus in the hierarchy. The more key presses, the less the subject was supposedly aware of whether s/he had scrolled through the whole hierarchy level already (without finding the target). The smaller the number of key presses, the better the users were able to maintain their mental focus.

Preference for the new sounds ('*would like to have these sounds in own mobile phone*') was assessed after the new sound condition (2 or 3). After both new and old sound condition (2 and 3), distractiveness and irritativeness of the sounds were queried. After each test condition (1 to 4), the experienced ease and effort of finding the targets and the frustration related to the tasks were assessed.

2.3. Hypotheses

1) MTII sounds would both decrease key press frequency and increase key press time, from silent condition. This is expected because MTII tones are completely abstract.
2) EndInd sounds would decrease key press frequency from silent condition and increase key press time, but less than MTII. This, because EndInd tones have a categorical ("natural") characteristics.
3) Default sounds would increase key press time from silent condition and not have effect on key press frequency.

2.4. Results

2.4.1. Key press time

Between conditions, the time between two successive key presses was shortest for EndInd tones and longest for MTII tones. Silent condition and the Default tones produced the second and third longest response times, respectively. 95%

confidence intervals indicate that the difference is likely to be real and not chance variance. Only the last silent condition was used for comparison, for reasons stated later. See Figure 2 below.

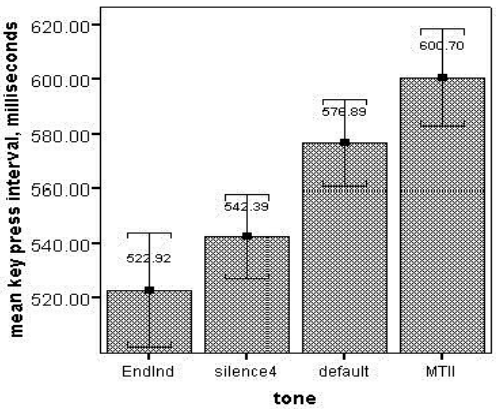

Figure 2. *Mean response times per tone condition.*

2.4.2. *Key press frequency*

Between conditions, no significant differences were detected in key press frequencies. Table 2 below describes each condition's mean percentage of 'Down' key presses in all key presses.

tone	Down, %
Default	59
EndInd	56
MTII	59
silence	62

Table 3. *Percentage of 'Down' key presses of all key presses between conditions.*

Since no difference was shown in 'Down' key press ratios, the possible difference between the key press frequencies in each sound condition was compared to the average key press frequency across all sound conditions. This reveals that MTII tones had 1.16 times the frequency of *Down* key presses in the sound conditions on average. EndInd tones had 0.81 and Default tones had 0.94. So, with MTII tones, subjects used more key presses on the tasks than with the other two sounds.

2.4.3. *Subjective preference and task load*

Questions on the *ease and effort of finding targets* gained positive responses from half or less than half of the subjects. Likewise, half or more of the subjects announced that the tasks were *frustrating*. Both findings apply to all tone and silent conditions.

None of the subjects replied MTII or Default tones *useful*, but 2 subjects found EndInd tones to be so. None of the subjects found any of the three tones *pleasant*. One subject would *prefer* to have EndInd tones *in her/his own phone*, but no one preferred MTII tones.

For Default tones, the preference question was not stated. Instead, subjects were asked at the end of test whether they usually have the keypad tones in their own phones off or on. 5 subjects reported to usually have the tones off, and the rest three said they are on but at the quietest sound volume level.

2.4.4. *Learning effect*

From first to last condition, it could be suspected that performance time enhanced by experience gained by the subjects. This actually happened: in first (silent) condition, the inter-key press time was 588 ms, 2nd and 3rd condition (with tones) the times were 570 and 573 ms, and on the last (silent) condition, 542 ms. The difference is significant by one-way ANOVA test ($F=5.084$, $p=0.002$). Since the time between both conditions with tones did not differ, learning may have counter-balanced some of the effect of tones.

