Synthesis of Alert Sounds

Jari Kleimola
HUT
jari.kleimola@fonet.fi

Abstract

This paper examines the most popular sound synthesis methods and their application to user attention grabbing, source identification and suggested action indication in auditory displays. Synthesis parameters are mapped to the parameters of alerting system via psychoacoustic sound properties, and an example mapping is used to demonstrate how these basic principles can enhance the usability of a mobile phone. Earcons are still considered more competent information conveyors, however. Investigation of implementation strategies in mobile environment show that the sonic palette of current mobile devices can be enriched with more advanced synthesis techniques.

1 INTRODUCTION

The popularity of mobile computing has increased rapidly, as mobile phones, PDAs, handhelds, portable music and video players, and other lifestyle devices are easy to carry along almost anywhere we go. Their small size comes with side effects in form of limited screen estate, which can make interaction with the device difficult. Mobile nature dictates also that our focus is not necessarily on the display, and therefore alerting sounds are often used to draw our attention to the device.

As more and more sophisticated applications are appearing to the market, and as the number of portable devices continues to grow, the auditory world surrounding us will become more crowded. Adding everyday background noises to this scenario makes it difficult to pick up alerts addressed to each individual. The brute-force solution to increase the loudness of alert sounds does not obviously solve the problem, so better attention grabbing mechanisms need to be employed.

Chapter 2 explains the basic concepts behind an alerting system and describes two paradigms that are convenient when modelling alerts in auditory user interfaces. Chapter 3 first analyses alert sounds of a simple mobile phone, then discusses the most suitable synthesis techniques, modulation routings and effects for generating such sounds, and presents a conceptual model that can be used to realise those synthesis methods. Chapter 4 lists psychoacoustic attributes of sound and discusses their ability to convey information, and via these attributes, provides an example mapping of synthesis parameters to the alert parameters that were presented in chapter 2. Chapter 5 discusses specific problems related to mobile environment, and investigates possible implementation strategies using current technologies. Finally, a conclusion is drawn in chapter 6.

2 AUDITORY ALERTS

2.1 Alerts

The aim of an alerting system is to notify the user that something is happening, to indicate what it is, and to suggest an action that the user should perform as a response.

Alerts are triggered by the system, and as such they must interrupt user's current foreground task. To grab his/her attention any multimodal interaction channel can be used. Focus of this paper is in the auditory channel.

The source of the alert and the suggested action are the main classifiers of alerts that the system is able to trigger (Ronkainen 2001), but they can also be further classified by their urgency and importance (Häkkilä et al, 2003), which in this study are combined as a priority concept. The source depends naturally on the mobile application, but the action classes are common across most applications. *Alarms* and *prompts* request the user to pick up the device and interact with it, while *warnings* and *notifications* are more subtle 'for your information' -type of messages not requiring further interaction. Priority is often encoded into the action class.

Mobile phones are examples of devices that generate alerts covering all classes discussed above. Table 1 lists system triggered alerts in a simple retro mobile phone, excluding incoming call and SMS notifications. Note the similarity between action class and Java's Alert() API call.

Table 1. Examples of alerts generated by a mobile phone.

Source	Action class	When triggered	Action
alarm clock	alarm	preset wake-up time has been reached	press confirm
reminder	alarm	preset date and time has been reached	read note and press confirm
timer	alarm	preset time lapse has elapsed	press confirm
hourly chime	notification	hour changed	nothing
application	notification	reminders cleared	nothing
battery (low)	warning	few minutes talk time left	nothing yet, but be prepared
battery (empty)	alarm	no more talk or standby time left	recharge battery
security	prompt	power up pin code query	tap in the pin code

2.2 Auditory Icons and Earcons

Auditory display has traditionally been modelled using two competing paradigms. *Auditory icons* are natural, everyday sounds which are used to represent actions and objects within an interface, while *earcons* use more abstract, synthetic tones in structured combinations to create auditory messages (Brewster 1994). The idea behind auditory icons is based on the fact that people listen to sound sources and their physical properties rather than the properties of the sound itself. The original concept of auditory icons was to use fixed recordings of real-world sounds, which Gaver later extended by suggesting parametrized auditory icons that are synthesized in real-time (Gaver 1993). These would allow creation of sonic families, and are certainly more flexible than pure playback samples. For synthesis, he proposes use of additive, modal and FM techniques.

