

DYNAMIC AUDITORY CUES FOR EVENT IMPORTANCE LEVEL

Jonna Häkkinen

Nokia Mobile Phones
Research and Technology Access
Elektroniikkatie 3, P.O.Box 50,
90571 Oulu, Finland
jonna.hakkila@nokia.com

Sami Ronkainen

Nokia Mobile Phones
Research and Technology Access
Elektroniikkatie 3, P.O.Box 50,
90571 Oulu, Finland
sami.ronkainen@nokia.com

ABSTRACT

Auditory output signals are currently used for various alarming functions. We introduce an approach, where a known auditory signal is manipulated by a factor, which indicates the importance level of the considered signal. The level of manipulation can be changed dynamically according to the required importance level. The independent variables used in signal manipulation were length, vibrato, high-frequency filtering, and reverberation. User tests show that auditory cues can cause changes in perceived importance level. The clearest results of increasing perceived importance were attained by decreasing the high-frequency filtering and increasing the frequency of the vibrato effect.

1. INTRODUCTION

The demand for effectiveness and intuitiveness of the user interface (UI) is growing with increasing complexity of applications. This creates a need to expand the conventional UIs towards more effective use of human modalities. It has for example been shown that adding auditory information to the visual interfaces improves the user's efficiency to take into account the forthcoming events and act on them [1].

With the expanding number of computers and other information appliances available for the user, the amount of multitasking is increasing. By multitasking, we mean that the user is not exclusively concentrating on the usage of one device, but rather monitors multiple devices simultaneously. Multitasking can also occur between the user and different things in the environment, e.g. when walking.

When the interaction is initiated by the device, the user's attention must be caught. Interaction can be initiated due to various kinds of time-, place- or communication-based events. The usage of sound in the UI provides one obvious method for catching attention. All events in a device are not equally important, and the user must only be alerted according to the level of importance. In this paper we present ways of modifying auditory signals according to different importance levels, allowing the user to decide whether to react or not. Also the difference between the importance of an event and its urgency is considered.

2. IMPORTANCE OF AN EVENT

Different events occurring in various devices obviously have many different levels of importance. For a fighter pilot, the information that an enemy missile is locked on his jet is very important. On a mobile phone, the information that message sending was successful is of much lesser importance.

The importance of an event can also change over time. For instance, the first warning of a low battery on a mobile phone is less important than a later signal, when the battery has almost completely run out and the user will soon be unavailable for phone calls.

The urgency of an event is not always the same as its importance. For instance, an incoming phone call from a friend can be more important than one coming from an unknown person. But if the call is to be answered at all, both calls are equally urgent - they demand immediate reaction. As another example, consider two emails arriving from different sources. Neither of them is urgent, requiring immediate action, but one coming from a friend (especially if one is expecting a message) is more important than one that e.g. contains an advertisement.

As a summary, the urgency of an event relates to the time during which user action is required. The importance relates to whether action is taken at all, or the event is simply ignored, or reacted upon at a later time.

2.1. Auditory presentation of urgency and importance

Traditional research on auditory warnings has been concentrating on studying varying levels of urgency, as for example in a previous study by Edworthy *et al* [2]. However, less research has been conducted on presenting event importance without necessarily creating a sensation of urgency.

Obviously the more alerting a sound is, the more important the event causing the sound will be perceived. However, as stated before, events can have varying importance levels but the same level of urgency. Furthermore, as has been pointed out also in previous studies, it is easy to startle people by using sounds that are too alerting [3]. Therefore it would be beneficial to find ways of presenting varying levels of importance, without affecting the alertness of the user.

There are also specific problems in utilizing previous research on urgency in the context of mobile devices. For instance, in a study by Haas *et al* [3] it has been found that the highest levels of

perceived urgency were related to sounds having a high frequency, a fast inter-pulse speed and a high level of loudness. Especially the utilization of loudness is difficult in alarms for mobile devices. The ambient sound level varies a lot in different situations. Therefore it is hard to decide the correct level of loudness for the alarm, in order to be audible and alerting enough, but not to be startling. The device can also often be buried in the user's pocket or a handbag, which makes measurement of the ambient sound level difficult, unless an external measuring device is available.