2.5. Conclusions

Hypothesis on key press frequency changes did not get support from the results. Hypothesis on time increase with MTII and Default tones was confirmed, but not on EndInd sounds. EndInd produced faster inter-key press times than silence.

As for subjective preference, EndInd tones were just slightly better tolerated on the average than the rest of the tones.

3. DISCUSSION

Informational redundancy or functionalness can be considered attributes of sounds, independent from the abstract/categorical distinction. The tones in the test were all redundant and somewhat abstract. They should have affected the performance in similar ways: reducing key presses and increasing performance time. Instead, the abstract MTII tones increased both measures. Categorical EndInd tones *decreased* both measures. Categorical Default tones increased performance time.

3.1. Informational value and semantic content

One explanation for the results is, that MTII tones' information was for most part not useful to users and distracted them. What MTII tones offered and other tones not, was the direction of scrolling and menu hierarchy position. Now, if this information were useful to the user, it would be reflected in the key press ratios. Apparently it was not. This is probably because the information on direction of scrolling and the current menu level is redundant in *relation to user's own knowledge* on what s/he is doing and what selections s/he has made previously. This redundancy can't be expected to turn into functionalness with very high probability, since user's attention from his/her own actions is not very easily distracted.

As for indicating the end of list, *p* for MTII sounds being functional in relation to the visual information is assumedly none higher than in the EndInd scheme. Both tones facilitated the perception of advancement in scrolling and provided users with useful information on crossing the list end.

Since MTII tones contained information in very abstract form, and default and EndInd tones were less abstract, their better performance in reaction times could be related to this semantic difference. Lack of useful information combined with complexity and abstract nature explains MTII's negative effect on user's performance in the test, according to this viewpoint.

3.2. Other factors

Why, then, did not silent condition produce fastest scrolling times? This could be due to reaction time to different stimuli. It has been noted that reaction time to auditory stimuli is usually faster than to visual stimuli, on the range of 40 milliseconds [6]. So, if users reacted to the tones instead of the visual display when scrolling, this would explain the difference between EndInd tones and silence. But this effect did not apply to default tones, so the explanation seems dubious.

Another possibility that should be examined is the possible effect of tone length. MTII tones were longest (75 ms), Default tones were second longest (25 ms) and EndInd the shortest (4 ms). Although, based on this line of thought, silent condition should have had shortest performance times, there might be some process that slows down the subjects' performance with sounds of suitable duration.

4. APPENDIX

Examples of the tones used in the test are to be found on the proceedings CD-ROM.

5. REFERENCES

[1] S. A. Brewster and M. G. Crease, "Correcting menu usability problems with sound," *Behaviour & Information Technology*, 1999, vol. 18, no. 3, pp. 165-177

[2] S.A. Brewster, P.C. Wright and A.D.N. Edwards, "The Design and Evaluation of an Auditory Enhanced Scrollbar," in *Proc. CHI'94 Human Factors in Computing Systems*, Boston, USA, April 1994, pp. 173-179.

[3] S. Helle, G. Leplâtre, J. Marila and P. Laine, "Menu sonification in a mobile phone – a prototype study," in *Proceedings of the 2001 International Conference on Auditory Display (ICAD2001)*, Espoo, Finland, August 2001, pp. 255-260.

[4] P. M. C. Lemmens, M. P. Bussemakers and A. de Haan, "Effects of auditory icons and earcons on visual categorization: the bigger picture," in *Proceedings of the 2001 International Conference on Auditory Display (ICAD2001),* Espoo, Finland, August 2001, pp. 117-125

[5] M. P. Bussemakers and A. de Haan, "When it sounds like a duck and it looks like a dog… Auditory icons vs. earcons in multimedia environments," in *Proceedings of the 2000 International Conference on Auditory Display (ICAD2000),* Atlanta, USA, April 2000, pp. 117-125.

[6] M.S. Sanders and E.J. McCormick, *Human Factors in Engineering and Design*. New York: McGraw-Hill, 1993.