Earcons are more musically inclined, and they must be learned (and remembered) before they can be associated with the sound source or a particular interaction task. They are serially constructed from smaller motives, which are short note sequences parametrized by rhythm, pitch, register, timbre and dynamics.

Brewster argues in his thesis, that the use of auditory icons is not always efficient because many times there is no natural sound that can be linked to a particular object or task in the user interface. Also, it is difficult to map a group of unrelated natural sounds to a single action class, and furthermore, a link between source object and associated alarm must also be learned, although previous knowledge certainly helps when memorizing the associations.

3 SYNTHESIS OF ALERT SOUNDS

3.1 Analysis of Alert Sounds of a Mobile Phone

This section analyses warning and alarm sounds of Nokia 3310 mobile phone, which is one of the older models (announced 2000), and is equipped with simple beep tones.

Figure 1 shows the earcon view of the tones. Alarm clock and timer alerts [A] consist of 10 beeps, each with 100 ms duration and a 90 ms pause in between. Reminder alert [B] is similar, but consists of only five beeps. All three of them are continuous, separated by a 570 ms rest, unlike the remaining four which are played only once. Application state notification [C] is a single beep with duration of 210 ms. Battery low warning [D] lasts for 220 ms, and it is repeated three times for battery empty alarm [E] with 90 ms gap in between. [F] is the familiar SMS indicator.

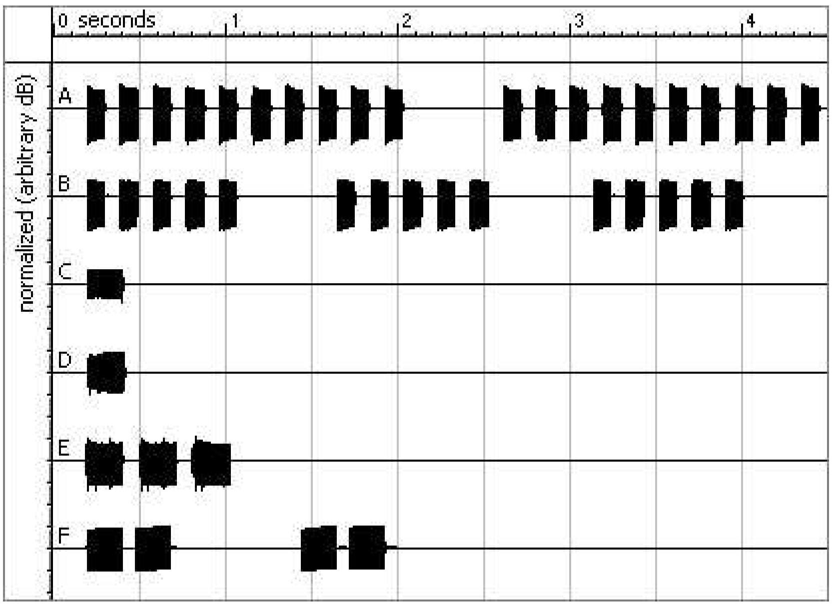

Figure 1. Temporal sequences of alert sounds in a mobile phone.

Figure 2 shows the waveforms and spectra of the alert tones. Alarm clock, timer and reminder tones [A] are identical, and consist of five harmonically related partials with fundamental frequency of 2681 Hz (e7, i.e. 2 octaves and a fifth above a440). Application state notification [B] is nearly sinusoid at 1750 Hz (a6). Battery signals [C] consist of two legato notes, first at 2792 Hz (f7) and then 2399 Hz (d7). The spectrum is taken from the longer (i.e. latter) tone, and the waveform plot is captured at transition time. The SMS tone [D] has all harmonics, and its fundamental frequency lies at 1499 Hz (f#6).

All tones have very fast onset and offset times, are spectrally harmonic, and fairly high pitched. There is no temporal modulation involved.

