Manne-Sakari Mustonen

INTRODUCING TIMBRE TO DESIGN OF SEMI-ABSTRACT EARCONS

Master's thesis in
Information System Science
03/07/2007

University of Jyväskylä
Department of Computer Science and Information Systems
Jyväskylä

ABSTRACT

Mustonen, Manne-Sakari

Introducing timbre to design of semi-abstract earcons
Jyväskylä: University of Jyväskylä, 2006
78 pages
Master's Thesis

Interface sounds have a relatively modest role in present day human computer interaction. The visual display is considered the most important part of the interface. Several studies of non-speech sounds have demonstrated encouraging results of how effective sounds could be. Despite these results, commercial interface design has not adapted sound usage. This maybe be partly due to the fact that some aspects of earlier studies have not been well thought-out, and the sounds designed with the help of earlier guidelines may not be functional.

This content analytic literature study focuses on studying the earlier research of non-speech interface sounds, literature of the semiotics, audiovisual design, musicology and psychoacoustic.

The literature of semiotics shows that the division of the auditory signs of the interface as representative or abstract sounds is, to some extent, misleading. The intuitiveness of the auditory signs in interface has been excluded. If intuitiveness or connotation of the sound is not considered, it may lead to problems. If the auditory sign does not fit into the context, or to the meaning it conveys, it may be distracting and non-functional.

The problems introduced can be solved. The literature research of audiovisual design, musicology and psychoacoustics demonstrates that timbre may be a key dimension in improving usability of auditory signs, and careful consideration of the connotations and other effects of timbre can improve non-speech sound usage.

KEYWORDS; Auditory Display, Auditory Icons, Earcon, Semiotics, Timbre

TIIVISTELMÄ

Mustonen, Manne-Sakari
Introducing timbre to design of semi-abstract earcons
Jyväskylä, Jyväskylän yliopisto
78 sivua
Pro gradu- tutkielma

Käyttöliittymä-äänet ovat nykyisessä tietokoneen ja ihmisen vuorovaikutuksessa vaatimattomassa roolissa. Visuaalista käyttöliittymää pidetään yleensä tärkeimpänä osana käyttöliittymää. Useat tutkimukset ovat kuitenkin osoittaneet että ääni voisi olla hyvin tehokas keino välittämään informaatiota käyttöliittymässä. Ääntä ei ole kuitenkaan otettu aktiiviseen kehitykseen kaupallisissa sovelluksissa. Osasyynä tähän voi olla se että vaikka aikaisemmat tutkimukset ovat osoittaneet rohkaisevia tuloksia, kaikkia ääneen liittyviä asioita ei ole otettu huomioon tutkimuksissa, ja näin olleen aikaisemmilla ohjeilla ei välttämättä synny toimivia käyttöliittymä-ääniä.

Tässä sisältöanalyyttisessä kirjallisuuskatsauksessa käsitellään tutkimuksia aikaisemmista ei-kielellisistä käyttöliittymä-äänistä, semiotiikasta, audiovisuaalisesta suunnittelusta, musiikkitieteestä sekä psykoakustiikasta. Semiotiikan kirjallisuus osoittaa että nykypäiväinen käyttöliittymien äänien jako esittäviin ja abstrakteihin ääniin on jossain määrin harhaanjohtava. Äänen intuitiivisuus on jätetty kokonaan huomiotta. Kun intuitiivisuutta ja konnotaatiota äänestä ei oteta huomioon, ääni ei välttämättä sovi kontekstiin eikä sovi yhteen kantamansa viestin kanssa, joka voi jota siihen että ääni voi olla jopa häiritsevä eikä ollenkaan toimiva.

Kuvatut ongelmat voidaan kuitenkin ratkaista. Kirjallisuuskatsaus audiovisuaaliseen suunnitteluun, musiikkitieteeseen sekä psykoakustiikkaan osoittaa että äänen väri voi olla mahdollisesti avainasemassa ratkaisemaan käyttöliittymä-äänen käytettävyysongelmat.

Avainsanat: Earcon, Äänikäyttöliittymä, Äänenväri, Ääni-ikoni, Semiotiikka

CONTENTS

1 INTRODUCTION

1. 1 Description of research field

In human computer interaction, concerning for example ordinary home PC use, the two main sensory modalities are vision and hearing whereas haptic sense is usually left unconcerned in design process. The hearing sense has some benefits compared to vision. The hearing sense is always in use, and it cannot be closed, as the eyes can be shut or directed elsewhere. In addition, range of hearing is much more comprehensive than vision. Range of vision is only 130 degrees high and 180 degrees wide cone, which is almost unusable outside inner cone of 30 degrees. Accurate and usable range of vision covers only five degrees' cone, which we use to reading etc.(Hyrsykari 2001) Hearing sense covers 360 degrees, all directions, which means: one does not have to direct ears to certain direction for sound localization. Auditory system is extremely accurate on perceiving the location of the sound source. (Warren 1999, 29-55)

The accurate range of vision can be focused on one point and direction at time, whereas hearing sense can observe multiple sound sources at time and focus on one – even when there exists plenty of other noise and sound sources, for example the cocktail party -effect, which refers to ability to hear one speaker in a loud crowd. When we are in a public place, there are several speakers in our surrounding and despite of them talking simultaneously we can still hear what our friend is saying.

Hearing sense can distinguish different sources effectively and recognize sources accurately; even if sounds are very similar to each other, such as two violins in concert, they are perceived as two different instruments and can be picked out separately. (Bregman 1990, Haykin & Chen 2005, Iverson 1995, Yost 1997) This ability of parallel listening and observing multiple sound sources simultaneously can be counted as a considerable benefit compared to vision, which is a very limited sense in this perspective. If all the interface's events are represented only on screen, some important events may remain unnoticed if eyes are directed somewhere else.

Sound conveys information effectively. For example from everyday life acoustical environment, also known as *soundscape*, we can infer from the sound from the mailbox if we receive one small letter, or a colossal book. Another example of information deduced from soundscape is different sounds of motors, computers and other devices. From these sounds, or more like from the changes in these sounds, we can hear if something is working correctly or not. Even the smallest changes can be effective in conveying plenty of information of our surrounding environment (Norman 1988, Schafer 1994, 15-29.) Human auditory system is accurate on perceiving even small changes in sound dimensions, which means loudness (volume, intensity) (Moore 1997, 63–88), pitch (Moore 1997, 177–212) and timbre (Moore 1997, 245–249).

Sounds are playing a more important role in human computer interaction than before, for example in the case of small mobile devices with small screens. Due to lack of screen space, it is necessary to convey information by other means than on screen. Small screen is not the only reason to use sounds as a part of interface. Information can be hidden also because of the complexity and size of a system, for example the monitoring room of a big power plant or the cockpit of an airplane. Not all the information can be displayed on screen, or it would become too difficult to observe all meters and screens. If all the information is dependent on attention of vision, there is a risk that fatal errors can occur and remain unnoticeable. In addition, use of sound can enhance usability by the means of feedback. Even in a common high-end home computer the use of sounds could prevent some errors by giving accurate information of the input that has been received, or of the problem that has occurred (Preece, 247-252).

Familiar examples of interface sounds are PC environment's various error sounds, incoming email sounds, clicks to indicate successful pressing of buttons etc. Common interface sounds can be found also on mobile phones: SMS – message incoming, battery full/low, problems in sending SMS, and so on. Almost all devices have some interface sounds. GPS-devices used in cars have often speech for giving navigation information. Watches have sounds to indicate pressed button, alarms, even hours etc. Videogames have sounds to indicate events and successful performances, events and progress of the game and so on. Interface sounds, also called the *auditory display*, has a vast amount of existing and potential implementations. Yet, despite of several different sounds that can be encountered in interfaces, and compared to

visual interface research, sound design in interfaces is relatively underdeveloped area, and active development has remained outside conventional technological environments and remained almost only in the field of academic human computer interaction research. (Brewster, Wrigh & Edwards 1995, Murphy, Pirhonen, McAllister & Yu 2006, Pirhonen, Murphy, McAllister & Yu 2006).

In auditory display research, non-speech sounds are divided in two rough categories according to present-day definitions; sounds are either representational sounds imitating real world sounds, like door closing etc., or abstract, usually musical sounds. Representational sounds are called nowadays as *auditory icons*, and abstract sounds are called *earcons.* (Brewster, Wright & Edwards 2007) One example of auditory icons that imitate real life soundscape is Sonic Finder (Gaver 1989) where natural soundscape imitating sounds were used to describe events of application such as scraping sound when dragging items on the screen. Other example of using described means to convey information has been studied in ArKola bottling factory simulation by Gaver, Smith & O'Shea (1991). In simulation, events of the factory were sonificated with auditory icons, and observer was able to hear all the processes of factory simultaneously and point out changes in soundscape, meaning changes in processes somewhere in the simulated factory.

Earcons we encounter in everyday computer use are usually musical warning sounds and other indicators of events, which usually do not provide much other information than "something happened" and user has to look on the screen in order to find out what happened. Researchers suggest that symbolic non-speech sounds can, and should, be used in ways that are more effective. Problem is that there are no decent instructions or guidelines to sound designers. At present, sound designers use predominantly ad hoc – solutions. What sounds good to designer should suit everybody.

There are some previous studies to create guidelines for earcon design. (Blattner, Sumikawa, Greenberg 1986; McGookin and Brewster 2004; Brewster, Wright, Edwards 2007) Despite of these studies, design of earcons remains ad hoc based and not as effective as it might be. Additionally, sometimes even the earcons designed by these guidelines can still become irritating or not functional. Previous guidelines do not provide enough aid to avoid some flaws in interface sound design. Signs, which are incongruent with the connotation and

meaning they are intented to convey, may lead to distracting situation. (Bussemakers and de Haan 2000; Lemmens, Bussemakers, de Haan 2000) It is crucial to consider connotations of sounds and apply that information into design of intuitive earcons in order to improve sound design. Present day guidelines do not provide help for design semantics of sounds. Lack of intuitive recognition of sound meanings and dissonant interface sounds can lead to errors and frustration. If we continuously hear some sound when using computer, and do not know why it frequently keeps on occurring, it may become irritating and frustrating.

Present day auditory sign categorization into arbitrary earcons and iconic auditory signs is difficult and raises problems. Most of the visual signs we encounter in everyday life include forms of representational, iconic signs and forms of symbolic, abstract relation between the sign and the object it represents. Connotations are considered as important content of signs. Signs can be distinguished as purely abstract or purely representational only rarely. (Fiske 2000, 72) For example, icons in PC desktop usually consist of some representative parts as well as some iconic parts. When examining for example Microsoft Word Document icon, it is an image, which represents a written paper document, and at the same time a letter W indicating it to be a Word document. Image of document is representational, iconic, and it depends of the features of actual paper document. Letter W is symbolic; there is no reason why W is representing the syllable we know to be W. Neither there is any natural reason why letter W represents Word. We just have to learn that in order to know what it represents. Without knowledge of our language system and knowing the products of Microsoft, W remains unsolved wave drawing. With that kind of mixture of abstract and representational part in the icon, it comes more intuitive; even if someone sees the icon for the first time, he/she can probably infer from it what it represents, due to his/hers knowledge of what a real document looks like.

This intuitive feature of auditory signs has not been considered present day in auditory display research, but forcefully is indented to exclude the signs that are to some extent intuitive and have non-abstract features. It is crucial to discuss the issues of intuitive character of earcons further and apply the knowledge to design of earcons. In studies of music science *timbre* (also; colour of the sound, quality of sound, and identity of sound source) is considered as the main primitive dimension of music evoking meanings and emotions. Thereby it could be the key dimension also in developing the intuitiveness of interface sounds.

In latest guidelines of earcon design by Brewster et al. (2007) timbre is considered only to help perception and to distinguish different earcon families by selecting clearly different instrumental timbres from MIDI. An aspect not considered in these guidelines is *why to select some certain timbres*. Timbre allows us to recognize different auditory sources accurately. In contemporary music, musicology and auditory scene analysis, timbre plays a central role in contribution to emotions, separation, identification, recognition and meaning of music and sound. Different types of meanings can be conveyed by using different instrumental timbres: one instrument playing "angrily" can convey a message of angriness. (Haykin & Chen, 2005, Jensen, 1999a, Lucassen, 2006, Padova, Santoboni & Belardinelli 2005, Zagorski-Thomas, 2005) Timbre as quality and colour of the sound can improve the coherence of auditory display and all contexts where earcons are used. If compared with music science and auditory scene analysis studies, the importance of timbre is underestimated in human computer interaction research, and provided guidelines only scrape the surface.

1.2 Conceptual frameworks and key terms

In auditory display research, interface sounds used to convey information are divided into two main categories, speech sounds and non-speech sounds. Using speech in interfaces is widely studied area; speech has been studied both as input and output method. In addition, commercial applications have been released for example for people with visual impairment.(Yankelovich, Lewov & Marx 1995) Non-speech sounds were at first considered as *auditory icons*, which means representative, iconic sounds that imitate natural sounds, as in ArKola and Sonic Finder by William Gaver (1986, 1991).The definition of auditory icon excludes symbolic, abstract auditory signs. After the auditory icon, Blattner et al. (1986) introduced the term *earcon* which was defined as:"…*earcons, which are audio messages used in the user-computer interface to provide information and feedback to the user about computer entities.*" Three different types of earcons were defined: representational earcons, abstract earcons and semi-abstract earcon, which is a combination of abstract and representational.

Later auditory icon was roughly separated from the term earcon; earcons were described as symbolic, structured sounds, and definitions of the semi-representational and representational

earcons were excluded. In the first definition, representational earcon refers to same kind of auditory sign than present day auditory icon, abstract earcon being counterpart for present day earcon, and semi-abstract earcons do not belong to any definition of non-speech auditory signs. In the auditory display research field, earcons are now considered as purely abstract sounds that are mostly structured by the western tonal music syntax. Semi-abstract level is excluded from both of the definitions, according to contemporary definition of earcons as abstract, synthetic sounds that can be used in structured combinations to create sound messages to represent parts of interface. (Brewster, Wright & Edwards 1993, Brewster, Wright & Edwards 1995) Due to the problems described more accurately later in this thesis, the first definition of earcons (Blattner et al. 1986), which includes the representational and semi-abstract earcons, is used in this thesis.

Earcons and auditory icons are *auditory signs*. As discussed earlier, earcons and auditory icons are roughly divided by semiotic differentiation into arbitrary, symbolic *earcons*, and iconic, representational *auditory icons*. However, there are severe flaws in this description. Distinction between symbolic and iconic sounds is not simple; one sign can have both symbolic and iconic features at the same time. (Fiske 2000, 81-85) Earlier earcon studies do not stress this dilemma enough, and in order to create design principles, this semiotic approach has to be studied thoroughly. Another aspect not considered well enough is how earcons together with the context form a meaningful whole where sounds are not dissonant.

Audiovisual contract is a term introduced by Michel Chion (1994) and it refers to the relationship of sound and visual representation. Audiovisual contract refers originally to the cinematic audiovisual representation, but to some extent, problems and issues are similar in audio user interface and cinematic narration. If audio representation is somewhat incongruent with visual representation, it may lead to negative responses. Audiovisual contract indicates a similar meaningful whole such as the one considered by semiotic approach.

Psychoacoustics is a discipline that studies how human auditory system works and how brain processes different kinds of sound stimulus, distinguishes sound sources and changes in sound stream. Human auditory system is very complex and to understand basics of timbre processing systems in human auditory system, it is necessary to consider to some extent basic studies of timbre. Masking effect is a process where a sound’s audible threshold is raised by

other sound, which is played simultaneously. Usually it means that a sound is masked by another louder sound so that the masked sound is inaudible. (Moore 2001, 89-97) The knowledge of masking effect is crucial to understanding how it is necessary to design earcon to avoid it to be masked by other sounds and to identify key issues of timbre usage in earcons.