Yet another issue is the dynamically varying level of event importance. For instance, the battery low warnings get more important over time, but they still indicate essentially the same event: the user needs to charge the battery. So, good auditory warnings for the successive battery low events would incorporate something that would lead the user to understand that the event is basically the same, and something that would indicate an increasing level of importance. It has been shown, that the user is able to gain rapidly more than one piece of information from an auditory signal [4]. So, the level of importance could be coded into the event sound by manipulating it in a suitable way.

In the study by Edworthy *et al* [2], various parameters for creating an increasing perception of urgency were presented. For instance, different selections in melodic and rhythmic structure were found to affect the perceived urgency. However, if those structures change too much between sounds, there is a risk that two sounds are not recognized as indications of the same event with varying importance. Instead they can be recognized as two completely different sounds, resulting in confusion. Furthermore, the parameters presented are rather constraining for a sound designer.

In this research we studied ways of modifying a sound in such a manner that the importance level changes, but the sound still maintains its identity. Our approach was to find ways of manipulating a basic, fundamental sound in such ways that different levels of perceived importance are created. The design of the fundamental sound itself can then be left for a sound designer.

3. SOUND DESIGN

Altogether five different approaches for indicating a varying level of importance were utilized. In a previous study [5] there had been indications that length of a sound could be one affecting factor. From previous research e.g. by Edworthy *et al* [2] we adapted the idea of varying inter-pulse interval, but modified it so that varying speed of *vibrato* in a sound was used. Another affecting parameter pointed out, but not confirmed in the same study [2], is the weighting between high- and low-frequency components in a sound. In our test, we utilized low-pass filtering of sound as one modifying parameter.

Filtering of high-frequency components is also a known distance cue, utilized e.g. in virtual auditory systems [6]. The metaphor of distance was seen as a potential way of indicating importance of a sound. Often in real life situations, sounds emanating from far away are less crucial e.g. for one's personal health than sounds coming from nearby.

As another way of presenting the illusion of varying distance we used varying levels of reverberation. No 3D processing of sounds was utilized, though. This was decided because of real life implementation issues. 3D sound usually requires headphones or

multi-speaker systems, neither of which can always be assumed to be available. Hence, all sound processing was carried out using monophonic processing.

3.1. Tested sounds

Artificial sounds were used to avoid associations, which could affect classification of the auditory cues. The fundamental sounds were single notes. Two different timbres were used in the fundamental sounds. One was a square wave of 294Hz (D4), the other a synthetic sound resembling a pan flute ("PanPeople" on an Alesis QS6.1 synthesizer), also played at D4. Both sounds had rapid onset and offset times, but the pan flute sound was richer in harmonic content.

These fundamental sounds were manipulated using different parameters. The same processing was applied to both of the timbres. All processing was done on Sonic Foundry Sound Forge editing software.

In all tests, the RMS power of each compared sound was normalized (to -10dB in Sound Forge, utilizing "Use Equal Loudness Contour" setting which utilizes the Fletcher-Munson equal loudness curves), in order to prevent loudness from affecting the perceived importance.

Parameters to be used were sound length, vibrato speed, filtering of high-frequency components, and direct-to-reverberated sound ratio, both with and without utilizing filtering of high-frequency components in the reverberation algorithm.

3.1.1. Sound Length

The different sound lengths used in the test were 500ms, 1000ms and 1500ms.

3.1.2. Vibrato Speed

The amplitude of vibrato was kept constant (10% setting in Sound Forge). Vibrato frequencies used were 3Hz, 5Hz and 7Hz.

3.1.3. Filtering

Three levels of filtering were used: one with no filtering, another using a low-pass filter with a cutoff frequency at 2kHz and another using a cutoff frequency at 4kHz.