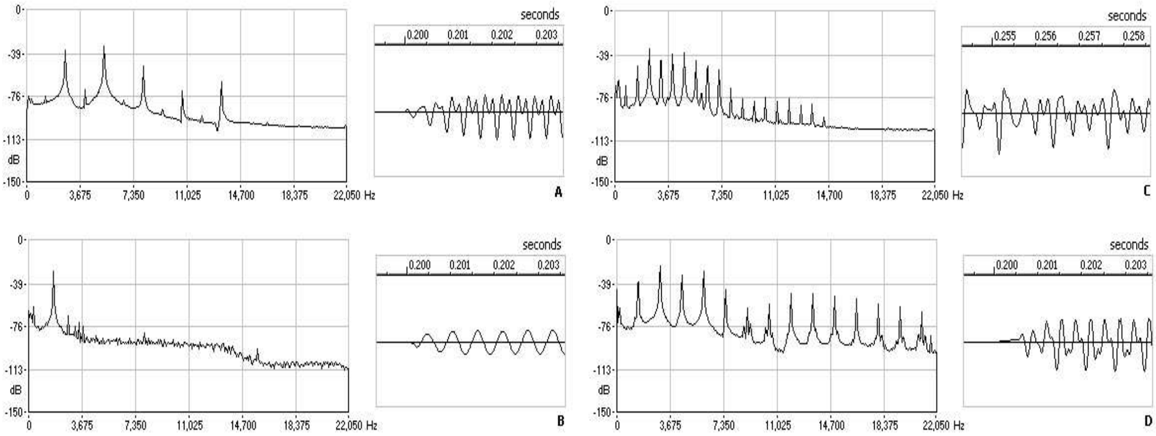

Figure 2. Spectra and waveforms of alert sounds in a mobile phone.

3.2 Overview of Synthesis Techniques

An extensive evaluation of modern synthesis techniques has been carried out by Tolonen, Välimäki and Karjalainen (Tolonen et al, 1998), based on taxonomy of methods proposed by Smith, and evaluation criteria after Jaffe. Smith groups digital synthesis techniques into four categories, which are abstract algorithms, sampling, spectral models and physical models. Jaffe proposes ten criteria for the evaluation of these techniques, addressing synthesis parameters, properties of produced sounds, and the implementation issues. A subset of the results of the evaluation work is represented next in the context of this paper.

Because latency should be minimized in order to achieve the best responsiveness of the auditory user interface, only those techniques that were rated as good in this sense were included, and those rated as poor in the computational cost and control stream density criteria were also discarded. Additive technique was included however, because of its generality, although in strict sense it would have been outruled by control stream density. The models for human voice synthesis were excluded because of their limited timbral variation possibilities. From evaluation criteria, perceptibility and behaviour of parameters, robustness and generality of produced sound and efficiency properties (i.e. computational cost, memory requirements and control stream density) were picked up. The results are summarized in table 2 below.

*Table 2. Evaluation of sound synthesis methods. Ranking is * poor, ** fair, *** good.*

	Parameters		Sound		Implementation		
	Percept	Behav	Robust	General	Comp	Memory	Control
Abstract							
FM	***	*	*	***	***	***	***
Waveshaping	*	**	-	**	***	***	***
KS	***	***	***	*	***	***	***
Sampling							
Sampling	-	-	**	***	***	*	***
Multiple-WT	-	-	-	***	**	*	***
Spectral							
Additive	*	***	-	**	**	*	*
Subtractive	**	**	**	***	**	**	**
Physical							
Modal	*	***	***	**	**	*	***
Waveguides	***	***	***	**	**	***	**

Evaluation of the implementation issues revealed, that the abstract algorithms are superior in the implementation category, while sampling techniques perform nearly as well (albeit with the cost of higher memory requirements). Waveguides seem to outperform modal, subtractive and additive techniques within implementation criteria.

The generality of produced sound reflects the timbre space produceable with the technique. Karplus-Strong algorithm simulates properties of a vibrating string, so even with the later extensions to the model, it is still for restricted use only. Physical models and waveshaping techniques are not as timbrally varied as other techniques. The robustness determines how well the sound identity persists when the synthesis parameters are modulated. Physical models and Karplus-Strong algorithm succeed very well, while FM might lose the focus entirely. Multiple wavetable, waveshaping and additive techniques were not rated for robustness, for reasons see (Tolonen et al, 1998).