In everyday listening, we hear thousands of different sound sources, which are not all relevant nor noticed. *Auditory scene analysis,* or *auditory stream segregation,* is a process in human brain that separates whole soundscape automatically into different units to process them separately. Usually we are interested in one sound source at time and we can hear it separately and discard others without active consideration. (Yost 1997, Bregman 1990) For example, when one's friend is talking in the bus where there exists plenty of other sounds like other people or engine roaring, one can still hear what his/hers friend is saying. One can simultaneously ignore other streams, and still notice when somewhere else something sudden happens, which may need ones attention – for example tyre exploding or someone says the listener's name. Human perception of auditory streams is incredible accurate, for example, spectrum of high-pitched piano tune is very different from low-pitched piano, and almost similar to flute, but human auditory system still recognizes source instrument accurately.

A field of musicology studies the relation between music and emotions, and these studies are very relevant to issues concerning auditory display. In musicology the main interest is to study what is the importance of timbre in emotions and meanings conveyed as well as in connotations of sounds. As Padova et al. (2005) argues that decoding information that timbre includes, would give us information about the producer of the sound, his intentions and his emotional states. According to this, in this thesis I suggest that timbre may be the primary dimension in intuitive earcon design.

1.3 Research question and assumptions

Earlier earcon studies have not much considered connotation of sounds, and their effect on earcons. Present day earcon definition excludes all kind of intuitive meanings or representative significations of sounds called earcons. However, first definition of earcons includes semi-representational and presentational earcons. The first definition of earcon is used in this thesis, due to the possibility to improve usage by motivation of the sign. The

research aim in this thesis is to study *how research on timbre from the disciplines of psychoacoustics, musicology and auditory scene analysis can be applied to design of semi-abstract earcons?*

Underlying reason for this study is that present day earcon studies do not consider nearly all of the possibilities of intuitiveness in earcon design, and that has caused some problems. In addition, problems are raised by the strict earcon definition, which does not include earcons with intuitive meanings reflecting the motivation of the sign. Terminological problems have implied that earcons are abstract auditory signs, and auditory signs are representational real-life soundscape imitating sounds; neither of those terms includes semi-abstract, intuitive auditory signs. This is absurd, though almost all other signs we encounter in everyday life do have some elements of abstract signs and some elements of representational, iconic sign. Thus contemporary earcon definition does have recognizable flaws, which have yet to be resolved.

As the basis of this thesis, there is a hypothesis that the most important element in creating meanings in sound is timbre; which is the dimension of sound that is also referred as quality, identity and the colour of sound. There are strong intuitive mappings between some timbre characteristics and some emotions and meanings. If timbre is incongruent with the entity it represents, it may become disadvantageous. The assumption is that from the studies of musicology and psychoacoustics it is possible to derive some guidelines to improve earcon design and use.

Due to the multidimensional and complex nature of timbre, it is important to study extensively timbre dimensions, characteristics and the phenomena, which it is involved with. Because the definition of timbre also refers to quality and *colour* of the sound, timbre has strong influence on aesthetics and correspondence with the whole context. This thesis introduces some approaches for consideration.

1.4 Research methods and structure of this report

Research method of this study can be described as content analytical literature review. In the first chapter, I have introduced the auditory sign research field in general, the reasons for the audio usage in interfaces, and described shortly how it is done today. The second chapter

presents the review of earcon studies and the separation of representational auditory icons and abstract, structured earcons and discusses the problematic nature and earcons and the flaws of their definition.

Chapter three provides semiotic approach to earcons, which discusses the semiotic distinction of different types of signs and defines more accurately the flaws of the present day earcon definition and use, mainly considering the unusable definition of timbre, which has led to excluding connotations and intuitiveness of earcons. The semiotic approach chapter includes audiovisual relationship studies that consider the meaningful whole and dilemmas of signs and their context. In the chapter of semantic approach to earcons, I have defined more accurately the semiotic nature of earcons and auditory signs. Bearing that in mind, the chapter introduces and discusses the previous earcon design guidelines, and examines the imperfection of the guidelines in designing functional signs.

Chapter four displays a psychoacoustical approach to timbre, reviewing the principles of sound dimension called timbre, which is introduced as the key dimension in answering the previously described problems of earcons. The general nature and perception of timbre are examined. Some relevant and similar issues to psychoacoustics are discussed in chapter five, through auditory scene analysis, which reviews and introduces the nature of timbre in segmentation of the complex auditory stimuli into understandable units, where we can hear different sounds produced by different objects separately.

A different approach to timbre is introduced in the chapter six: timbre, emotions, and meaning. The relevant issues for sign creation from research on emotions and timbre are presented. This approach discusses the connotations and other important aspects of timbre perception.

Finally, in chapter seven, conclusions of discussed earcon issues and timbre relevance to design of intuitive earcons are drawn and proposed.

2 EARCONS

In research, interface sounds used to convey information are divided into two different main categories, speech and non-speech sounds. Using speech in interfaces is a widely studied area; speech has been studied as input and output method. (Yankelovich et al. 1995) Both aspects have been studied and commercial applications have been released for example for people with visual impairment. Non-speech sounds were at first introduced as *auditory icons* (Gaver 1986), which indicate *iconic sounds that imitate natural sounds*, such as door closing representing the closing of file or program, or scraping sound when dragging items on desktop. Later, after the auditory icon, the term *earcon* was introduced. It covers all non-speech interface sounds, including auditory icons, known as *representational earcons.* (Blattner et al. 1989) After the term was introduced, next definition of earcons excluded representational and semi-abstract earcons. This led to a confused auditory display field where no levels of intuition of auditory signs were taken into account and earcons were mostly studied as musical sounds, designed by western tonal syntax. (Hankinson & Edwards 1999, D'Incá & Mion 2006, Walker, Nance & Lindsay 2006, Brewster et al. 1995).

However, representational and abstract earcons (or auditory icons and earcons, depending on the terminology) are not always so easy to distinguish. As signs everywhere around us, there are hardly any purely abstract signs or purely representative imitating signs. This dilemma has not been considered in the auditory display field. This has led to some problems in the auditory display field when intuitive sounds do not match with any definition. Bearing in mind differences and difficulties in differentiating between symbolic and iconic sounds, one sound can have different signification levels, so definitions of earcons and auditory icons are to some extent insufficient.

2.1 Definition of auditory icon

Auditory icons are interface sounds that imitate real world, everyday soundscape sounds. For example a sample of closing a door or engine starting used in interface are auditory icons. Auditory icons represent directly some everyday soundscape event, object, or attribute of

object. In everyday soundscape, we can separate out for example large or small letter dropping in mailbox simply by listening to it. *Mapping* refers to the relationship between two things, in this case interface sound and its object (Norman 1988). Auditory icons have natural, direct mapping by physical analogies that leads to immediate understanding of event/object/attribute represented. Mappings used in auditory icons are somewhat strongly naturally mapped, which signifies that sign is strongly motivated. (Fiske 2000, 77) For example of this usually familiar mappings of certain sounds to some other object or its attribute; rising level refers to more and diminishing level refers to less, louder sound can represent a larger amount, and lower sound while drawing icons on desktop can refer to a larger (heavier) file.

In everyday soundscape, we get many auditory messages that tell us what occurred, or reveal information of attributes of an object. Soundscape tells us whether things are working properly, working poorly or not working at all. Soundscape tells us also if something occurs immediately or abruptly. Using auditory icons in interfaces relies on these mappings of everyday listening. (Norman 1988) For example of sound conveyed meanings, we can hear by the sound if a car engine is running properly or if there is something odd in the sound that may tell us that it is not working. As another example, we usually can hear some sounds from our computer hard drive writing, and we can infer from it that the drive is still writing to disc. When someone is dragging a luggage in airport, we can hear from the sound of wheels if the luggage is heavy and big, or light and small.

Examples of using these kinds of mappings and knowledge of sounds in auditory icons introduced above are William Gaver's ArKola and Sonic Finder. (Gaver et al. 1991, Gaver 1986, Gaver 1989) ArKola was a bottling factory simulation where auditory icons were used to simulate events of factory and to enhance collaborative work. Auditory icons represented simulated operating machines and occurring events; such as the sound of clanking bottles representing bottle dispenser. In addition, if action stopped, sound stopped. Process indicator sound also presented ongoing process speed by changing rate of sound: the faster the clanking, the faster was the process. Moreover, if some product was wasted, system indicted that by playing a sound; if bottles were wasted system played a sound of crashing glass.

Auditory signs provided the users a possibility to monitor processes simultaneously from a distant room. Users were able to adjust different machines to same rate by listening to sounds. This experiment proved that this kind of usage of representational auditory signs, auditory icons, enables effective complex process monitoring and maintenance (Gaver et al. 1991).

Then another example of auditory icons experiment: experimentally designed device called Sonic Finder was used in Macintosh environment to provide auditory icons for desktop use. Scraping sounds indicated dragging items and another sound implied when an item was dropped to trash (Gaver 1989). Those kinds of auditory signs proved to help users to avoid some common errors and mistakes such as dragging an item next to trash and not in to trash can.

2.2 Definition of earcon

Blattner et al. (1986) first introduced the term *earcon* as a term that included earlier described auditory icons:"…*earcons, which are audio messages used in the user-computer interface to provide information and feedback to the user about computer entities*." Thus earcons were describes as structural non-speech auditory counterparts for visual icons of interface. Blattner et al. define earcons as *representational*, *abstract*, and *semi-abstract* earcons, corresponding with the terminology of visual icons. The earcon term is a wordplay led from the icon term. There is some deficiency in this definition due to the differences between symbols and icons. Although Blattner et al pointed out that icon is a term that is widely used in interface design to mean symbols and icons, term *icon* may be to some extent misleading when signs covered are also symbolic.

Representational earcons are like representational visual icons, similar to auditory icons introduced by Gaver (1986). Representational visual icons are simple pictures of objects or operations, but not as representational as photographs. However, as Blattner et al. (1989) state out, representational icons and representational earcons do have limitations, because not all operations, events or entities used in interface have their pictorial or sonical representation. Moreover, some that have, may be hard to illustrate or sonificate. In that case, abstract representations are used. However, selecting between a representational and an abstract earcon is not that simple always, and unfortunately, it has been not discussed in earlier earcon

studies. Some applications or devices are not able to play complex sounds such as representational earcons might be. Even though there is possibility to map the meaning to some real life sound, it may not be possible due to the limitations of the device. In addition, the aesthetic issues may be the reason to choose between representational and abstract earcons.

Compared to the visual icons, abstract icons are combined of geometric marks and shapes to represent a specific entity that cannot be portrayed with representational icon. Semi-abstract icons are combined of the representational and abstract icons. Abstract earcons are combined by western music syntax, rhythms, melodies and instruments used in a similar way than geometric shapes used in visual abstract icon construction. (Blattner et al. 1989)

As Blattner et al (1989) describe abstract earcons: "*…structured sounds earcons which are defined as non-verbal audio messages that are used in the user-computer interface to provide information to the user about some computer object, operation or interaction*". Abstract earcons are constructed from motives, which are defined as "*a rhythmicized sequence of pitches. Rhythm and pitch are* fixed parameters *of motives, and timbre, register, and dynamics are the* variable parameters *of motives.*"

After Blattner et al. next attempt to define earcons was by Brewster et al. (1995). They defined earcons as abstract, synthetic sounds that can be used in structured combinations to create sound messages to represent parts of interface. Earcons are composed of motives, which are short, rhythmic sequences of pitches with variable intensity, timbre and register. Motives usually consist of notes that vary on rhythm, duration and dynamic contour. Earlier studies of earcons emphasize musical nature of earcons, and design is widely based on the western tonal music syntax (Hankinson & Edwards 1999, Murphy, Pirhonen, McAllister & Yu 2006, Pirhonen, Murphy, McAllister & Yu 2006).

Pirhonen et al. (2006) suggested that earcons should be defined as "*all the non-speech sounds that are not auditory icons*". In this definition earcon covers all the auditory signs (symbolic/arbitrary sounds) that are not speech nor try to imitate real world, everyday soundscape. This definition is different and to some extent more practical, due to the previous empirical testing and studies that consider earcons as sounds based on a western

tonal music syntax, excluding all the auditory signs that are not auditory icons, nor somehow intuitive or clearly not arbitrary. Examples of these non-abstract and non-representational signs that are excluded from the definition are speech intonation imitating *spearcons* introduced by Walker et al. (2006) and behavioural gestures imitating auditory signs.

Still, these contemporary earcon definitions have severe deficiencies. As an analogy to visual icons, this kind of definition would mean that all the signs used in visual interface would be basic geometrical shapes with primary colours, or photos of objects, never their combinations or mixtures. This earlier strict segmentation of auditory icons and earcons has lead to confusing terminological field and may be distracting design issues of earcons. Almost no other signs we encounter in our surrounding environment are purely symbolic, but they have some forms of both abstract and representational meanings (Fiske 2000).

This strong forced segmentation of earcons to purely arbitrary auditory signs may lead to non-functional sounds and at some cases to performance disturbing *auditory distracters* (Bussemakers & de Haan 2000, Paul M.C. Lemmens) when connotations and important motivations of sign are not considered. After Blattner et al (1989), the levels of coding meanings into auditory signs have been flattened. The latter definitions of earcons have been considered only at the Blattner et al. definition of the abstract earcon level. Due to the definition of several scholars after Blattner et al. (Brewster et al. 1995, Brewster et al. 2007, Bussemakers & De Haan 2000, Hankinson & Edwards 1999, McGookin & Brewster 2004, McGookin 2004), representational or semi-abstract earcons are not associated with earcons, and neither with auditory icons.

Although, it is suggested in this thesis, that auditory icons should be, again, merged to term earcon to form one term including all the auditory signs, in order to clarify design issues and avoid terminological tribulation. In that case, the definition of Blattner et al (1989) as earcons including *representational*, *abstract*, and *semi-abstract* earcons conforming the terminology of visual icons could be usable. In that definition, representational earcon is a current auditory icon, abstract and semi abstract earcons included in current earcon and intuitive level coding of meaning. This first, more extensive definition of earcons is used in this thesis.

2.2.1 Types of earcons

According to Blattner et al. (1989), abstract and semi-abstract earcons have four categories.

1. *One-element earcons* are simple earcons that are used to convey only one simple meaning. This type of earcons cannot be decomposed to create meanings that are more complex. Semantic analogy to this type would be words of language.

2. *Compound earcons* are combined by connecting one-element earcons in sequential order in order to create complex messages. As semantic analogy, this type would be sentences of language, consisting of words.

3. *Hierarchical earcons* are a type that is based on a certain grammar-like structure. This structure can be considered as a tree-like structure that starts from the top and inherits meanings from earcons above. (see: fig 1.)

4. *Transformational earcons* are also constructed from certain grammar. In this context, one dimension of sound represents one meaning. Therefore, by combining dimensions it is possible to create various complex meanings. Each auditory dimension can be modified to change meaning of an earcon. For example, frequency of sound can represent size of an object; rhythm can represent families of objects and so on. An example of this type of use, by McGookin (2004) suggests that low-pitched piano rhythm can be used to represent inexpensive rollercoaster ride in a theme park, and by changing instrumental timbre to a violin, meaning is changed to represent inexpensive water ride. In this earcon-type it is not necessary to learn each individual earcons but only different meanings of different auditory dimensions.