3.1.4. Reverberation

In Sound Forge, reverberation model "Cavernous Space" was used, with decay time of 2 seconds. Early reflections were not utilized, but only late reverberation. Reverberation was varied by modifying the direct-to-reverberant sound ratio in the sounds. Settings varied from a completely dry sound to a completely effected sound, with an in-between sound consisting of a mix of dry sound at 0dB (compared to the overall normalized sound level) and reverberated sound at -30dB.

To compensate for the slower onset time of the completely effected sound, 15ms onset and offset slopes were applied to all of the sounds where reverberation was the modified parameter.

A problem with reverberation is that the reverberation tail obviously increases the total length of the sound. The possible

effect was studied by examining the possible correlation between the effects of sound length, and reverberation amount to the perceived importance level of a sound, as will be shown later.

3.1.5. Reverberation, non-filtered

As a default, the reverberation algorithm used applies a low-pass filter that attenuates frequencies over 2kHz. In the test, we also used a setting where this attenuation of high-frequency components was turned off. This was done to reveal a possible difference in perceived importance between sounds utilizing reverberation with high-frequency filtering, and sounds without. If such a difference were to exist, then it could reveal that reverberation does not work well as a parameter for perceived importance. Instead, the possible differences in perceived importance could result from changes in the high-frequency filtering and sound length.

4. TESTS

Tests included listening to and evaluating the importance of presented sound series, and a short interview after that. Ten subjects (students from various fields, 6 male and 4 female, aged between 20 and 30) participated in the tests. The subjects were paid with two movie tickets each. All subjects reported normal hearing.

The test set-up included an IBM Thinkpad T21 laptop computer, where the test environment was running, and Sennheiser HD25 closed-back headphones. Testing occurred in a quiet room even though the headphone design reduced the amount of external noise.

The test subjects did not hear the fundamental sounds before the test. This was presumed to correlate well with real-life use of off-the-shelf devices where people usually do not rehearse listening to event sounds.

4.1. Sound series

The sounds were presented in pairs of three-sound series, labelled A and B in each pair. Inside a three-sound series, one sound parameter was modified in unidirectional manner over three steps, e.g. the length was constantly increasing in the three successive sounds. The same sounds were presented in reversed order in A and B series (fig. 1) so that if the A series consisted of sounds of increasing length, the B series consisted of sounds of decreasing length. The timbre used in a pair was always the same, i.e. the timbres were never compared against each other.

The subject was asked to listen to each AB pair one by one. Alternative sound series were allowed to be listened several times. Then the subject selected the series (A or B) of the pair, in which the sound importance was perceived to increase. The selection was binary: for each pair the subject selected either A or B. Then the subject was asked to add written comments on the choice before continuing to the next sound series pair. Subjects were instructed only to judge the importance of an event causing the sound – not e.g. its urgency, nor the aesthetic values of the sounds.

The test consisted of ten AB pairs. Five pairs had fundamental signal of square wave and five resembled a pan flute. The pairs

were arranged so that all sounds of one timbre were listened in the first half of the test, and the sounds of the other timbre in the second half.

The order of the listened sound pairs was chosen so, that the same parameter was not modified in two successive pairs. To avoid possible effects caused by presentation order, all the sound test pairs were presented in inverted order to the subjects 6-10. The presentation order (AB or BA) for all parameters was also inverted for different timbres.

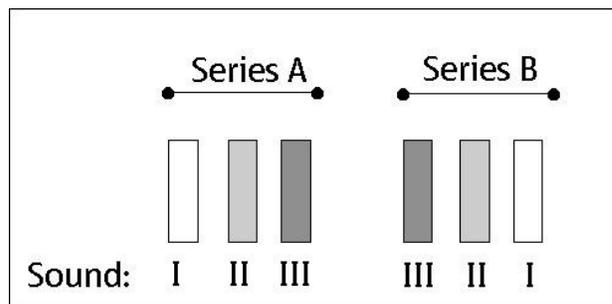


Figure 1. Order of sounds in test series.

4.2. Interview

After the test, subjects were requested to orally comment on the sounds, e.g. by asking what they remember from them.