Sampling techniques were not evaluated in the parameter category, because they are controlled by mere on/off gates and gain attributes. For the other groups, the behaviour criteria describes the 'linearity' of the parameters, i.e. a change in parameter should affect a proportional change in synthesized sound. Physical models and Karplus-Strong behave most adequately, and as expected, FM behaves in the most flexible manner. For strong parameters, the perceptuality rank is high, meaning that a significant change in parameter value is reflected with a significant change in the produced sound. Modal, waveshaping and additive methods have the weakest parameters, while the subtractive method is once again at par.

When looking at the results from mobile device's perspective, currently popular techniques (i.e. subtractive wavetable, sampling, and FM) are the most general methods, thus offering the widest timbral space. They perform also well in the implementation category. However, Karplus-Strong and waveguides do have an excellent overall score, and these might be considered as attractive additions to the sonic palette of current devices. This is also interesting because they may be useful in synthesis of parametrized auditory icons (Brewster 1994). Waveshaping, additive and modal techniques seem to be the most inappropriate of the evaluated methods.

3.3 Modulation and Effects

Dynamic waveforms can be produced by connecting a slowly floating control signal to any synthesis parameter. Most common modulation generators are the envelope generator (EG) and the low frequency oscillator (LFO), which can be patched in to generate vibrato, trill, tremolo, loudness contour, brightness sweeps and other time-varying enhancements. The usability of modulation depends on the synthesis technique, the number of parameters available, and on behaviour and strength of those parameters. In general, FM, waveshaping and subtractive methods are the most useful methods in this regard, although real-time modulation is possible with any of the techniques described above.

In dedicated synthesizers, raw synthesized sound is often still post-processed with effects, such as delay, reverb, equalization, compressor and 3D algorithms. Some of these effects are computationally complex tasks, but they are appearing also in modern mobile devices.

3.4 Hybrid Synthesis Model

Although synthesis methods discussed in 3.2 are basically quite different, they can conveniently coexist inside a single conceptual model, as realized in a hybrid synthesis engine prototype (Kleimola, 2005). A simplified model of a prototype voice is shown in figure 3.

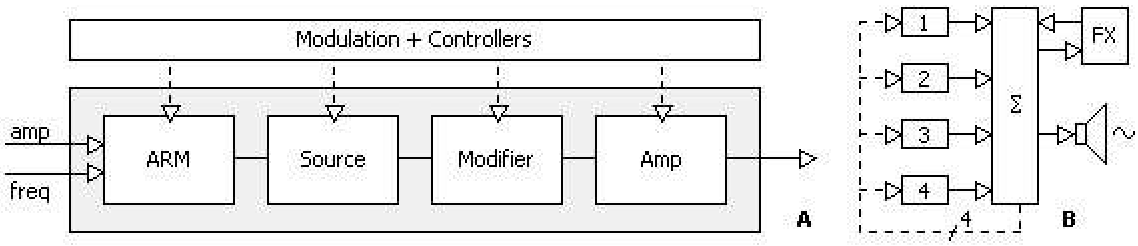

Figure 3. Hybrid synthesis model. A) component B) voice built out of four components. Dashed lines denote modulation.

There are four identical components in a voice, each of which has structure as shown in figure 3A. ARM block mixes audio rate modulation inputs (for FM) from other components. Source block provides the basic sound material (oscillator or waveguide), which is routed through Modifier (for filtering or waveshaping) into the Amplifier block. Control rate modulation sources are routed into appropriate patched blocks. Each component is connected to output mixer, which adds post-processing effects to the signal, as shown in figure 3B.

4 PARAMETER MAPPING

4.1 Sound Parameters

The basic attributes of sound are pitch, loudness and timbre. Although the audible *pitch* range is large, and the frequency resolution of human ear is about 5 cents (Rossing et al, 2001), pitch alone is a poor choice to encode information. The reason for this is that pitch needs a reference to another pitch in order to be recognizable. *Loudness* is another poor choice for the same reason, and moreover because it is dependent on the environment noises, so absolute measurement is impossible. The most adequate basic parameter to convey information is therefore *timbre*.

When real-time modulation of basic parameters is added, also pitch and loudness can be used to carry information, as these changes are relative in nature. Some tests have for example shown that fast vibrato or trill is perceived as more urgent than simple steady tone (Häkkilä et al, 2003). Guillaume et al (2002) performed tests on urgency mapping, and concluded that acoustic parameters are less important cues than mental representations, although temporal structure is important. According to their experiments, fast sequence rate, high pitch, random melody, irregular harmonics and fast onset were perceived as increased urgency.