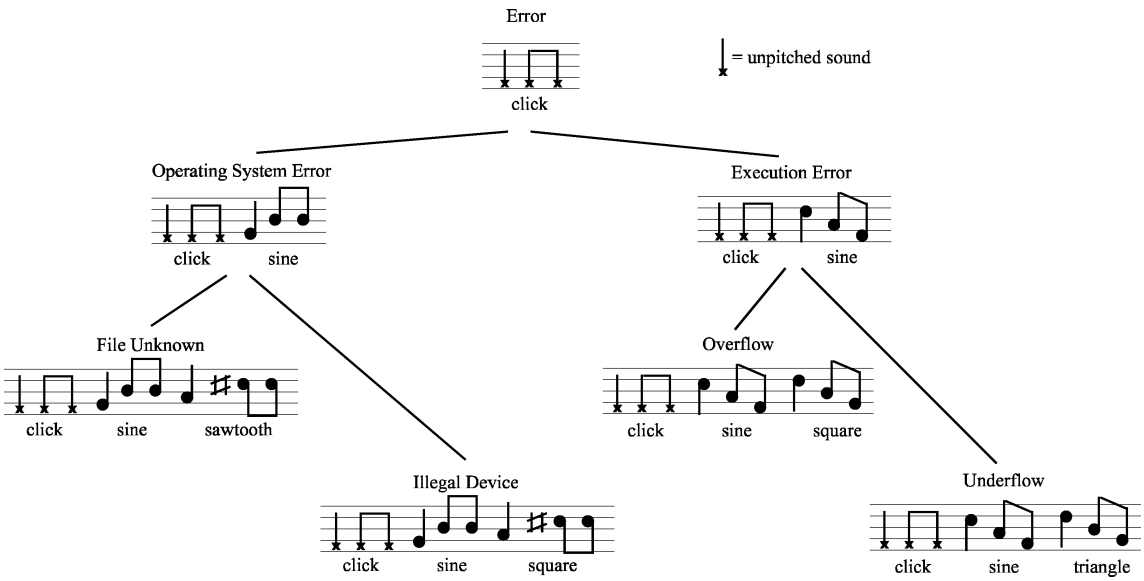

Fig 1. Structure of hierarchical icons, taken from (Blattner et al. 1989)

2.2.2 Evaluations of earcons

Earcons are an effective mean to convey information of the user interface (Brewster et al. 1993). There has been very little evaluation of earcons, but some has been previously introduced. Brewster criticizes earcons because they are slow to use and dependent on the time when an event happens (Brewster 1995). Whole earcon has to be played in order to convey meaning, and it needs to be detected at time it is played; there are no cues left after the sound occurred. If simultaneously occurs several events that require sound to be played, some of them may become too late for the event and therefore become to some extent useless.

There has been almost no evaluation of *usability* of earcons. Usability consists of five things. There are few aspects important to consider when evaluating systems usability; learnability and efficiency of use (intuitiveness), memorability, errors and subjective satisfaction. (Nielsen 1993) Few evaluations of earcon usability have been presented: Lemmens, et al (2000), and Bussemakers & de Haan (2000) introduced studies that considered the efficiency of use. Bussemakers & de Haan demonstrated in a study that earcons are normally evaluated in an environment where earcon is presented alone. In that case, there is no visual or other counterpart presented at the same time. If information is presented in multiple streams, in visual and auditory streams, user has to compose these streams into one experience.

Lemmens, et al (2000) conducted experiment of earcons played to participants when they pictured categorization tasks. Participants categorized line drawings of entities that were animals or non-animals. These two types of pictures were presented, and participants were supposed to say “yes” when they see an animal. Sound was played simultaneously. Sounds used were major or minor chords. Major and minor chords have strong connotation in western musical syntax; major chords are usually said to evoke happy mood, and minor are usually said to evoke sad mood. A hypothesis for the study was that minor chords presented were supposed to improve situations where subjects were expected to indicate that presented picture was non-animal, that is to say “no”, and major chords were supposed to improve situations where subjects were expected to indicate that picture presented is animal, that is, to say “yes”. Lemmens, et al. (2000) proved that sounds have no effect on performance speed, unless sound is somewhat incongruent. In that case, sound usage may hinder the categorization task. This study, however, studied earcons as short chords, and sounds used were chords in major or minor, which was the only altered dimension in sounds.

Brewster & McGookin (2004) demonstrated encouraging results. When simple audio feedback was played of successfully pressed button in PDA device, sound usage allowed size of the buttons to be increased. However, in this study, the connotations of sounds used were not taken into account, which can be counted as the severest problem in earcon studies.

2.3 Comparison of auditory icons and earcons

.

Regardless of the definition of earcons used in this thesis, in this chapter I will discuss the auditory icons and earcons as different terms, conforming to the present day definitions, in order to avoid misunderstandings. There has unfortunately been very little evaluation of earcons or comparison between auditory icons and abstract earcons. When comparing when to use earcons and when to use auditory icons, there has been only little discussion of that issue, but some examples can be found. For instance, McGookin & Brewster (2004) suggest that auditory icons can be used when it is possible to use an intuitive link between object and sound. This means that an object has a counterpart in the everyday soundscape. Abstract earcon can be used when there is no intuitive link between an object and sound to represent it (Brewster et al. 1993, Brewster, Wright & Edwards 1994, McGookin & Brewster 2004). This means that there is no everyday soundscape counterpart for the object or event to represent.

There is one severe weakness in this approach for choosing between earcons and auditory icons. Some mobile devices cannot produce all types of sounds due to the limits of sound reproduction means. For example, high-end wrist top computers used by athletes' need effective means to convey information without increasing size or battery consumption of device. In such devices, earcons may be better than more processing power and storage capacity demanding audio samples of everyday soundscape. Even if there is a clear indicator to use representational sounds, it may be impossible. This aspect has not been discussed in earlier earcon studies.

One benefit of earcons when compared to auditory icons is that they can be combined to create complex auditory messages. For example, earcons "image file" and "edit" can be combined to represent "edit image file" (Brewster 1995, Brewster et al. 1993). Even though it has been proved that earcons can be played simultaneously to some extent, without decreasing recognition rates, ability to recognize simultaneous earcons is limited. (Brewster 1995, Mcgookin & Brewster 2004) Auditory icons cannot be combined as e.g. hierarchical earcons, but auditory icon is more useful when it is necessary to combine different ongoing events simultaneously such as in ArKola bottling factory simulation (Gaver et al. 1991). Auditory icons have the benefit of conveying complex meanings in one single sound or soundscape; it is easy to recognize even at the first use and map them to meaning, and the meaning is usually unambiguous. However, auditory icons have been reported to be sometimes annoying to users after longer use.

Abstract earcons need to be learned before becoming useful, therefore they may not be suitable for first-time users or applications used infrequently. In addition, if the earcon is dissonant to the meaning it intends to convey, it may have the opposite effect that has been the intention. This can be inferred from the studies of Bussemakers et al as well as Lemmens et al. introduced earlier (Bussemakers & de Haan 2000, Lemmens et al. 2000).

However, it is important to consider that all these evaluations refer only to abstract earcons and representational earcons. None of them discusses semi-abstract earcons. Semi-abstract earcons may have benefits of the both. Semi-abstract earcons may be designed to be an intuitive, fast and effective way to use sound in interfaces.

3 SEMIOTIC APPROACH TO EARCONS

Earcons are signs, so it is necessary to describe their semantic nature to understand their usages and design principles. Earcons and auditory icons are now roughly divided by semiotic distinction to arbitrary, symbolic *earcons*, and iconic, representational auditory icons, and type where their combination, semi-abstract earcons is excluded. Separation to symbolic and iconic sounds is not simple; one sign can have symbolic and iconic features at the same time. In addition, it is important to notice the way the listener interprets current sounds played. If sound does not fit into the whole soundscape, or sound is incongruent and dissonant with other sounds, context and entities in interface, problems may occur.

Earlier earcon studies do not stress the semiotic analysis of auditory signs extensively. In studies where earcon design guidelines are under development, these aspects are usually not considered. This has lead to artificial discrimination of different auditory signs. In addition, in order to create design principles, this approach has to be studied accurately.

3.2 Communication process of auditory signs

There have been stated three problems in communication process (Fiske 2000, 19-21)

- 1. Technical problem; how accurately signs can be transferred.
- 2. Semantic problem; how accurately signs can convey wanted meanings.
- 3. Effectiveness problem; how effectively meaning interpreted of the sign guides the intrepretents to operate in a desired way.

Technical problems in auditory signs may be other distracters in the whole soundscape of the surrounding environment. If surrounding environment is noisy, it may make sounds of the interface inaudible. This is called a masking effect. Masking effect occurs easier on sounds that have narrow spectrum. Therefore, usage of timbres consisting of rich spectrum may help to avoid this problem (Brewster et al. 2007, Moore 2001, 89). Technical problem can be, for

example, the limitation of the device's sound reproduction. Small wrist top computers cannot play complex tones, perhaps only one sine wave. Not all devices have high fidelity speakers.

Perhaps the most important approach, which is hardly at all considered in the field of auditory display studies, is the semantic problem. In the context of earcon design, semantic problem can be introduced by the question: How non-speech and non-representational sound can convey meanings effectively? In order to answer that problem, the semiotic nature of auditory sign needs to be considered.

There are several different definitions of signs in the semiotic studies. The main structure of the different semiotic theories is mostly the same: it includes sign, an object wherein the sign refers to, and an interpreter. When speaking of the sign, it refers to some physical object which we can sense, such as visual sign, sound, or haptic sign. Sign refers to something else than itself so to understand sign's actual meaning, one has to understand the connection between sign and an object it refers to. The relation between the sign and its object has to be inferred by the users of interfaces

3.3 Semiotics approach to auditory signs

3.3.1 Peircean sign

In the field of semiotics, two models are the most influencing: C.S. Peirce's triadic model and Ferdinand De Saussure's dyadic model that differ from each other. Peirce's model considers sign to consist of three parts, which are the physical form of the sign, object of the sign and *intepretant* as a triangle (Fig 2). In this model, every point of the triangle is connected to each other. As Nathan Houser (1992, xxxvi) defines the Peirce's sign; "*The sign relation is fundamentally triadic: eliminate either the object or the interpretant, and you annihilate the sign.*" *Sign* refers to an *object* that is outside the object itself. For example, the sign can be a sound or a visual sign, and the object can be a computer file. Someone interprets the sign, so the sign affects to the mind of the person interpreting; that is, the mind that infers the sign is in this case called the *interpretant.* The concept of the interpretant is to some extent complex; Peirce has referred to it also as "*the proper significate effect*" (Fiske 2000, 64-71). The interpretant is a nonmaterial concept that is created by the sign and the user's experience of the object of the sign.

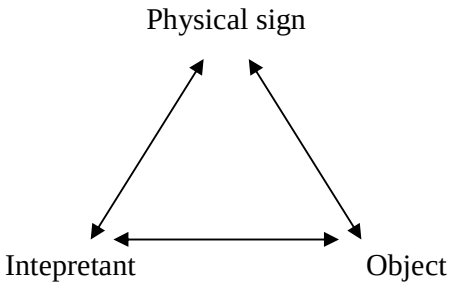

Fig 2. The sign triangle according to Peirce's model

According to Peirce, the relation between a sign and an object it is referring to can be iconic, indexical or symbolic. Iconic sign means that it resembles the object somehow: it looks, sounds or feels like the object. Indexical definition referes to a sign, which has direct mapping to its object, when there is a clear connection between a sign and the object it stands for; that is to say, it indicates the object. Representative earcons have iconic-indexical nature. Symbolic sign has no connection with the object. This means that the meaning of the sign has to be learned. Symbols are usually based on agreed code or language. This is also the nature of abstract earcons.

3.3.2 Saussurean sign

Ferdinand de Saussure's model considers signs from a different point of view. Being a linguist, he considered only relations of signs. Saussure did not consider the object of a sign as important as Peirce did. In the model of Saussure, the physical object was not considered at all. More important to his approach is the relation of a sign to other signs. The main emphasis is therefore on the sign itself.

In de Saussure's model sign is composed of *signifier*, which is the form that sign takes, or a mental image of the sign; and of *signified*, which is the concept it represents. Sign is a result of association between signified and signifier. Relationship between these two, signified and signifier is called signification, see figure 3. The signifier of de Saussure has been referred as the mental image of sign, but also as the physical form of sign. Some semioticians refer to the signified as the physical sign, such as letters on a paper or sound waves on an ear drum, for instance Fiske (2000, 66) and Cobbley & Jansz (1998). The signifier, as de Saussure defined,

(in French; *signifiant*) means the psychological, immaterial concept (Tarasti, 1996, 23). Saussure sees the sign more as a *form* rather than *substance* (Saussure 1983, 12).

In Saussures model signified denotes only the concept of signified, not any physical object itself. This means that one signifier can stand for multiple signified if a concept for mapping a signifier is different. Signifier "open" can stand for example for the fact that certain shop is open or for the suggestion to open a door, depending where the word "open" is seen. Sign is therefore seen as a vehicle for conception of the objects and for the concept, depending on where and how the sign is interpreted.

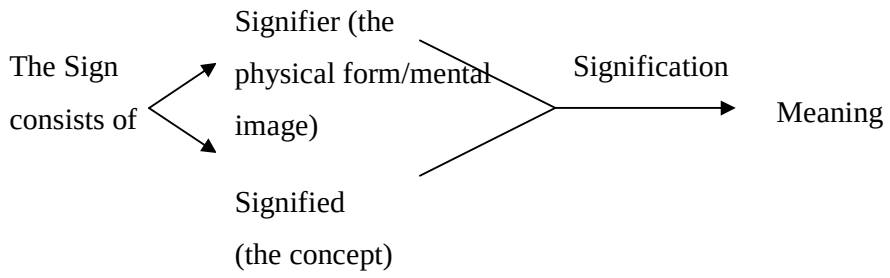

Fig 3. The composition of the meaning according to the model of de Saussure.

Saussure's model used two different categorizations for signs: they are either arbitrary signs, which denotes artificial symbols having no clear intuitive connection between sign and an object (such as abstract earcon); or iconic signs which seem or sound like object (Fiske 2000. 76-77).

Saussure's approach is considered more practical than Peirce's approach, despite similarities. Peirce's approach deals more with ontology (nature of existence) of the signs than their relation with each other, like de Saussure's model does. Model of Saussure is simpler, it deals more with relations of the signs, than ontology, and therefore it has been basis for analysis on languages and other such sign systems, like music.

3.3.3 Levels of meanings

Meanings have two levels. First, one that was the only interest of de Saussure, considers relation between signifier and signified, and signs relation to its object. Saussure was interested in language system and did consider only secondarily signs relation to real life object. This first level was called *denotation* by post – Saussure semiotician Roland Barthes (Fiske 2000, 113-114). Denotative meaning of the sign is the sign that is the most common and mostly accepted, a meaning that can be found for example in dictionaries. When speaking for example of a picture, denotative meaning is what stands in the picture, a dog or a tree for example. Denotative meaning is not arguable.

Barthes noted that denotative meaning is not usually the only meaning conveyed by some certain signs. There is also another level of signification: *connotation*. This second level refers to extra meanings that receiver of the sign attaches from his/hers own intertextual experience to the denotative meaning of the sign.

It is almost impossible to avoid connotative meanings when using signs. Connotations are personal because they depend on the person's earlier experience. At some level, connotations are common inside cultures. Denotation of a sign, whether it is a word, picture or sign, is usually unambiguous. Although from the sign, it can be also considered how it has been made, who has made it, what kind of feelings it conveys. Denotative meanings are not necessarily robust. Connotative meanings can be attached closely to denotative meanings and merge into different kind of meaning that was firstly in denotative meaning. (Fiske 2000, 112-121)

When we are speaking our tone of voice can connotate what we feel about a subject we are speaking about, and therefore change the meaning to something else. As considering earcons it has to be studied, what kind of connotative meanings are conveyed by the sounds we use, in order to design efficiently different types of earcons. If one focuses only on the quality of sound, on the timbre, it can convey different types of meanings than when sound is heard only as arbitrary/symbolic sound.

The studies of instrumental timbre induced emotions usually can be considered as a connotative meaning conveyed by the sound. There is no denotative mapping of some instrumental timbre being sad, or some being romantical or happy, nevertheless some instrumental timbres are considered so. (Lucassen 2006)

3.3.4 Motivation of the sign

Saussure considered sign's relations to its signified as iconic or arbitrary. He emphasised strongly the idea of language's arbitrary nature, that there is no natural mapping between a sign and signified; instead, he argued that all is based on rules and codes. Whereas Saussure emphasised his arbitrary signs as the most sophisticated sign class, so did Peirce consider his symbolic signs, which matches to Saussure's arbitrary. While arbitrary and symbolic refer to same kind with both Saussurean and Peircean approach, so does also iconic: they both consider iconic signs as the same.

Rarely a sign encountered in everyday life relies only on purely symbolic mapping. Almost all of the signs include levels of symbol and iconic signification. Motivation of the sign is always present. Letters and words are good example of the purely symbolic signs, but for example, traffic signs include levels of different types of signification. (Fiske 2000, 77-78) Therefore, they cannot be purely segmented as symbolic or iconic.