Each subject was also asked to evaluate their musicality, and whether they played any instrument or sang. The musicality was quantified with values 0, 0.5 or 1, so that subjects who said they were not musical and did not play an instrument nor sing got 0. The ones who claimed to be musical, as well as those who admitted to either playing an instrument or singing, were considered musical and were given value 1. The rest were given value 0.5. The average value in the group was 0.65.

5. RESULTS

The distribution of the answers is presented in fig. 2. The bars in the figure represent averaged values of subjects' binary selections, so values near 1 or 0 mean consistent selections among subjects, while values near 0.5 mean that opinions were divided. Value 1 means agreement with our pre-assumed direction of increasing importance, while value 0 means agreement with an opposite direction. The pre-assumed directions were: decreasing amount of filtering (*i.e. brighter sounds*), decreasing amount of reverberation (*i.e. shorter, less reverberated sounds, both in filtered and non-filtered cases*), increasing length and increasing speed of vibrato. The order of presentation was not found to affect the answers.

In fig. 2, it can be seen that amount of filtering caused most results in agreement with our pre-assumed direction, while reverberation caused most disagreement.

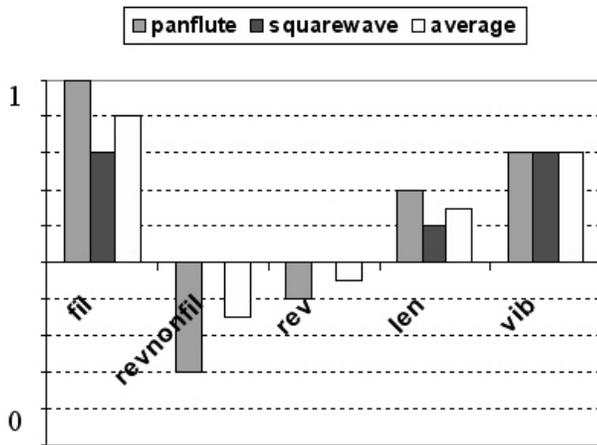


Figure 2. Averages of subject selections in the test. (Note: 1 and 0 refer to pre-assumed direction of increasing and decreasing order of importance, respectively).

It can be seen that subjects agreed more strongly in opinions concerning the pan flute timbre than the square wave. However, any parameter did not result in systematically opposite directions between different timbres.

As the selections made by test subjects were binary, the answer distribution was binomial and simple significance testing could be utilized. The average support (both timbres combined) for increasing importance was 0.9 for the least amount of filtering, and 0.8 for the fastest speed of vibrato. The findings are statistically significant (two-tailed test, $p < 0.01$ in both cases).

As is visible in fig. 2, different subjects gave contradictory answers to the effect of increasing length or increasing amount of reverberation to the perceived urgency. The average support for increasing importance was 0.65 for increasing sound length, and 0.35 for decreasing reverberation, neither of which are statistically significant results.

No significant correlation was found between the subjects' reported musicality and the importance level related to any of the sound parameters, or the consistency of their selections for the same parameter in sounds consisting of different timbres.

5.1. Inter-parameter effects, timbre and distance

As mentioned, it was known before the test that reverberation and length of a sound could have a correlation because reverberation tail increases the total length of a sound. Furthermore, also the filtering embedded in the reverberation algorithm was also seen to possibly correlate with filtering sounds without reverberation.

No significant correlation in the perceived importance between filtered sounds and reverberated sounds utilizing filtering in the reverberation algorithm were found for either of the timbres. This hints that filtering and reverberation could be treated separately. This is backed up also by that the directions for increasing importance were decreasing filtering (direction 1 in fig.2, i.e. brighter sounds) and increasing reverberation (direction 0 in fig.2, i.e. more muffled, longer sounds).

A positive correlation (because of the pre-selected 0 and 1 directions) between the test subjects' selections for increasing importance for different sound lengths and reverberation amounts

would have indicated that the subjects clearly treat length and reverberation independently. However, no clear positive or negative correlation was found. In written comments there were often mentions of changing amount of reverberation, and that the reason for the selection was the length of the reverberation. Some interdependency therefore seems to exist between the effect of the length and the amount of reverberation.