Spatial location cues can also be used to convey information, in form of distance (reverberation and echoes, loudness and brightness), and as virtual 3D positioning. This requires headphones, as the speakers of mobile devices are often small, close together (if there are more than one in the first place) and of poor quality

Musical parameters of earcons are the most effective information containers, and these include duration, rhythm, tempo, melody, harmony and instrumentation. These fall outside the scope of this paper, however.

4.2 Mapping Synthesis Parameters to Sound Parameters

Table 3 presents a mapping scheme, which is derived by considering the primary parameters of the synthesis techniques and the basic attributes of sound. The subcolumn marked by ‘x’ denotes that the parameter is commonly used as a modulation destination. The table shows that sampling techniques used in auditory icons cannot be adequately parametrized in real-time, as even the loudness depends on the original sampled waveform. Only some methods are capable of parametrizing the timbre, which is the most important information encoding sound parameter.

Table 3. Primary parameter mapping.

Technique	Pitch		Loudness		Timbre	
FM	carrier freq	x	carrier amp	x	carrier freq / modulator freq, modulation index	x
Waveshaping	source freq	x	source amp	x	transfer function shape, distortion index	x
KS	delay length		initial excitation	x	loop filter, initial excitation	
Sampling	sample freq		sample amp	x	sample	
Multiple-WT	sample freq		sample amp	x	sample	
Additive	fund. freq	x	partial amp	x	partial amp and freq	x
Subtractive	source freq	x	source amp	x	filter cutoff freq, slope, mode	x
Modal	source freq	x	source amp	x	filterbank freqs	
Waveguides	delay length	x	excitation amp	x	depends on model	

4.3 Mapping Sound Parameters to Alert Parameters

Recalling the aim of an alerting system from chapter 2, the first task is to grab user's attention. This would work best with sounds that have fast attack time, like the ones observed for example in tuned percussion or plucked string instruments. To indicate the source of the alert, an obvious choice for the first level organizing parameter would be timbre, which can also be used to group related sources together into timbral families (e.g. marimba and vibraphone both belong to the tuned percussion instrument family, both quite easily synthesized using the FM method). Action and priority are combined in action class, which would probably best be identified with a short earcon motive, or alternatively using temporal modulation of pitch, loudness or timbre.

Many mobile devices allow at least part of the mapping to be configurable by the user. This can be done by connecting the alerting events to sound via a matrix that persists as a part of personal user profile or sound skin. Organizing the entire matrix as an addressable entity, one can then easily change the setup by switching to another profile or theme according to change in environment or mood. Personal profiles also minimize the confusion between alert sounds addressed to other individuals.

4.3.1 Example Sound Skin

The mapping presented in tables 4a and 4b can be used to define the sounds for the mobile phone alerting scenario described in chapter 2. The action class, i.e. whether the alert is an alarm, warning, notification or a prompt, is indicated by using a repetitive note sequence for alarms (user should pick up the device), and by a one-shot sequence for other actions ('by the way' -events). The urgency is further encoded into the tempo and brightness of the sound (faster tempo and higher cutoff frequency in filter means more urgent alert). All sounds have fast attack and decay times in general, although some artistic freedom can be allowed for this non-discriminating parameter.

Table 4a. Action and urgency cues.

Action class	Sequence	Notes
alarm	looping	fast
alarm clock	"	"
reminder	"	"
timer	"	"
battery (empty)	one-shot	"
warning	one-shot	slower
battery (low)	"	increasingly faster
prompt	one-shot	slower
appl query	2 notes	upwards motion
notification	one-shot	slow
hourly chime	2 notes	downwards motion
appl message	1 note	

Source indication is mapped into the timbre. Three source families can be identified from table 1, namely clock (alarm, reminder, timer, chime), battery (low, empty) and application alerts (prompt and message box -style notification). The first family is realized as sounds with metallic percussive transients, and the application alerts using more mellow wooden characteristics to avoid annoyance (user's focus is already on the device display). Both families can be efficiently synthesized using the FM technique, and they do contain inharmonic partials for easier recognition in noisy environments and improved attention-grabbing qualities. The sounds employ also a warm brassy or pipe-like body, which can be synthesized by subtractive or 2-operator FM methods. This is to enforce the sustaining weight, and to make the tones more pleasing to the ear. The sources inside a single family are differentiated by earcons.