Motivation and restriction are terms from Roland Barthes. They refer to a level of how much signified controls signifiers (or the object controls its sign). The terms of Barthes, Peirce and Saussure correspond strongly. Strongly motivated sign is iconic; a photograph is much more motivated than a traffic sign. Arbitrary and symbolic signs are not motivated. Term restriction refers to limits signified sets to a sign; if a sign is strongly motivated, the signified sets some restrictions to the sign. (Fiske 2000, 77-80)

Photograph of a person is strongly motivated of the person's features. Painting of a person is not as strongly motivated, but still to some extent depending of the person's features. It can be more altered than the photo. Caricature of a woman is even less motivated than painting, but it is still *iconic*. Next level might be a picture familiar from the bathroom door; it is not that motivated by some certain person's features. On the other end can be a *symbol* of the person's gender; the sign of gender is not motivated at all, not restricted by the person's features, and

the sign is now depending on code. Examples of this are introduced in the figures below. The more unmotivated the sign is, the more important it is to know the code. If the interpreter does not know at all the code or habits, sign may not be understood, or it may be misunderstood.

Figure 4. Photo is a highly representative, iconic picture of a couple. Photo is strongly depending on the features of persons in the photo.

Figure 5. Painting of the couple is not so depending of persons' features. Some features may be altered considerably but models may still be recognizable.

Figure 6. This symbol is to some extent dependent on the human caricatures, but it is more symbolic than iconic.

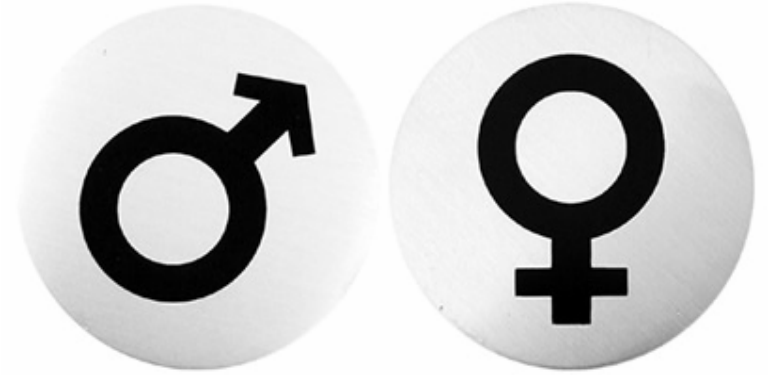

Figure 7. These familiar images are purely symbolic signs representing man and woman.

It is impossible to measure the level of motivation of a sign. Extremes may be found, but everything lying between is impossible to place on ordinal scale. Sometimes it is difficult to distinguish whether a sign is motivated by habit or if it is iconic. Some instruments are considered to be sad, but are they considered as sad just because there is sad spectral dimensions, or is it due to the habits; is it just because we have heard that kind of instrumental timbre in sad context? (Lucassen 2006)

Signs can rarely be divided into two rough categories of being symbolic/arbitrary or iconic. Different levels of motivation/limitation or symbolic/iconic should be considered on sliding scale, as presented below. When motivation becomes stronger, so does restrictions, and highly motivated sign is iconic, see the figure 8. The more the interpretation depends on the habits and conventions, i.e. codes, the more symbolic the sign is. Symbolic sign can be motivated to make interpretation easier, more intuitive, and yet remain symbolic. Hardly any sign we

encounter in everyday life is purely symbolic. Instead, they usually have some levels of motivation.

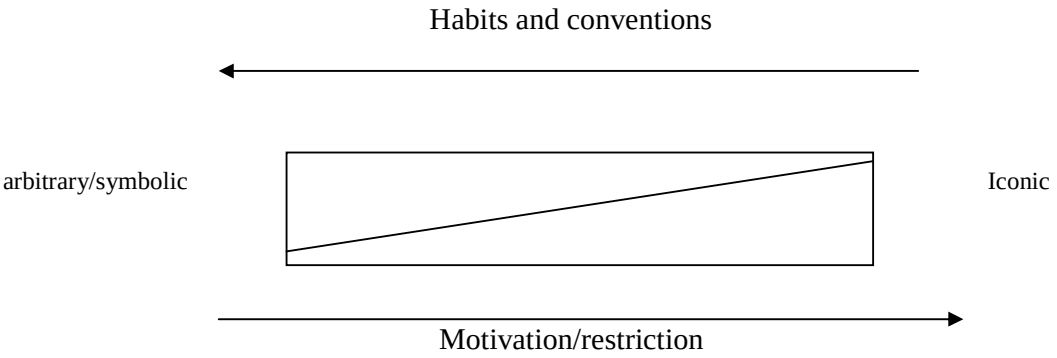

Fig 8. The Motivation and the restriction levels can be portrayed on sliding scale. On the rightward end would be a photo, and a gender symbol is on the leftward end.

Similar approaches discussed here with the image examples can be applied to sounds too. Door slamming in movie soundtrack represents event that someone just slammed a door. Even if sound is not natural, but it is composed of different sound samples and altered with computer, it can be very accurately interpreted as a door slammed. The opposite may be the familiar SMS message incoming beep-sound in mobile phones. There is no intuitive link in our surrounding natural soundscape to such sound. If one hears that type of beep at first time in his/hers life, unlikely it is interpreted as SMS incoming sign. By motivating a sign, its interpretation may become easier, more accurate, less demanding, and less dependent on knowing a certain code. In addition, motivated sign may prevent errors and improve recognition of the sign.

3.4 Defining a semiotic nature of earcon

Peirce's symbol and Saussure's arbitrary sign refer to a similar kind of sign. One needs to know some code, language or syntax to interpret sign correctly. Without such knowledge, it is impossible to understand what words mean or what syllable is some letter. This is the type of coding that is dealt with in contemporary earcon design, where earcons are purely abstract signs.

Blattner, et al. argued (1986): "*earcons are sonic counterparts for the visual icons in user interface*." Rarely a visual sign, like the icons on a desktop, relies only on purely symbolic or iconic signification. Almost all of the interface icons include levels of symbolic and iconic forms of sign. Motivation of the sign is usually present. (Fiske 2000, 72) Purely symbolic signs are rare considering the interface and other visual signs. For example, traffic signs include different forms of signs. For example, Finnish traffic sign that represents crosswalk has clear human caricatures and lines to represent the crosswalk itself. None of these sign parts is purely abstract, but they rely to some state on the characteristics of the entities they represent. Most of the signs we see are like that. Therefore, signs can be rarely segmented as symbolic or iconic. Visual icons on a computer desktop may have some features of iconic signs, and some features of symbolic sign.

Present day earcon and auditory icon discrimination is contradictory to the term that Blattner et al. (1989) introduced. First earcon description included iconic sounds under the term earcon. Earcons were distinguished as representational, abstract and semi-representational earcons. Present day definition of these terms causes confusion. Current earcon design issues are strongly dependent on this erroneous and unnatural segregation of the auditory icon and earcon. Earcon description does not include different levels of coding meanings into earcons, and all intentions to create more sophisticated intuitive auditory signs lead to new terminology, such as speech intonations imitating spearcons (Walker et al. 2006), or behavioural sound gestures imitating sounds (D'Incá & Mion 2006) that do not fit into earcon or auditory icon definition.

According to semantic difficulties of earcon and auditory icon description, it is recommended to use again the definition of Blattner et al. (1989) where earcons include auditory icons, which are called representational earcons. Other earcon levels are semi-abstract and abstract, semi-abstract earcons including the auditory signs that are in present day definitions excluded from the auditory display research field. These semi-abstract earcons are earcons that are motivated, intuitive auditory signs, which are not representational. In addition, the previous earcon definitions do not emphasize almost at all how earcons fit into context where used. Only few researchers have considered that, for example Pirhonen et al. (2006).

3.5 Taking the context into account

3.5.1 Paradigm and syntagm

Saussure considers coding of signs by defining *paradigm* and *syntagm*. Paradigm is a group of signs where units have something common together, something that makes them to be a group. For example, letters used in a language form a paradigm. Group consists of letters, not of numbers or special characters. In addition, units of paradigm need to be unique, and clearly distinguishable from other units of paradigm. An example of this are different handwritings: even if letter M is written in two different handwritings, it is still distinguishable from letter N. Differences which make units of paradigms distinguishable are called distinctive features. Paradigms are composed of different paradigms. Words are a paradigm, which is composed of e.g. nouns, verbs, and adjectives. (Fiske 2000, 81-85)

Usually units of paradigm are not represented alone; instead, they are usually presented with other units of paradigm. For example, words are rarely represented alone, but usually represented with other words, thus creating a logical sentence. Such a harmonious combination of different chosen units of paradigm is called a syntagm. Units of paradigms chosen to syntagm are often determined by rules, habits and preferences.

Syntagm of a sentence requires one unit from the verbs paradigm, one unit from the noun paradigm etc. Syntagm of clothing requires shoes, pants and shirt, at least. Certain words do not suit with some other words, stylistic characteristics and appropriate meanings limit characteristics for sign selection. In auditory displays, this may stand for the fact that all

sounds in auditory display should be considered in one context, where sounds should be congruent with other sounds and entities they represent. Usually it may be a visual icon, or some hidden entity. If presented earcon does not fit in the context, represented earcons may become irritating or disadvantageous. Whereas visual display's coherence usually is constructed by colour schema, timbre may be the key dimension in creating coherent auditory display, due to its nature of being the colour of the sound. Nevertheless, not only the context of other sounds is important: this approach is important to consider in the whole context of use. (Murphy et al. 2006)

Paradigmatic choice refers to a process where it is possible to combine from all the selected signs those signs that can create desired meanings. There is usually a vast array of different signs possible to choose. Some signs have bit different meanings, but still they may be among the same paradigmatic possibilities. For example, the stallion representing Ferrari sports cars, can be probably from the same paradigm which includes mule and donkey, but is it the same thing? Not all the alternatives are possible. This may be due to the connotations of the donkey not being as masculine as stallion. The same thing exists when considering sounds. Distorted guitar is usually used to represent youth and strength in movies, but one has to consider, if a flute or mouth harp is suitable to represent that? They are from the same paradigm, but still differing. This aspect has to be considered in auditory sign design too.

To some extent, similar approach to this is to think paradigmatic choices as metaphors. Metaphor's function relies on the paradigm. Metaphors are defined as *"a new and unfamiliar entity is expressed with the help of familiar concepts by pointing out the similarities between the two."* (Pirhonen 2005) An example of the metaphor is in this case to use the stallion to represent the power of the sports car. The stallion is renowned of its power and masculine nature. The association of the power of the stallion is transferred into sports car. The power of the metaphors is that the attributes of a familiar concept (which is called *vehicle*) is associatively transferred into the unfamiliar concept (which is called *tenor*). The tenor and vehicle has to be similar in the way that they can be placed in the same paradigmatic group, but different enough to be contrasted. (Fiske 2000, 122-131) Horse and car are from the same paradigm, means of transportation. However, they are different in the way that car is not realistically masculine. The masculine attributes are transferred into car. Metaphors are not

realistic, but rely on imagination. Metaphor consists of two domains, which have something common together. Metaphors are used in order to support the process of conceptualisation.

In this metaphorical approach, some attributes of sounds can be associated with entities an auditory sign represents. Metaphors are in an important part of our everyday experience conceptualization. When spoken of the “higher class of society” or “higher morale” we are using metaphors to conceptualize abstract entities. “Higher” refers usually to “better”. This has no natural reason, another way to consider “higher” could be worse. If we fall from a high tree, we get hurt worst than from lower. In this case, we use counterparts “high” and “low” in order to understand something that is abstract. In sounds, this metaphorical level is used commonly; rising level of sound refers to increasing amount.

When there are paradigmatic choices (signs that are possible to combine), there is potential to create meanings, and when there is syntagmatic combination there is a realized meaning, which is combination of units of paradigm. Syntagmatic analysis refers to analysis of the *meaningful whole*, including how signs fit into their context. If considering the stallion example above; the syntagmatic analyses considers the most suitable to choose - a stallion, a mule or a donkey to represent the values of Ferrari.

Syntagmatic analysis are not quite often used in studying interfaces. In the field of auditory interfaces only few have taken into account the syntagmatic analysis of context when using sound in interfaces, for example Murphy et al. (2006). Interpretation of the sign leans on other signs, and on the whole context. In sound design, all the sounds of interface and all surrounding soundscape combine the meaningful whole. If one sign does not fit into syntagm, it may lead to disadvantageous, dissonant sign. Important aspect in syntagmatic analysis is to take into account the whole context where sound is used. To some extent, similar approach to paradigmatic choice and syntagmatic analysis is introduced in the audiovisual contract.

3.5.2 Audiovisual contract

When a sign is presented in user interface, it is meant to direct user’s behaviour to desired way. If the sign’s conceptual model, which refers to the way how system is understood as a system, is not matching to user’s prior knowledge, or is in conflict with earlier a priori knowledge and connotations, sign may lead to undesired result and errors. This problem is

strongly similar to syntagmatic analysis introduced above. It is also to some extent more advanced.

Approach is called audiovisual contract, (Chion 1994) which considers auditory output and visual output as the same context. Neither can be considered alone, they form interactive relation. This approach has been applied widely to cinema and theatre, where audiovisual narration is in an important role, but user interface studies have not considered this. Lemmens and Bussemakers et al. stated out an example of ventriloquist speaking and simultaneously moving the mouth of the puppet. This is how audience is deluded to believe that the voice is coming from puppets mouth, and the ventriloquist is not speaking. (Lemmens et al. 2000)

Nevertheless, this kind of illusion is very easy to break. In the ventriloquist case described, synchronicity and proximity of ventriloquist and the puppet are used in order to create congruent visual and auditory perception that can be combined into a realistic perception. If this does not happen, sign may become distracter instead of assistance. An example of errors that can be made in the earcon design which lead into this kind of problems is that connotations (conceptual model of sound meaning) are not taken into account, and signs are maybe even used in a contradictory meaning.

In studies of audiovisual design (theatre, cinema), different modalities supporting each other can be referred as *added values*. Especially media based on visual output has added value from the sound. An expressive and informative value enriches the other modality. For example if sound is presented together with an image synchronically, value is added via principles of synchresis. Such as in ventriloquist example above, synchresis, for example in movies, is illusion, but powerful illusion. Sounds presented alone are very different than sound presented with images. Images presented with sounds are different from images without sound.

Phenomenon introduced here, added value, has importance in human computer interaction, and auditory display design. As the father of the audio visual contract, Michel Chion argues: "*Added value is what gives the incorrect impression that sound is unnecessary, that sound merely duplicates a meaning which in reality it brings out, either all on its own or by discrepancies between it and the image*." (Chion 1994, pp 5)

Chion introduces several ways to use sound in audiovisual context; Unification, punctuation, anticipation and separation. Unification is the most widely used function of sound in films. This means that sound is chaining the flow of images, and unifying the atmosphere by some certain soundscape. Third role of the unification sound is unifying the whole context by non-diegetic sound which has no visual nor known source) music, sounds or soundscape.

At the extensive definition of punctuation, it refers to all the placement of commas, semicolons, question marks, periods etc. in grammatical context. Punctuation is a very effective mean to change and derive different gestural roles. By using exclamation mark, one sentence can be more obtrusive than sentence without exclamation mark. Different kinds of gesture imitations have been used in cinema to create punctuation via the audio.

Anticipation, which considers divergence and convergence of sound, refers to the feature that sounds and images have tendencies. They indicate certain direction, actions, entities etc. Usually sounds and images follow certain pattern, and evoke expectancy of something to follow. Observer expects a certain form to be followed. These forms may consist of styles, habits, conventions etc. For example, musical form leads the listener to expect cadences, and listener expects the silence to be filled with matching cadence. Certain sounds lead observers into state of expectation of pleasant continuum.