The answers related to the sound with reverberation effects were contradicting. Subjects disagreed, if greater reverberation meant decreasing or increasing level of importance. In addition, the subjects, who selected the same direction for both timbres with other effects, could choose opposite choices for reverberation sounds. The timbre difference is clearest in the case of reverberation with no filtering in the algorithm: average values were 0.2 and 0.5 for the pan flute and square wave sounds, respectively.

The original rationale for using reverberation and filtering was to arouse impressions of distance and thus cause an effect on the sound importance level. In written comments to test sounds, 3/10 subjects mentioned anything related to distance of the sound source in the reverberated sounds. In filtered sounds, 2/10 subjects mentioned something distance-related. The exact word "reverberation" was mentioned in 7 and 6 cases out of 10 with non-filtered reverberation effect for square wave and pan flute timbres, and 5 times for filtered reverberation sounds for both timbres. In the post-test interviews, reverberation was the most often mentioned (5/10 subjects) parameter that was commented by the users.

6. CONCLUSIONS AND FURTHER WORK

This study examined the effect of different manipulation factors on the perceived importance level of the sounds. The effects used on sounds were vibrato, length, reverberation with and without high-frequency filtering in the reverberation algorithm, and high-frequency filtering alone. The effects most clearly related to increasing importance level in the auditory signals, were the fastest speed of vibrato and the least amount of high-frequency filtering. The result of a sound with faster vibrato indicating a more important event agrees with the results of a previous study [2], in which a faster temporal inter-pulse pattern was found to be more alarming than a slower one. The same study suggested an effect of weighting the high-frequency components, which was also confirmed in our study.

The reverberation effect resulted in contradicting answers both among different subjects, and between timbres in answers of individual subjects. Many of the subjects' verbal comments suggested that the reverberation effect was recognized, but the overall sound length was still dominating the perception of importance. However, the effect of length on the perceived importance also varied among subjects.

The aim of the study was to examine the perceived importance of auditory signals, not the perceived urgency of them. However, as the test relied on individual selections and verbal comments, the findings cannot be guaranteed to be separate from those affecting urgency. In future work, the urgency effect of the most prominent parameters found in this study should be examined e.g. by using reaction time tests.

In this study, some subjects mentioned the notion of distance in the sounds, and some of the used sound parameters were also

selected based on a metaphor of distance. However, its actual effect was not studied. In the future, utilizing 3D-processed sounds or real 3D recordings, the effect of distance on the perceived importance should be studied.

Yet another issue is the lack of reference sounds in real usage situations. In our test, sounds were presented in series where the subjects could compare the changes. In real life, sounds are usually presented one at a time. In future research it should be studied if the appropriate importance information can be recognized from one individual sound alone.

ACKNOWLEDGMENTS

We would like to thank our colleague, Mr. Nicholas Martin for his help in formulating this paper.

7. REFERENCES

- [1] W.W. Gaver, R.B. Smith, T. O'Shea, "Effective Sounds in Complex Systems: The ARKOLA Simulation", in *Proc. of CHI'91*, Human Factors in Computing Systems, pp. 85-90. Addison Wesley, 1991.
- [2] J. Edworthy, S. Loxley, I. Dennis, "Improving Auditory Warning Design: Relationship between Warning Sound Parameters and Perceived Urgency", *Human Factors Vol 33(2)*, 1991, pp. 205-231.
- [3] E.C. Haas, J. Edworthy, "Auditory Warnings. Designing Urgency into Auditory Warnings Using Pitch, Speed and Loudness," *Computer & Control Engineering Journal*, pp. 193-198, Aug. 1996.
- [4] M.L. Brown, S.L. Newsome, E.P. Glinert, "An Experiment into the Use of Auditory Cues to Reduce Visual Workload", in *Proc. of CHI'89*, May 1989, pp. 339-346.
- [5] S. Ronkainen, "Earcons in Motion - Defining Language for an Intelligent Mobile Device" in *Proc. 7th Int. Conf. on Auditory Display (ICAD)*, Helsinki, July 2001, pp. 126-131.
- [6] D.R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press, 1994.