Battery signals are identified by detuned and flanged timbres, with sweeping low-pass filtering that is quite intuitively associated with electricity. Mixed on top of that is a single pitched note sequence indicating the current charge level of the battery. Over time, multiple battery warnings are generated, and as the battery is losing its energy, the tempo and the length of the sequence is increased, indicating that the warning urgency level is becoming higher. The battery empty alarm is similar to the warnings, except that a slightly different fast tempo motive is used instead. A touch of ambience (i.e. post-processing delay/reverb effects) is added to all sounds to make them more appealing. The timbral map is shown in table 4b.

Table 4b. Source identification cues.

Family	**Transient**	**Body**	**Brightness**
clock			
alarm clock	metallic	breathy pipe	all (in)harmonics
reminder	metallic	breathy pipe	“
timer	metallic	breathy pipe	“
chime	metallic	breathy pipe	“
battery	synthetic	detuned fat	brighter each time invocated
application			
appl query	wooden perc	soft brass	less (in)harmonics
appl message	wooden synth	soft brass	few harmonics

The example sound skin presented above is relatively conservative, and time-variant modulation possibilities are only marginally exploited. Simple modification to the theme could be made by realizing plucked string and air jet timbres using physical modelling, or by replacing the earcons entirely by extreme control rate modulation. Nevertheless, the example theme is certainly less annoying than the analyzed scenario of chapter 3.1, while still maintaining a good perceptuality score even in noisy environments.

5 MOBILE ISSUES

5.1 Problems

Masking phenomenon emerges when the alert signal becomes inaudible in a noisy environment. This can happen if a disturbing signal is louder than the alert signal, and if both signals are close to each other either in time or frequency domain. Temporal masking occurs when louder sound follows the weaker sound within 10 ms time window, or when the weaker sound follows the louder one within 50 ms window. The reason behind temporal masking is that both sounds are associated with the louder source. Frequency domain masking occurs when two concurrent sounds share the same frequency band. Closely located spectrum components are perceived as a single component. It has also been found that masking amount is not linear function of sound level.

Limited hardware resources are another problematic area when dealing with mobile devices. Memory constraints and the lack of CPU and DSP power require highly optimized synthesis algorithms, can cause severe latency problems, and even make certain techniques impracticable. This problem is present especially in devices that do not have DSP as a mandatory component. Lack of floating point support poses further complications. Small portable size means also smaller speakers, which are often of quite poor quality. This should be taken into account when encoding alert parameters into the synthesis parameters.

5.2 Implementation Possibilities

By far, the most popular synthesis method used in mobile phones of today is the subtractive wavetable technique. Some models manage also FM synthesis using a proprietary DSP chip, and of course, there are still a lot of older models that use the simple buzzer sounds. Most modern devices are programmable however, and this section provides a brief overview of the possibilities that are available for realizing the synthesis techniques discussed in section 3.2 (see web links Sonify 2005, Sparks 2005, Boyce 2005).

5.2.1 Synthesis Engines and DSP chips

Commercial synthesis engines do not usually make their APIs open to general public, but if the engine claims to implement the MobileDLS standard, then it can be assumed that custom wavetables and samples can be appended to the repertoire that resides in the device ROM, and that customized patching is also available. This applies only to techniques in the sampling and spectral category. Examples of such synthesis engines are mobileBAE, SonicEAS, Mobileer and MicroQ (for more information, see the links in Sonify 2005). These synthesis engines employ also a Midi sequencer, which is a must when working with earcons. Some engines feature also audio effects.

A notable exception to the closed-API engines is TAO iSS (Tao 2003), which seems to have an exceptionally flexible API. All techniques discussed in section 3.2 can be implemented using its callback-based audio stream and plugin interfaces, but although it has been included in some mobile phones, it needs an operating system to stand upon.

On the hardware side, Yamaha's MA-series chips offer 4-operator FM in addition to the forementioned set of techniques. As they are co-holders of various waveguide patents, it wouldn't be surprising if some future chip design would add physical modelling to the feature set. The chip contains also a streaming hardware MIDI sequencer, relieving CPU load (but not substantially).