Finally yet importantly, usage of sound in audiovisual context is silence as separator or obtrusive effect. To some extent, silence can be used for effect. Silence, however, is rarely alone. In order to perceive silence, some other noises or sounds has to exist at first. If listeners are firstly exposed to clearly audible noise or sound mixture, silence may be very effective in order to arrest attention. (Chion 1994, 47-62)

Why these means have not been widely accepted to auditory display research? These aspects are strongly been affecting the audiovisual design for example in the cinema, but these aspects have been not considered at all with auditory displays. Although there is a wide gap between motivations of auditory display and cinematic narration, these aspects cannot be ignored. Visual icon and symbol design has taken extensive influence from arts and visual narration. There are strong connotations of sound and syntagm-related expectations in

audiovisual context that cannot remain unconcerned when creating effective non-speech interaction.

An important aspect to audio display design, Chion argues, is the audiovisual dissonance. He argues that there are hundreds of matching images to one sound. Moreover, form this vast array of possibilities some are wholly predictable. Others, which do not fit into these matching possibilities, may lead to contradicting or unwanted connotations. Which images and sounds match each other, is ambiguous. Some images and sounds match better than others, some not at all. If sound does not correspond with the visual representation, it may lead to auditory signs to be disadvantageous. (Bussemakers & de Haan 2000, Chion 1994, Lemmens et al. 2000)

How does these different matching and mismatching sounds differ from each other? It is hard to discriminate whether the sound presented with visual counterpart is causing audiovisual dissonance or not. Can auditory display designers trust on learning purely abstract, symbolic sounds? In the next, I will present reasons whether sounds used can be purely abstract, randomly selected sounds that only need to be learned in order to be effective.

3.6 Modes of listening

Chion argues that humans have three different, overlapping listening modes. Example of these different modes of listening is an easy field test. If we ask someone to tell, what he has heard during the past day, we get several different answers. Some people tell about the sound sources they heard, some tell what things their friends have told them, and some tell what kind of sounds they have heard; noise , loud sounds, or they describe elevators squeaky sounds which included some hissing sounds.

Aspect of these listening modalities takes into account that human perception system is not “passive”, meaning that everything is perceived by the ear and by the mind. Due to the primitive all-aware nature of hearing, sound always affects and overwhelms mind, sometimes even surprisingly. Sounds work on us directly, physiologically. Some sounds may affect our behaviour, for example breathing noises may affect our own respiration. Three listening modes are not separable, but always overlapping each other.

3.6.1 Semantic listening

Semantic listening refers to listening and interpreting some certain codes or languages into messages. When we listen to our friend talking about weather, we are listening to them with semantic listening mode. This mode of listening is interest of linguistics and it is very complex. Syllables and phonemes are not listened individually and judged alone, but they are interpreted in a context with other syllables; organizing syllables into words, words into sentences and finally sentences are interpreted as messages. (Chion 1994, 25-29)

This listening mode does not take into account acoustic features, such as how loudly words are said or how words are pronounced. This listening mode can be referred also as the listening mode how abstract earcons should be interpreted. However, one has to take into account that this listening mode is not in use alone.

3.6.2 Causal listening

Whereas semantic listening is the most common mode used in communication, causal listening is the most common mode in listening soundscape. Causal listening consists of the listening to a sound in order to gather information of the source itself. For example, causal listening is used when hearing the mailbox when mail carrier is dropping mail into it. By the sound itself, we can hear whether there was a big book dropping into mailbox, or whether it was just a small letter.

Causal listening refers to various levels. When we recognize person's voice or some unique sound from certain object, causal listening is used. Other level of the causal listening is that we do not listen to identity of the source, but listen to what is the category of the source, e.g. categories can be of male voices, big motors, song or fire. (Chion 1994, 28)

3.6.3 Reduced listening

The last listening mode is reduced listening, which refers to listening where listener is focusing on the dimensions of the sound itself. Sounds are interpreted as itself, without considering whether it may be a sign for something else. When subjects are asked to describe what they hear by the means of reduced listening, overlapping of the listening modes comes

obvious. When subjects describe sounds as “stinging” or “mellow”, one has to consider whether subject is describing the sound itself, or the effect the sound has on him/her.

Because of these different modes of listening, and their inevitable overlapping, it becomes obvious that symbolic, abstract sounds cannot be purely abstract, but connotations, subjective satisfaction of sound and all the effects of sounds should be considered. (Chion 1994, 29-30) Whereas in spoken communication pronunciation has an effect on listener’s interpretation of the messages by connotations, so may be the situation with symbolic sounds when no connotations are considered. If present day definition of earcons excludes possibilities of considering these aspects of connotations and subjective satisfaction and the meaningful whole, earcons should be defined more accurately.

3.7 Guidelines for earcon design

Interface sign design is not necessarily a simple task. As in all interface and communication design, the design and creation of earcons depends on the interface itself and processes the application behind does. Exactly same design principles cannot be utilized to mobile entertainment applications and power plant control applications. There are some attempts to derive guidelines for earcon design.

Blattner et al. used a strong analogy to visual sign design, which consists of motives. In their guidelines of earcon design Blattner et al (1989) suggest that careful design of motives enables good design of earcon families, which enables effective use of earcons.

As where any symbol is created from elements, so are earcons. These guidelines suggest that motive’s fixed parameters, rhythm and pitch has to be constructed by the syntax of western tonal music. According to Blattner et al (1989), earcons are identified by these parameters and variable parameters as timbre, register and dynamics are not that important. Blattner et al argue that to recognize familiar tune these variable parameters may vary greatly. In these guidelines it was suggested that rhythm is the most important factor on recognition of sound sources, so earcons were designed by the motives which were described as; “*sequences of pitches that create a short, distinctive audio pattern often characterized by the simplicity of its*

rhythm and pitch design". Moreover, sounds were suggested to be pure tones, rather than with richer spectrum timbre.

Next attempt to derive guidelines for earcon design were by Brewster et al (1993), these guidelines of earcons were differentiating from the first guidelines by emphasising to some extent the importance of timbre. They suggested also that musical timbre should be used in order to help perception and avoid masking. These guidelines emphasize rhythm usage in earcons, due to the importance of rhythm in differentiating structured and non-structured sounds.

Next derived guidelines for earcon design were by Brewster et al. (1995) and Brewster et al (2007). These guidelines suggest that creation of earcon families should start with timbre, register and rhythm. These are the basic structures of families. Each family of earcons should have different timbre and default register. Every created family of earcons have their own register, rhythm and/or timbre. Timbres should be chosen to be clearly distinctive from each other, as Brewster et al. (2007) suggest using for example musical instrument timbres "brass" and "organ" rather than "brass1" and "brass2".

Brewster et al. (2007) suggest that using instrumental timbre, register and rhythm is the best way to create families of earcons. Using timbres with richer spectres (timbres with multiple harmonies) rather than pure tones, (which Blattner et al. suggested to use) it is possible to avoid masking. Other main dimensions for creation of earcons are register and rhythm.

Also, in earcon designing it is necessary to create pitch between 125 kHz and 5 kHz to avoid possible masking and to avoid being inaudible to some listeners with weakened hearing. Rhythm and duration needs to be created as different as possible, in order to create distinguishable earcons. Also too short sounds may not be noticed, so duration needs to be long enough to be perceived as a sound. Brewster suggests notes longer than 0.0825 sec, and if tone used as earcon is simple, it can be as short as 0.03 sec. In order to decrease time taken to perceive earcons and to recognise meaning. (Brewster et al. 1995)

Brewster (2007) discusses the intensity level of earcons; all the sounds used in auditory interface need to be in the same rather narrow intensity range so if user adjusts main volume,

all auditory signs should remain audible and not become too loud. Too loud sounds may become too irritating and too quiet sounds may become masked by other sounds.

Brewster et al. (1995) criticised earcons because they are so strongly dependent on the exact time when event occurs. They conducted experiments of parallel, shorter earcons. Experiment proved that recognition rates did not fall significantly until three sounds were presented. Two earcons played simultaneously provided reliable recognition rates. Also combining earcons (such as earcon meaning "open" combined with "file" earcon created "open file" earcon) reduced duration of earcons.

McGookin and Brewster (2004) developed idea of parallel auditory signs further and created some guidelines for designing concurrent earcons. First guideline suggests keeping number of earcons limited. Increasing number of concurrently presented earcons reduced amount of successfully identified earcons. Second guideline speaks out register as parameter:"*If register is used to encode a data attribute, it may be beneficial to ensure that in harmonic intervals are used between earcon concurrently presented in different registers."* Third guideline suggests principles of timbre selection. Each concurrently presented earcon should have different timbre. But if two earcons are carrying a similar timbre-encoded message, timbres should be selected from the same instrumental timbre group, for example "organ1" and "organ2". Fourth guideline suggests that concurrently presented earcons should not start exactly at same time. At least 300-millisecond gap between onsets of sounds was recommended.

3.7.1 Criticism of the guidelines

Guideline is defined in Oxford dictionary: "*Rules or instructions that are given by an official organization telling you how to do something, especially something difficult, guidelines are something that can be used to help you make a decision or form an opinion.*" (Oxford Advanced Learner's Dictionary 2005)

Although studies presented earlier bring out the problem that there are no decent guidelines for earcon design, and they attempt to create some, they actually do not provide much aid for designers. These guidelines introduce some issues to be considered in order to distinguish different earcon families, but do not help to make decisions while designing meanings, which

are inarguable, the most important part of the sign. With these guidelines designer can design *sounds* that are to some extent distinguishable, but do not provide any reasons for creating *signs*. By following these guidelines, semantic design of earcons yet remains ad hoc – based. Earcon studies have forgotten to a certain degree what earcons are for in human computer interaction. These studies do not take into consideration the usability issues, of how to improve learnability, efficiency of use and memorability of earcons, and how to prevent errors and maintain subjective satisfaction.

If visual icons were designed with similar guidelines and restrictions as earcons, they would consist only of basic geometrical shapes, squares, triangles, points and lines, with no consideration if the images fit in the other context, or the entities they represent.

Earlier earcon design guidelines have no attempt to provide aid for sign creation, and they take no account of motivation of the sign, or context of use. Due to these aspects have not been considered, earcons have been proven to be possibly very disadvantageous. To avoid the kind of distraction earcons may cause, design guidelines should consider the connotations, which sounds have. Previous guidelines of earcon creation do not emphasise this at all.

The severest flaw of all of the guidelines is the not so well considered question "*Why to select some certain timbre to some certain earcon.*" After the Blattner et al (1989) earcon description that included representational, semi-abstract and abstract earcons, the more intuitive meaning of the earcons was forgotten. Since then there has been no attempt to improve intuitiveness with earcons. This may be due to the present day distorted earcon and auditory icon definition, which neither includes semi-abstract, to some extent an intuitive nor motivated level of sign. This discrimination is very harsh, due to the multilevel signification of symbols and icons. This has led to severe flaws in earcon design, which leads to reported errors (Bussemakers & de Haan 2000, Lemmens et al. 2000). This is a clear indicator of neglected usability evaluation of earcons. To criticize previous approaches it is necessary to think earcon itself: is it only the melody that makes difference when creating meaning of sounds? Can earcon designers trust that all users learn all purely abstract earcons?

The earlier research of Lemmens et al. (2000) introduced in chapter 2.2.3 showed that sounds have no effect on performance, unless sound is somewhat incongruent. In that case, sound

usage may hinder the categorization task. This study, however, studied earcons as short chords, and sounds used were chords in major or minor, which was the only altered dimension in sounds. The experiment of Lemmens et al. (2000) proves flaws of the earlier earcon guidelines. Earlier guidelines do not provide aid to design sounds that are congruent to meaning it conveys. One, effective and not considered mean is to use timbre.

Timbre has been poorly understood earlier in earcon design, perhaps partly because of timbre's complex nature, and in comparison with other perceptual attributes, especially pitch and loudness, timbre is more multifaceted to understand and study. That may be the reason why it has been so little considered in earcon design guidelines. Yet, some studies have been introduced of timbre, and those studies may be important in the context of earcon design .(Poulin-Charronnat et al. 2004, Padova et al. 2005, Padova et al. 2003, D'Incá & Mion 2006)

Timbre has proved to be a key dimension in emotions and meanings induced by music (Yost 1997, Padova et al. 2005, Iverson 1995). One of the most effective ways in evoking emotions and conveying meanings in musical timbre is behavioural gesture imitating playing style, e.g. instrument played angrily is always interpreted as angry, due to the spectral characteristics of timbre. (Zagorski-Thomas 2005)

Some guidelines consider how sounds can be as distinguishable from each other (Brewster et al. 1995, 2007, Mcgookin & Brewster 2004) Auditory stream segregation, auditory scene analysis and identification of the musical materials are nevertheless been in modest role. Guidelines suggest that sounds should be selected from MIDI instruments as different from each other as possible. Source detection and identification is more complex, but well studied area.

According to guidelines from Blattner et al (1989), it is easier to discriminate melodies and rhythms from each other and using different timbres are not important. In the studies of auditory scene analysis and musicology, timbre has proved to be the most important dimension in distinguishing sound sources and in recognizing sound. This has been taken into account in guidelines developed after Blattner et al. (1989). If sound must be identified by only one varying dimension (in this case meaning temporally varying pitch), it is possible to

name only 5-6 at time to identify more than 5-6 different sounds, so to create sufficient amount of earcons it is necessary to use more than one dimension. (Moore 2001, 26)

According to studies of psychoacoustics, it has been proved that if a sound varies only by one dimension, for example by frequency, human auditory system is able to recognize about 500 different frequencies and for intensity, 100 different intensity levels can be perceived. Moreover, if sounds vary by these both dimensions it is possible to tell apart between 300 000 and 400 000 distinguishable tones. This discrimination requires that sound can be compared to another sound, so subjects name differences between sounds. Absolute judgement, which identifying without comparing to other sound, provides different results. If there are no counterparts, human auditory system can distinguish five to nine different sounds that vary only by one dimension. Ability to identify stimuli that varies along one dimension is very limited. According to experimentation of Pollack and Ficks when varying sound dimensions were raised to eight observers, they were able to identify about 125 different sounds (Handel 1989, 269-270). It has been proved that melodic changes do not affect identifying characteristic features of different timbres of instruments. (Deliege and Sloboda 1997, 272-3)

Sounds with instrumental timbre are recognized more accurately than simple pure tones; as Blattner et al (1989) suggested. Using pure tones rather than sounds with complex spectres has proved to provide unusable guidelines. Several studies (for example in musicology) have taken account of musical timbre as being the most important dimension in recognition of musical materials (Padova, Bianchini, Lupone & Belardinelli 2003, Padova. et al. 2005, Poulin-Charronnat, Bigand, Lalitte, Madurell, Vieilleard & McAdams 2004, Zagorski-Thomas 2005). This has been taken into account to a certain extent on later earcon design guidelines, but not extensively.

Present day guidelines still emphasize importance of the rhythm as the most important dimension in distinguishing sounds from each other. These guidelines suggest that rhythm should be as differentiating from each other as possible in order to create easily recognized families. (Brewster et al. 1995, Brewster et al. 1993) Studies, which are used as the basis of this argument compare rhythm with unstructured sound bursts. (Deutch 1980) This approach is to some extent inaccurate because earcons are all structured sounds, not unstructured sound bursts. E.g. Poulin-Charronnat et al. (2004) demonstrated that timbre may be more important

than pitch and rhythm in identification and recognition of structured sounds. The same kind of results are given in auditory scene analysis (the study of human auditory system's ability to discriminate auditory sources from complex soundscape) and in musicology. (Padova et al. 2005, Bregman 1990, Jensen 2001)

To resolve all these flaws introduced above, timbre may be the main dimension. Timbre has not been studied sufficiently in earcon design, but the literature of psychoacoustics, auditory stream segregation and musicology have relevant issues to design of semi-abstract intuitive earcons.