5.2.2 Java

MMAPI does allow freeform synthesis using a dynamically synthesized wav file structure, but as the entire sound must be precalculated before the playback can begin, the latency figures might render this approach impractical. More efficient methods are available in sampling and spectral categories using the synthesis engines via Java MMAPI. AMMS enables the use of effects and virtual 3D audio.

5.2.3 *Operating Systems*

Most freedom for synthesis implementation is available when using native OS APIs. They feature callback-based audio streaming interfaces that operate near hardware level, thus providing lower latencies.

6 CONCLUSION

The analysis of a mobile phone without auditory icons showed that purely synthetic tones do not communicate very well, unless short motives are used as a first level organizing parameter. Although spectra of the analyzed tones were different, it would require too much effort to notice the difference aurally. Earcons are much more powerful information conveyors, so their use even with sampled instrument waveforms is desirable. This would require a learning process, though.

Static properties of synthesized sound are thus inefficient information containers, so control rate modulation should be used for better discrimination. Using more advanced synthesis methods would also probably help, but the processing power of a mobile device might become the limiting factor, and should first be measured in a native environment. OS-based devices offer the most flexibility in implementation.

Of the synthesis methods, FM and waveguides would be most interesting candidates to augment capabilities of wavetable-based techniques. The candidates are considered as enhancements, because currently dominating subtractive wavetable technique is the most appropriate one for the task, despite of its shortcomings.

REFERENCES

Boyce, K. 2005. *Cellular handset audio evolution, Part 3*. National Semiconductor, 2005. Available online from Mobile Handset DesignLine, <URL:http://www.mobilehandsetdesignline.com/howto/showArticle.jhtml?articleId=170703641&pgno=1>

Brewster, S.A. 1994. *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*. Ph.D. Thesis, University of York, 1994. Available online from <URL:http://www.dcs.gla.ac.uk/~stephen/publications.shtml>.

Gaver, W.W. 1993. Synthesizing Auditory Icons. *Proc. of the SIGCHI conference on Human factors in computing systems (INTERCHI'93).* Amsterdam, The Netherlands, 1993. ACM Press, pp. 228-235.

Guillaume, A. et al 2001. Perception of Urgency and Alarm Design. *Proc. of the 2002 International Conference on Auditory Display.* Kyoto, Japan, July 2-5, 2002. Available online from <URL:http:// www.icad.org/websiteV2.0/Conferences/ICAD2002/proceedings/index.htm>.

Häkkilä, J. and Ronkainen, S. 2003. Dynamic Auditory Cues for Event Importance Level. *Proc. of the 2003 International Conference on Auditory Display.* Boston, MA, USA, July 6-9, 2003, pp.233-237. Available online from <URL:http://www.icad.org/websiteV2.0/Conferences/ICAD2003/frames.html>.

Kleimola, J. 2005. *Design and Implementation of a Software Sound Synthesizer*. Master's Thesis, Helsinki University of Technology, 2005. Available online from <URL:http://www.acoustics.hut.fi/publications/theses/kleimola_mst/>.

Ronkainen, S. 2001. Earcons in Motion - Defining Language for an Intelligent Mobile Device. *Proc. of the 2001 International Conference on Auditory Display*. Espoo, Finland, July 29-August 1, 2001, pp.126-131. Available online from <URL:http://www.acoustics.hut.fi/icad2001/proceedings/navig/toc.htm>.

Rossing, T.D., Moore, F.R. and Wheeler, P.A. 2001. *The Science of Sound*. 3rd edition, Addison Wesley, 2001. 680 p.

Sonify.org. 2005. *Interactive sound for the web & wireless tutorials*. <URL:http://sonify.org/tutorials>.

Sparks, D. 2005. *Integrating High-Quality Audio into Mobile Design*. Sonic Networks, 2005. Available online from Mobile Handset DesignLine, <URL:http:// www.mobilehandsetdesignline.com/howto/170703432>

Tao Group Ltd. 2003. *intent Sound System Documentation*. Available online from <URL:http:// www.sseyo.com/tao_group/ave/iss/api.html>

Tolonen, T., Välimäki, V. and Karjalainen, M. 1998. *Evaluation of Modern Sound Synthesis Methods*. Report 48, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, 1998. Available online from <URL:http://www.acoustics.hut.fi/publications/reports/sound_synth_report.html>.