4 PSYCHOACOUSTICAL APPROACH TO TIMBRE

4.1 Timbre definition

Commonly used definition of timbre is by ASA (American standards association): "*the attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch are dissimilar*." (Deliege and Sloboda 1997, 257) A flaw in this definition is that it defines what timbre is not. It actually does not take any account on what timbre is. (Jensen 2001)

Difficulty with the timbre definition is that some instrumental timbres are similar with different instruments at the same pitch than with sounds originated from the same instrument but at different pitch. For example, spectral differences of high-pitched piano and flute are small, but high-pitched pianos and low-pitched pianos spectral differences are noticeable. Despite of the great difference between high-pitched and low-pitched piano and similarities between high-pitched piano and flute, human auditory system identifies sound sources accurately. In conclusion, rather than referring to timbre as wastebasket of dimensions that are not pitch, duration or loudness, timbre can be defined as identity of the sound. (Grey 1977, Jensen 1999a, Jensen 2001, Jensen 2002) Timbre does not refer only to the identity of the sound. In addition, timbre is referred to as quality and "colour" of the sound, including emotions and meanings induced by the sound. (Zagorski-Thomas 2005, Padova et al. 2003, Padova et al. 2005)

Timbre is understood to be complex and multidimensional by nature. Whereas other dimensions of sound, such as duration and pitch can be placed on the ordinal scale from high to low or from short to long, timbre cannot be placed on such scale. Timbre dimensions are affecting the spectral envelope and amplitude envelope etc. of the sound. Defining timbre is not as easy as describing the duration or pitch of a sound, but it is even more complex to identify one timbre over another.

4.1.1 Verbal attributes

A commonly used method for identifying timbre is giving verbal attributes to different stimuli's in order to create a semantic differential scale. Method was developed e.g. by von Bismarck. He had subjects to rate speech, musical and artificial sounds on 30 verbal attributes. By scaling result multidimensionally, he found that there are four axes. First axe related with the word pair dull-sharp, which was later found to be determined by the frequency position of the overall energy concentration of the spectrum. Second word pair was compact-scattered, which was found to be determined by the tone/noise characteristics of the stimulus. Third axe was the pair full-empty and the fourth colourful-colourless. Axes three and four were not found to be determined by any specific quality. (Jensen 1999a, Deliege and Sloboda 1997, 257-8, Grey 1977)

This is a very subjective method, but relevant, because it is important to be able to talk about timbres. This verbal opposition set makes it possible. For example defining piano and violin, piano can be defined as dull, hard, clean and decaying whereas violin may be defined as bright, soft, dirty and sustained. (Jensen 2001)

4.1.2 Dissimilarity tests

The dissimilarity test is aimed to find similarities between timbres of different musical instruments. Subjects judge dissimilarity of number of sounds. In order to find relevant timbre qualities of dissimilar timbres judged by subjects, a multidimensional scaling is used to analyze sounds. The aim of multidimensional scaling is to reveal relations inside the set of stimuli. Stimuli are represented in low-dimensional space so distances between the stimuli echo their relative dissimilarities.

By the multidimensional analysis, Grey (1995) found that the most important dimension in timbre is spectral envelope. In addition, attack-decay behaviour and synchronicity were found important, along with spectral fluctuation in time. (Grey 1977) Later on Krimphoff et al. found that most important timbre dimensions are brightness, attack time and the spectral fine structure (Jensen 1999a). Removal of the attack transient hinders widely the recognition of musical timbre, although some instruments, such as flute, are not that dependent on the attack transient than others. (Thayer 1974)

4.1.3 Multidimensional nature of timbre

Timbre is multidimensional unlike other sound dimensions and timbre cannot be presented in a single scale that is used for other dimensions; such as volume or pitch that can be situated on a single scale from loud to quiet and high to low. One instrument cannot be placed on a scale whether it is greater or less than any other instrument. (Moore 2001, Jensen 2002, Deliege and Sloboda 1997, Grey 1977) What are the dimensions that we use to identify sound source? How we identify some instrument timbre or familiar voice? Our ability to identify one source or object among many large set of objects depends upon several varying dimensions of timbre. Firstly, we can identify complex stimuli by the patterning of energy as a function of frequency. Secondly, a stimulus typically varies within time, and the temporal patterns can act in an important role in perception and identification of source, such as fluctuations over time and states of timbre.

It is likely that the number of dimensions required to characterize timbre is limited by the number of critical bands required to cover the audible frequency range. This would give the maximum of 37 dimensions. It appears to be generally true, both for speech and non-speech sounds, that the timbres of steady tones are determined primarily by their magnitude spectra, although the relative phases of the components may also play a small role. (Moore 2001, 245-255, Iverson 1995)

4.1.4 Time unvarying patterns

In contract of steady sinusoid, this can be defined accurately by two dimension; frequency and intensity. Tones with complex spectral envelope (meaning timbre), which we encounter in everyday life, are more difficult to define. Major determinant of the timbre is distribution of energy over frequency, spectral envelope. For example, sounds with strong lower harmonics are defined as soft and smooth, and tones with strong higher harmonics are usually considered as sharp and penetrating. (Moore 2001) As mentioned above, spectral envelope has been proved the most important dimension in timbre identification.

4.1.5 Time varying patterns

Spectrum of the sound is usually altered remarkably within time. Schouten suggested that these identification-affecting changes in time are; (a) whether the sound is periodic, having tonal quality repetition between 20 and 20 000 per second, irregular or does have noise like character. (b) Whether the waveform envelope is constant or does have fluctuation in time, and if there is fluctuation, what kind are they; (c) is the spectrum or periodicity changing in time and (d) what the previous and the subsequent sounds are like. (Yost 1997)

Differences in static timbre are usually enough to distinguish two *sounds* but more likely they are not enough to distinguish two different *auditory objects*. (Moore 1001, 247) Reason for that is that sound from an instrument can be greatly varied by the environment where sound was played, such as echoes, reflections and muting effect of soft surfaces in the room. Although in perception of earcons, this is not subject affecting the recognition of timbre and sound, but not only environment causes fluctuations and temporal variations within spectral envelope.

Timbre is usually considered to consist of different segments or stages. The used terms are varied, but basic idea is the same: timbre consists in temporal order of (a) onset-transient (or attack-transient, initiation noise) (b) sustain (or ready-state, static-stage) and (c) release-transient. (Jensen 2001, Handel 1989, Moore 2001, Thayer 1974) Recognition of musical

instrument or imitating digital sound depends on the onset transients and temporal structure of the sound envelope. As proved by Thayer, removing attack transient may hinder the recognition of the musical instrument. (Thayer 1974) One of the strongest forces in instrumental timbre recognition is onset transients and the temporal structures of the envelope of sound. For example, recognition of piano is strongly dependent on its strong and fast onset and rather slow, step-by-step decay. Few studies have demonstrated effect of changing temporal asymmetric envelopes. For example, the piano is not recognized when played backwards. (Moore 2001)

Patterson created sinusoidal carriers that were amplitude modulated by repeating exponential function in order to create two kinds of sounds: Sounds that were “damped”, which means that envelope increased suddenly and decayed slowly, and sounds that were “ramped” which means that envelope increased slowly, and decayed suddenly. For an envelope repetition rate of 4 ms, damped, sudden-onset, sounds were perceived as single source, like drum sound with echo, and ramped, slow-onset, sounds were heard as two different drum roll on a non-resonant surface with continuous similar tone to carrier frequency. Later Patterson and Ackeroyd made same kind of test with broadband noise as a carrier. In this experiment, damped sounds were described as drum struck with wire brushes, and ramped sound was heard as noise with hiss-like quality that was sharply decayed.

According to described nature of timbre and importance of spectral envelope in recognition of timbres, it is rather challenging to synthesize instrumental timbres. Many instruments have noise-like qualities influencing strongly to their recognition and discrimination from others. Flute, for example, has simple harmonic structure but starts with noise-like “puff” sound. Dynamic variations with time characteristics of instruments are important. If instrumental timbre is simulated only by on-state spectral envelope summation, it is practically not recognizable. A great deal of source recognition relies on the temporal noise-like characteristics. (Moore 2001) For example, *musical Instrument Digital Interface*, MIDI is “lifeless” compared to natural instruments spectral envelope, and using synthesized sound may hinder the recognition of auditory sources. (Zagorski-Thomas 2005) The same dilemma occurs if it is necessary to use imitations of behavioural expressions to create meanings.

4.2 Timbre recognition and memory

As referred earlier, usability of the systems consists of five things. There are few aspects to consider when evaluating systems usability; learnability, efficiency of use (intuitiveness), memorability, errors and subjective satisfaction. (Nielsen 1993) Nowadays earcon design does not take into account these aspects, whether earcons are easy to remember or not.

Commonly cited theory of human memory is that memory consists of two different stages, long-term memory and work memory. This two-stage memory division has received some criticism (e.g. Ericsson & Kintsch 1995) but for the needs of this study, this classification is sufficient. Work memory is a short-term memory, which storage capacity is only at the highest 2 seconds. It is used to store temporarily information that is processed in any of a range in cognitive tasks. Processed information that is considered important and relevant is entered into long-term memory, which is capable of storing vast amount of information. Storing information in long-term memory is not a conscious process; instead, it is usually an automatic cognitive process.

Information is retrieved from the long-term memory when needed for cognitive processes. The capacity of long-term memory is not the main issue, but how the information is retrieved, which is, how we remember things. The processing of the information afterwards, which is remembering, is dependent on how we interpreted and stored the material in first place. If the stored material makes sense in the first place, which means that it coheres with the information we already know and we consider it important and meaningful part of the context, it may be easily restored. If material was to some extent abstract, hard to understand and fit to other knowledge, it has to be processed and structured before it can be remembered again. Norman (2001, 66-74) discusses that memory is divided into three different memory types; memory for arbitrary information, memory for meaningful relationships and memory through explanation.

Relying on arbitrary memory may create problems, because it is arbitrary. Learning and remembering is difficult, it may take considerable time. Secondly, if information from arbitrary memory is forgotten, there are no associative cues to restructure information. Meaningful structures and associations simplify extensively memory tasks. Information can

be retrieved again with the help of associations. Memory through explanation refers to memory that is the most efficient according to Norman (2001). This memory relies on understandings. Interpretations of events are fundamental to human performance of remembering, learning and realizing some meaningful combination of information, as is the case with e.g. metaphors. After learning and becoming aware of something, it is memorized effectively, and it can help to retrieve and derive information. Considering this, connotative and metaphorical use of timbre is important in auditory signs. If the signs are considered to correspond to the context and have strong connotative associations with entities they represent, they are recognized easily.

Some studies e.g. (Padova et al. 2005) have proved that timbre is an important element in perceptual processing of sounds, and timbre can be referred as a fundamental source cue. To source information it is necessary to source recognition, and timbre is fundamental, high-source cue whereas shifts and pitches are referred as secondary, low-source cues and therefore do not play as important role as timbre in memorizing sounds.

This is because timbre is associated with some musical instrument and therefore timbre is a clear indicator of source. "*Timbre was used as source information to help retrieve the memory trace: timbre carries information that easily identifies the source from which the melody originated"*. (Padova et al. 2005) Human perception organizes timbres categorically, which is not the case with rhythm and melody. (Jensen 2002) It is possible to name timbres into different categories. Categorical memorization is very effective in human cognition. As we may find it difficult to remember the name of a person we have been introduced to just minutes ago, we can remember him as by his gender, height, or by other multiple features that we can use to categorize the person. (Preece 1994)

If the instrumentation of the melody is changed, subjects are less able to recognize melody played. Radvansky and Potter (2000) conducted experimental recognition task where melody was altered by two dimensions, timbre and pitch. It was found that features such as melody's timbre can be powerful method to improve recognition and retrieval of information. In memory tasks, timbre is proven the most important feature. Padova et al. (2005) proved in their study that in recognition tasks salience is more important feature in stimuli than tonality, and in absence of salience, timbre is used to aid recognition decisions.

4.2.1 Identification and recognition of timbre compared to other dimensions

It has been proved that if sound varies only by one dimension, for example by frequency, human auditory system is able to recognize about 500 different frequencies and for intensity 100 different intensity levels. Moreover, if sounds vary by both of these dimensions it is possible to distinguish between 300 000 and 400 000 different tones. This means that sound can be compared to another sound, so subjects name differences between sounds. (Deliege and Sloboda 1997)

Absolute judgement of identifying sounds provides different results. If there are no counterparts, human auditory system can distinguish five to nine different sounds that vary only by one dimension. Ability to identify stimuli that varies along one dimension is very limited. According to experimentation of Pollack and Ficks introduced in Handel (1989) when varying sound dimensions were raised to eight observers, they were able to identify about 125 different sounds. I melodic changes do not affect identifying characteristic features of different timbres of instruments (Deliege and Sloboda 1997)

Based on just noticeable differences (JND), human hearing sense can distinguish 1500 pitches and 325 levels of loudness, but amount of recognizable timbres remains unsolve.d (Deliege & Sloboda 1997) Obviously, as an object-identification dimension, amount of recognized timbres is vast.

Timbre can be described as the identity of sound source. Although e.g. high pitch piano timbre parameters are closer to flute than low-pitched piano, but timbre is identified correctly and accurately as piano. Humans organize objects taxonomically, and perceived timbres are organized with taxonomical hierarchies, with help of identities of sounds. (Jensen 1999a, Jensen 1999b). Considering this, timbre may be the main dimension to create distinguishable earcon families.

4.3 Masking effect

Relating again to usability of auditory display, one possible error that may occur in use, is that auditory sign becomes inaudible by some other sound, whether it is sound from the interface,

or from the surrounding environment. Threshold for audible sound can be raised by other sound or noise presented simultaneously. Sensation of current sound can be obscured or even become inaudible by the other sound. This phenomenon can be occurred for example while discussing in a car. It is necessary to raise one's voice in order to be heard because audible threshold of voice is raised by noise created by car. This process is called masking. The more complex the spectral structure of sound is, the more unlikely it is masked by other sounds.

The idea of *critical bands*, introduced by Fletcher, explains the masking of a narrow band (sinusoidal) signal by a wideband noise source. (Fig 9)

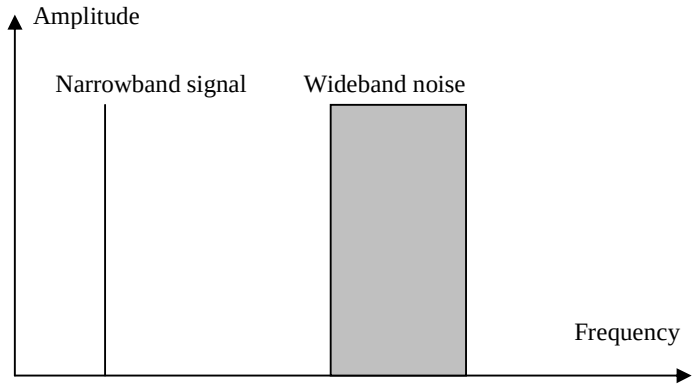

Fig 9

In this figure, the noise signal is remote from the frequency of the sinusoidal signal. In this situation, the wideband noise presented does not change the threshold of hearing of the narrowband signal. Figure 10 shows the noise signal surrounding on the sinusoidal narrowband sound. (Zwicker and Fastl 1990)

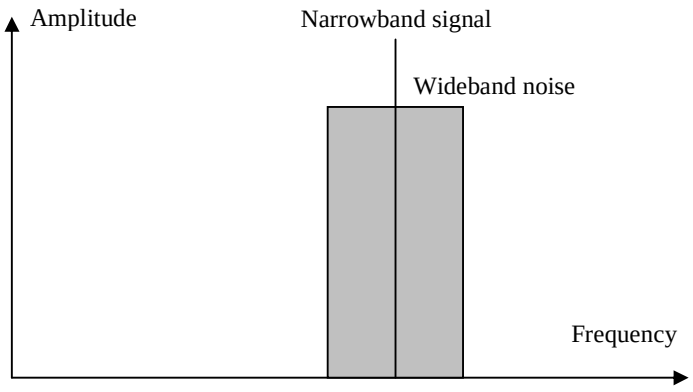

Fig 10

With the noise surrounding the frequency of the narrowband signal, the threshold of hearing of the narrowband signal is greater than before. The noise signal is therefore masking the narrow band tone at amplitude levels between the old threshold and the new. As the noises bandwidth is increased, the threshold of the sine tone will raise. However, there exists a point when an increase in noise bandwidth gives no increase in the threshold. (Zwicker and Fastl 1990)

Masking effect influences heavily when masker's frequency is close to masked ones frequency. To avoid masking of a certain sound, it should have more than one component frequency (single sinusoid). Timbre refers to this quality; different timbres mean differences in frequency spectrum of their harmonics. To avoid possible masking of auditory signs in user interface it is not recommended to use pure tones. To avoid masking and to help perception it is necessary to use timbres with harmonies if it is considered that environment where auditory interface is used might be exposed to noise or other sounds. (Zwicker and Fastl 1990)

This aspect has been considered to some extent in earlier earcon design, such as in (Brewster et al. 1995). However, the human perception of sound is more sophisticated than that. Although some sounds are close to each other, and may be considered that one just simply masks another inaudible, it may not be the case.

5 AUDITORY SCENE ANALYSIS

Auditory scene analysis is a process that addresses how human auditory system separates complex waveforms into meaningful representations. Usually we are interested only in one sound source, such as someone talking, but soundscape includes multiple other sound sources and streams. When hearing music, one is able to focus on one instrument at time even tough multiple instruments are making sounds and all of them reaches ear simultaneously. Earlier earcon studies argue only earcon identification when compared to other earcons. Guidelines suggest that timbre used in earcons should be instrumental timbres as different from each other as possible. The human auditory system is very accurate on distinguishing sound sources form each other from very slight cues, and rarely makes errors even though sounds presented are very similar to each other. (Jensen 1999b)

Auditory scene analysis studies these processes how and why we can concentrate only on one sound source when there is simultaneous flood of different sounds. Auditory scene analysis has roots in gestalt theories. It is relevant to consider these studies of grouping sounds in some categories in order to understand how earcons are to be designed. Auditory scene analysis states that sounds are grouped by similarities or dissimilarities along the different dimensions. Of all the sounds we hear (meaning everyday soundscape in all situations we face) are likely grouped those that are close to each other, similar to each other, follows each other, associated to each other or are dependent on each other. The more similarities with dimensions of sounds, the more likely they are grouped into one sound stream or source, and vice versa. (Yost 1997, Bregman 1990)

The field of auditory display research has not taken almost any account of auditory scene analysis effect on the earcon design process. Because human auditory system can identify sources from very fine cues, it is not necessary to select perceptually far distant timbres; slighter differences may be effective.

5.1 Selecting one sound from the complex soundscape

Nothing we hear is alone in the soundscape. Everything in our environment is making sounds. In an ordinary room, we can usually hear sounds of air conditioning, someone talking outside the room and distant car engine roaring. In a concert, we cannot hear violin, trombone, cello and drums. We hear a very complex sound that is a mixture, a sound stream, of all these sounds, echoes and noises.

Human auditory system is capable of parsing a soundscape to form a mental representation of each sound stream (or source) by mental auditory scene analysis. There are two conceptual processes in auditory scene analysis. First is segmentation, which is a process where auditory system decomposes the soundscape into sensory elements, segments. The second is grouping. This process combines segments into streams, combining streams that are originally from the same source. (Bregman 1990, 47-184)

Segregation process of pitch and loudness is relatively easier to examine than timbre. Timbre is more complex to explore due to its multidimensional nature. Acoustic correlates of pitch and loudness are well known and well defined whereas acoustic correlates of timbre are more uncertain. Timbre structures that are useful for the segregation are spectral separation, spectral profile, harmonicity, spatial separation, temporal separation (which consists of temporal onsets/offsets and temporal modulations, fluctuations over time, and time varying patterns).

Grouping involves two aspects: primitive grouping and schema-driven grouping. Primitive grouping refers to intuitive grouping, such as similarity, proximity, harmonicity, onset-time and continuation of the sound streams. Primitive grouping is a “bottom-up” process that is based on the simple perceptual acoustic properties. Schema driven process is a “top-down” process that acts on output of primitive grouping stage. Schema driven process relies on the learned knowledge of source models, which can become better and effective by learning. (Bregman 1990, 411) An example of the schema-driven sound stream organization is the organization process of speech into syllables, or musical structure organization, such as instrument sorting. Schema driven organization is more complex to define, due to the learned models.

Cocktail party problem refers to a phenomenon that is a magnificent illustration of human perception accuracy, which can be explained by the schema-driven segregation. We are able to listen to, and follow one certain speaker in the crowd when there are several other speakers speaking simultaneously. In the situations where all the voices and sounds are equally loud, speech of a certain speaker remains audible and understandable for listener. (Yost 1997, Haykin and Chen 2005, Bregman 1990) The cocktail party problem was first introduced in 1953, and it was argued that spectral separation, profile and harmonicity are on the important role in sound source determination. (Yost 1997)

What are the cues for the primitive grouping organization? Features that have influence on grouping of sounds are proximity in frequency, common periodicity, common spatial location (which is perceived by the changes it causes to spectral structure by echoes etc.) common onset and offset, and common temporal modulation, which refers to time varying patterns of timbre, introduced earlier. (Moore 2001, 249-262, Bregman 1990, 478-489)

Schouten has proposed that recognition of auditory object may include several factors:

- Whether sound is periodic, having a tonal quality for repetition rate between about 20 Hz to 20 KHz, or it has more noise like characteristics.
- Whether the waveform envelope is constant or fluctuates as a function of time (spectral envelope) (This refers to onset, steady-state and decay structure)
- If sound fluctuates, what kind of fluctuation it is.
- Whether any aspect of the sound is changing as a function of time. (The change of both of spectral envelope (formant-glide) and fundamental frequency (micro-intonation))
- What the preceding and following sounds are like.

Van Noorden stated in his studies that sound streams with frequencies close to each other were perceived as one sound. The more alternating frequencies were, the more likely they were perceived as two unique tones. This refers widely to the timbre characteristics, and may explain why different instruments are recognized in concert. (Yost 1997) Due to the differences in musicians playing styles, differences in instruments and spatial locations of the musicians, timbre characteristics are varying *enough* from each other in order to be perceived

as different object. As identity of the sound source, timbre is the key dimension in recognition and selection of sound sources.

6 TIMBRE, EMOTIONS AND MEANING

Huron's theory suggests that decoding information that timbre includes would give us information about the producer of the sound, its intentions and its emotional states. (Padova et al. 2005) How are music-induced emotions relevant to earcon design? Consisting the definition of Kleinginna & Kleinginna emotion is (Juslin and Sloboda 2004, 75):

> Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as perceptually relevant effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-directed, and adaptive.

Important to the earcon design topic is the part d, emotion as a factor to lead to goal-directed behaviour. In interface, it is important to convey meanings and emotions that lead user to operate in a desired way. In the terms of usability (Nielsen 1993), using behaviour directing emotional cues in earcons, it improves intuitiveness of use, as well as efficiency of use. As Bussemakers & de Haan (2000) proved in their study, if played earcon does not correspond to task (it is audio-visually dissonant (Chion 1994)) it may lead earcon to be auditory distracter, and hinder the task performance. Emotions discussed in music and emotion research are not equally corresponding to events, action or objects that may be represented with earcons (Nielzén & Cesarec 1981). However, as discussed earlier, incongruent and congruent connotations are important aspect to consider, in order to create intuitive earcons. Approach is to some extent corresponding to semantic sciences that distinguish between denotative and connotative meanings.

6.1 Concept of emotion

There are two parties arguing of the music's influence on listeners' emotions, cognitivist and emotivists. Cognitivists argue that music's emotions do not affect listeners emotional states, but emotions of the music can be recognized. Emotivists on the other hand contradict that emotions of music have influence on listeners' emotional states. Truth may lie somewhere in the middle (Lucassen 2006). Feeling evoked by music may lead to certain behaviour, even though emotion of the music does not affect the listeners emotion.

An important aspect to consider is that music can be emotionally charged by earlier experiences, as well instrument's timbre. Once we have learned to map some instruments in some emotions, it may be hard to separate whether emotion is induced or recognized; is it due to the spectral characteristics or are we used to think that some instrument is e.g. sad.

Major approaches to music emotion research are categorical approach, dimensional approach and prototype approach. In categorical approach experience of the emotions of the music is categorical, all the categories distinctive from each other. In this approach, there are a limited number of basic emotions. Each main emotion is defined functionally in terms of a key appraisal of goal-relevant events. There are some different views of the sets of basic emotions. According to Oatley, these are happiness, anger, sadness, fear, and disgust. (Table 1)

Table 1. *Key appraisals for basic emotions taken from Juslin & Sloboda 2004, 76)*

Emotion	**Juncture of plan**	**Core relational theme**
Happiness	Subgoals being achieved	Making reasonable progress towards a goal
Anger	Active plan frustrated	A demeaning offence against me and mine
Sadness	Failure of major plan or loss of active goal	Having experienced an irrevocable loss
Fear	Self preservation goal threatened or goal conflict	Facing an immediate, concrete, or overwhelming physical danger
Disgust	Gustatory goal violated	Taking in or being close to an indigestible object or idea.

While categorical approach emphasizes the characteristics that distinguish different emotions, dimensional approach emphasizes the identification of the emotions based on their placement in the field of dimensions such as *valence*, *activity*, *potency*. Russell has brought up circumplex model. This two dimensional model includes dimensions activation and valence (Fig. 11). (Juslin & Sloboda 2004)

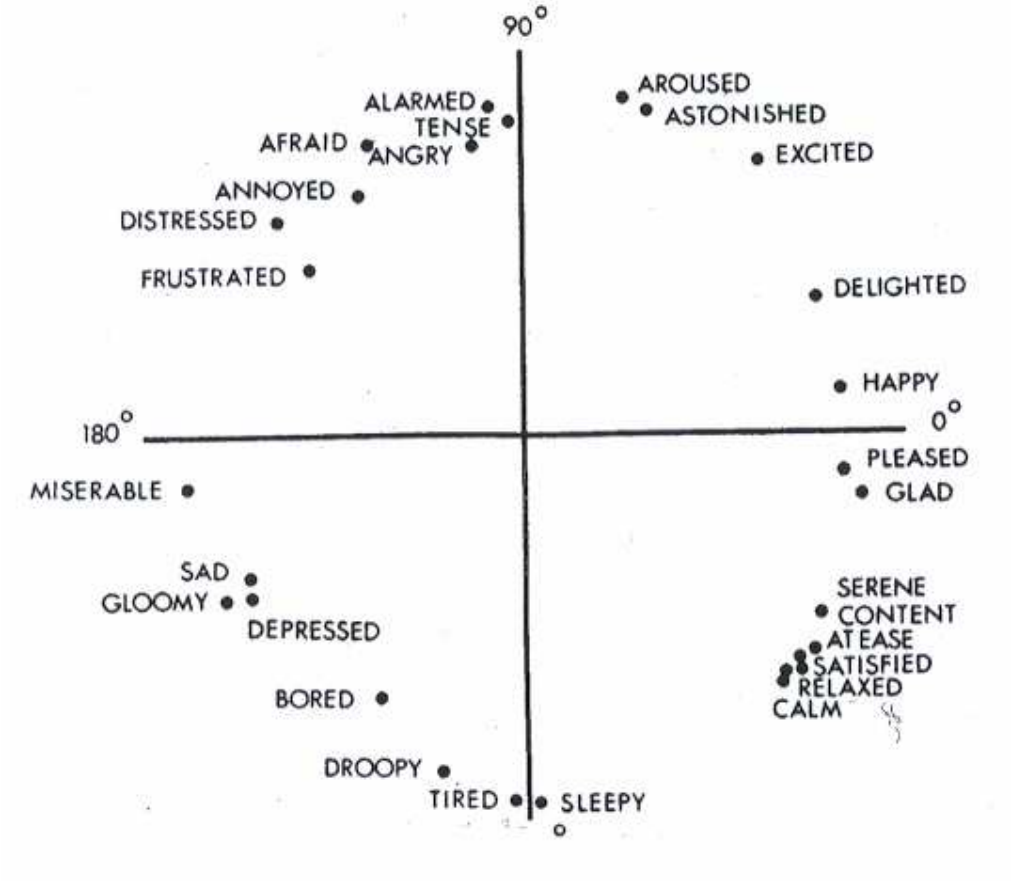

Figure 11. Within the structure in this approach, emotions that are across one another, are bipolar and have inverse correlations, such as sadness and happiness. (Juslin & Sloboda (2004) , 78)

Prototype approach to emotions is based on the idea that human perception conceptualizes and categorizes information by the shape that is given by language and knowledge structures associated with language, that is: membership in a particular category is determined by resemblance to prototypical exemplars. Prototype is an abstract image that consists of a set of entities. Some of the entities match better to prototype image than others, such as frustration is a better example of anger than jealousy.

This approach has, to some extent, features of the categorical and dimensional approach. It emphasizes the individual categories and the hierarchical relation among the categories. (Example in figure 12) In the figure, vertical dimension shows the hierarchical relations among the categories. The uppermost level is super ordinate level, which is defined by the

positive or negative valence emotions have. Middle level represents prototypes, basic-level categories, of emotions. Lowest, the subordinate category, consists of all the emotions related to the basic-level categories.

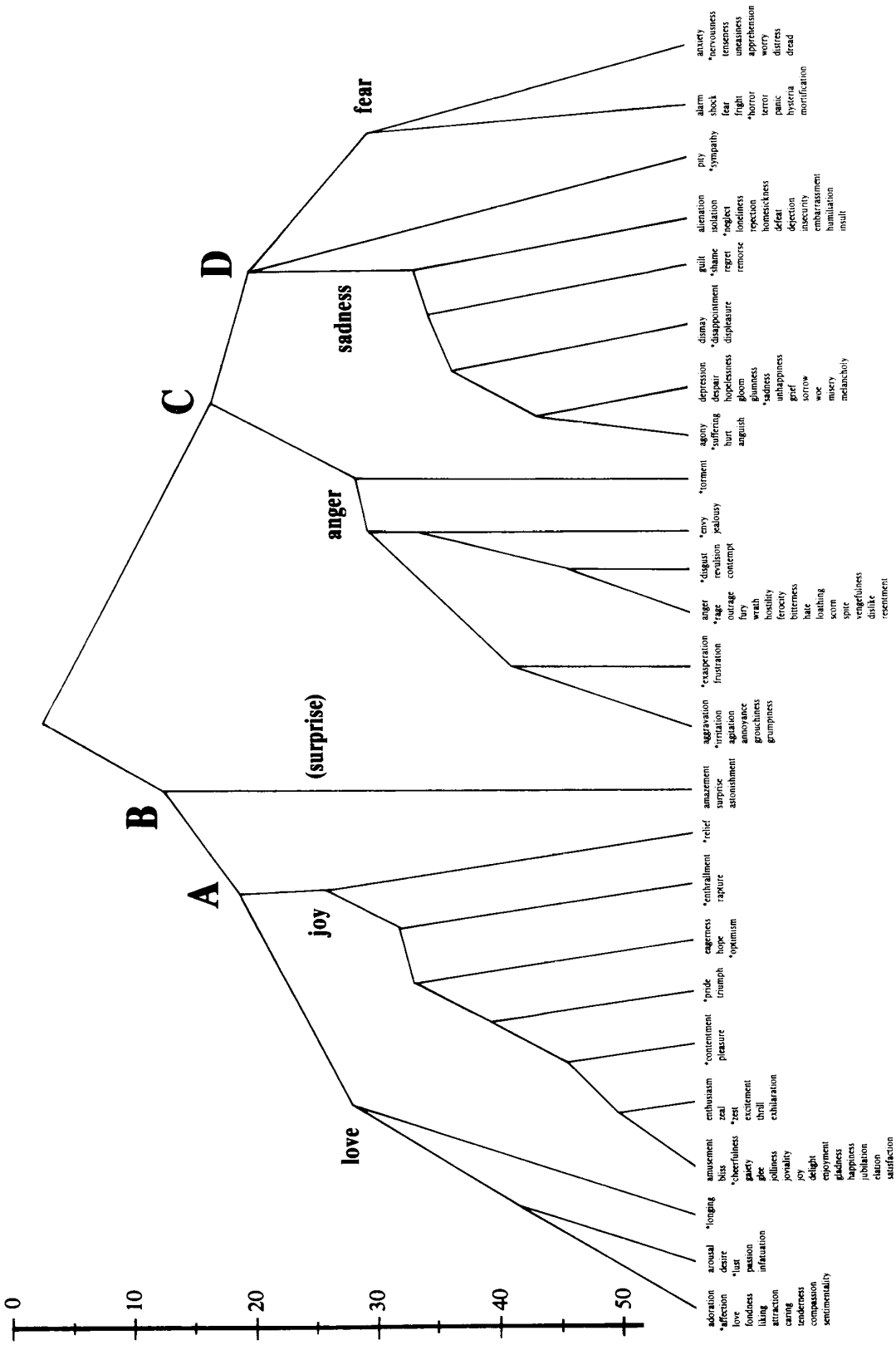

Fig 12. *Prototype analysis of emotions and words. (Juslin & Sloboda 2004, 80)*

6.2 Emotions evoked by instrument timbres

Huron's theory suggests that decoding information that timbre carries would give us information about the producer of the sound, its *intentions* and its *emotional* states. (Padova et al. 2005) There has been little research of emotions and meanings evoked by musical timbre. Two approaches are relevant to this thesis. Such as the semantics distinguishes denotative and connotative meanings, similarly has to be considered the music and timbre, in order to improve usability of auditory signs. Connotative meanings of instruments, and emotions of instruments' behaviour-imitating playing style, which means that instrument is played so that the meanings the instrument conveys is somewhat mimetic to human behaviour. A certain aspect of emotional gesture is cross-cultural, and another is culture-related, such as shouting, talking with tender voice etc.

Timbre plays the main role in musical expression and meaning recognition. Shouting or angrily played instrument is perceived as angry if sound intensity is lowered. It becomes quieter, but neither neutral nor happy; the meaning remains the same. Perception of the meaning is formed by the relative intensities of the higher partials in comparison to the fundamental tone, and timbre's noise-like characters. Angrily played instrument spectral envelope has certain mimetic similarities with shouting voice timbre. Expressive meaning of instrumental music depends widely on the similar perceptions of mimetic spectral behaviour, such as shout has fast onset decay and noise-like structures, and so has 'angry' instrument. (Lucassen 2006, Zagorski-Thomas 2005)

Padova et al. (2003) demonstrated in their study some spectral variations that affect the instrumental timbre variation on the evoked emotions. Changes of harmonic dynamic and harmonic ratios aroused negative meanings. Changes of distributions of spectral energy aroused happiness. They stated out that harmonic components and the spectral energy are fundamental factors in the task of perceptive judgment as of emotive response, in proportion to a study by Zagorski-Thomas (2005) introduced above. This proved that perception of the meaning is formed by the relative intensities of the higher partials in comparison with the fundamental tone, and timbre's noise-like characters.

Several studies and researchers demonstrate that music has clear and strong emotional content and affect. (Lucassen 2006, Sloboda 2005, Padova et al. 2005, Juslin and Sloboda 2004, Zagorski-Thomas 2005, Cook 2001) How much does timbre have effect on this? There is some intuitive links between instruments and emotions. We are used to hear cello in sad, slow music, whereas marimba is connected with happier and faster music. Is the learned musical context (earlier experiences) that evokes the happiness or sadness induced by these instruments, or is it instrumental timbre itself?

How much is music itself affecting emotional perception? There are not many studies of the instrument-evoked emotions. Lucassen (2006) conducted experiment with four instruments and concluded that there are some emotional meanings with instruments themselves, which are independent from the musical context. Piano proved emotionally neutral, marimba was considered very joyful, cello invoked sad emotions, and alt saxophone did evoke both, happy and sad emotions, which is, according to Lucassen, considered as consolation effect. Although there was only few instruments in this experiment, it clearly demonstrates that some emotions are intuitively linked to some certain instruments, and that should be taken into account when creating auditory signs.

According to studies of Lucassen (2006), Padova et al. 2003 and Zagorski-Thomas (2005) instrument's timbre itself can evoke meanings and emotions effectively. One can use emotionally loaded musical timbre, such as cello, to derive negative and sad meanings, or use "sadly" played instrument, that is mimetic to emotional behaviour. This aspect is not considered extensively in previous public earcon studies. By this kind of timbre usage, it may be possible to avoid problems introduced by Lemmens et al. (2000), Bussemakers, and de Haan (2000). In addition, it is possible to improve motivation of the sign, which will also improve learning, identification and recognition of the auditory sign, and events or objects it represents; in other words, it is the way to improve also intuitiveness of the earcon, thus also the usability.

7 CONCLUSIONS AND DISCUSSION

Semiotic analysis of earcon and their definition has exposed harsh imperfections of present day earcon definition. Strict separation to abstract or representational signs has left out the important combinatory level of these signs. By the semiotic analysis the contemporary earcon and auditory icon has proven to be unusable and erroneous. Original earcon definition by Blattner, Sumikawa et al. included all these different levels of auditory signs. This definition is suggested to be used again. According to this definition, earcons encompass all the non-speech sounds used in the interface, and they can be representational, abstract or semi-abstract earcons. Semi-abstract earcons can be defined as abstract earcons that have strong motivation, or in other words, they are intuitive.

The auditory display researchers have acknowledged the lack of decent guidelines for interface sound designers. However, the provided guidelines for earcon designers do not offer enough aid to designers, because the sign design issues are not considered. The connotations and the context are not considered either in the auditory sign design guidelines. As an analogy to visual sign design, if similar guidelines were followed, visual icons would be photographs of the entities they represent, or abstract basic geometric shapes which have to be learned. No consideration is given to analysis of context, if the sign is coherent with the context, or with the entity it represents.

When connotations or context are not considered, and if connotation is dissonant with context, earcons may become disadvantageous, unusable and even performance hindering distracters. As referred earlier, usability of the systems consists of five things. There are a few aspects to consider when evaluating systems usability; learnability, efficiency of use (intuitiveness), memorability, errors and subjective satisfaction. (Nielsen 1993) Earcon design guidelines do not give these aspects much attention.

An underestimated dimension of sound, timbre, is demonstrated in this thesis. According to timbre research, timbre is recognized and memorized effectively due to the nature of timbre why it is also referred as identity of the sound. As human perception memorizes timbre

categorically, it is more effective to use timbre as the key dimension of earcon before rhythm, pitch or melody, in order to enhance memory and identification of earcon families and thus to improve the learnability of auditory signs. This will help users to learn the meaning of the sign more easily and faster, and enables them to remember and recognize the earcon efficiently after a while from hearing the interface.

The syntagmatic analysis and presented audiovisual contract consider the meaningful whole. Of all the potentially usable sounds, only some are corresponding to the entity it is representing, and coherent with context. This should be considered also in earcon research and design. Timbre is the key dimension in this approach too, bearing in mind the nature of timbre being the colour of the sound, and being the key dimension in providing information of the intentions and emotional states of sound source. The emotion and intention conveying characteristics of timbre is important to consider due to their strong connotations.

If comparing colours and visual connotations used in visual signs, they are studied widely. Error messages and warning signs are usually red, due to the cultural stressing of red to mean stop. Blue is used to neutral meanings, yellow and red together are used commonly in warning signs and attention grabbing. None of similar aspects has been considered earlier in earcon design.

Some instruments have strong connotations to some emotions. Some instrumental timbres may be played by reminding some certain emotion, due to the imitation of human behavioural characteristic. The midi technique recommended to use in earcon design is ‘lifeless’ compared with natural timbres. Timbre characteristics are difficult to synthesize and e.g. emotional behaviour imitations, such as angrily struck instrument is almost impossible to synthesize. According to this, synthesized instruments such as MIDI are not suggested to be used if intuitive earcons are desired by the timbre’s features that imitate human behavioural gestures.

When these kinds of connotations are taken into account in timbre research and design, the usability of auditory signs will improve, and it can prevent errors. Timbre is also the one most important factor in subjective satisfaction to sound. In order to create appropriate syntagm of auditory display, timbre is an important feature to take into account, due to its important role in identity, meaning and aesthetics of sound. Although aspects of aesthetics

were not considered extensively in this thesis, except in the context of syntagm, aesthetics play an important role in auditory display field, according to hearing senses sensitiveness, and the introduced listening modes.

Instrumental timbre evoked emotions have not been studied extensively, but from the existing studies can be drawn some examples of different instruments being mapped to different emotions, so there may be some basic timbre characteristics affecting the emotion of the instrument timbre. In addition, human behavioural gestures imitative timbre characteristics are an effective mean to convey emotions or meanings. As considered in table 2, it is clear that some emotions can be used as useful connotations for the earcons. For example, the emotions related to happiness may be used as connotative counterpart for auditory messages like ‘success’, ‘proceed’ or ‘task performed’. Sadness, as opposing emotion for the happiness may be used to convey messages opposite, ‘unsuccessful’, ‘task not performed’ and anger and fear may be used as effective warning signs.

Table 2 *Basic emotions for meaningful earcons, examples derived from the key appraisals for basic emotions by Juslin and Sloboda (2004, 76)*

Base emotion	**Example meaning of auditory sign**	**Core relational theme**
Happiness	Success, task performed successfully, file loaded, file found	Making reasonable progress towards a goal
Anger	Error occurred and needs action immediately	A demeaning offence against me and mine
Sadness	Error occurred, does not need actions, task not performed successfully, file not found	Having experienced an irrevocable loss
Fear	Possible threat to user/device	Facing an immediate, concrete, or overwhelming physical danger
Disgust	Unwanted actions occur, warning of possible unwanted event	Taking in or being close to an indigestible object or idea.

As circumplex model of emotions demonstrates, some emotions are opponent, bipolar to each other, and some emotions are closely related, such as angry, tense and alarmed are very similar. Same results can be concluded from the prototype approach to emotions. Some emotions are more corresponding to the prototype of emotions, such as amusement is more corresponding to joy than relief, but still they both are corresponding to it, whereas amazement is to some extent incongruent with joy, even though it is closer to joy than to anger.

Emotions, which are similar, that is to say, similar with connotations, may be used in a similar way. Because there are no extensive research results of all the timbre characteristics that are affecting perceived emotions, conveyed or evoked by the instrumental timbre, timbres used in interface should be analyzed by subjective judgements in order to find their connotations. As in visual design, artist should be at least consulted when creating signs, and subjective evaluation should be conducted. There is no reason why sound design should be any different in that aspect. Due to the hearing senses direct affection to emotions and feelings, professionals should be in even more important role than in visual design.

The most important aspect is to consider that connotations of the timbre cannot be incongruent with the entity it represents. As Chion (1994) argued, there may be several corresponding connotations with e.g. image and sound, several sounds are matching with the image, and it may not be possible to determine best and second best sound to use. Nevertheless, there may be similarly vast array of sounds that do not correspond with the context, and are dissonant auditory distracters.

In view of auditory scene analysis and auditory stream segregation, it is not necessary to select timbres that are widely differing from each other, as Brewster et al. (2007) suggested. Human perception can distinguish the sound stream by very slight cues, and after a while of experience, this ability can grow even better. Whereas contemporary earcon design guidelines rely on learning of very abstract sounds, and does not trust on the accuracy of identification of sound sources and segregation of sound streams, according to literature review, it should be

the opposite. Providing intuitive meanings of earcons would prevent errors, besides of discussing the use of clearly different MIDI instruments, without considering the question “why this timbre”.

Although this thesis has mostly considered instrumental timbres, it should not be the only consideration in the future. Artificial timbres may be as effective as instrumental timbres, due to the efficiency of the schema-driven segregation of sound stream. However, the reason why artificial timbres have not been considered here is for the lack of proper analysis of timbre features that have effect on emotions and connotations of timbre.

Nevertheless, this kind of artificial sounds can be used and designed effectively if utilizing for example panel method introduced by Murphy et al. (2006) where it is possible to evaluate meanings and usability of the sounds by others than the sound designer. However, synthesized instruments are not recommended, if it is not possible to clearly evaluate that sound used is not lifeless, irritating and unusable. Whereas visual user interfaces are not constructed of web-safe colours and clip-art images, sound interfaces should not be constructed from the lifeless MIDI sounds, but designed to match the need and context.

8 SUMMARY

In this content analytical literature review I have studied non-speech auditory signs, earcons, and their application to problems not considered in present day auditory display research. Contemporary definition of earcons has been demonstrated to be to some extent flawed and unsuitable.

According to semantics, distinguishing the abstract and representational signs is not easy, and in most of the cases it may be even impossible. Connotations are important part of sign design. This aspect has not been considered extensively in present day earcon design, because the present day earcon definition excludes all forms of sign creation that are not abstract. That absurd discrimination has led to some problems where earcons can become dissonant and distractive.

The first earcon definition by Blattner et al. (1989) is suggested to be taken into account again, because the present day earcon definition excludes the motivation and intuition of earcons. Almost no other sign we encounter in everyday life is purely abstract or purely representational, but auditory research field has separated different types on these grounds, and that has excluded signs that have some levels of representational or abstract signification.

The context of use and the dissonance of the sounds and entities they represent have almost never been considered in present day earcon design. In order to avoid some serious problems in auditory signs, the suitability of the sound in the meaningful whole has to be considered thoroughly.

Timbre, dimension of sound that is referred usually as identity, quality or colour of the sound is the most important dimension of sound to be taken into pinpoint in design of earcons. This dimension of sound has been poorly understood and considered in present day earcon studies. A reason for the exclusion of timbre from the earcon research may be due to the complex

nature of timbre. This should not be the reason to exclude it, because the literature of the psychoacoustics, auditory stream segregation and musicology has studied timbre extensively.

The literature review demonstrates that timbre has strong effect on connotations in emotions and it should be considered more accurately in studies and design of usable auditory signs, and in creation of a meaningful whole of the interface and the context. With careful timbre-design, sounds used in interface become more intuitive and it is also potentially a good mean to improve memorability and assure the audiovisual harmony and at the same time prevent errors caused by audio-visually dissonant sounds.

9 REFERENCES

Blattner, M., Sumikawa, D. & Greenberg, R. (1989) Earcons and Icons, Their Structure and Common Design principles In Human-Computer Interaction, Vol. 4: 11-44

Bregman, A. S. (1990) Auditory Scene Analysis, The Perceptual Organization of Sound, The MIT press, Massachusetts

Brewster, S. (1994) Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD Thesis, University of York, UK

Brewster, S. (1995) Parallel earcons, Reducing the length of audio messages. International Journal of Human-Computer Studies, 43: 153-175

Brewster, S., Wright, P. & Edwards, A. (1993) An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. Inter CHI: 24-29

Brewster, S., Wright, P. & Edwards, A. (1994) In Human Factors in Computing Science Boston, Massachusetts, United States

Brewster, S., Wright, P. & Edwards, A. (1995), Experimentally derived Guidelines for the Creation of Earcons, in Human Computer Interaction: 155-159

Brewster, S., Wright, P. & Edwards, A. (2007) [Online], Guidelines for the Creation of Earcons University of Glasgow, cited 1.5.2007, available online <URL:http://www.dcs.gla.ac.uk/~stephen/earcon_guidelines.shtml>

Bussemakers, M. P. & de Haan, A. (2000) When it Sounds like a Duck and Looks like a Dog... Auditory Icons vs. Multimedia environments, in Proceedings ICAD 2000: 184-189

Chion, M. (1994) Audio-Vision, Sound on Screen, Columbia University Press, New York

Cobley, P & Janzs, L. (1998) Semiotiikkaa vasta-alkaville ja edistyneille, Jalava, Helsinki

Cook, N. (2001) Theorizing Musical Meaning. Music Theory Spectrum, Fall 2001, Vol. 23, No. 2: 170-195

D'Incá, G. & Mion, L. (2006) Expressive Audio Synthesis: From Performances to Sounds (2005) Proceedings. of the 12th International Conferecne on Auditory Display, London, June 2006:

Deliege, I. & Sloboda, J. (1997) Perception and Cognition of Music, Psychology Press, New York, NY

Deutch, D. (1980) The Processing of structured and unstructured tonal sequences. Perception and Psychophysics, 28: 381-389

Fiske, J. (2000) Merkkien Kieli, Vastapaino, Jyväskylä.

Gaver, W., Smith, R. & O'Shea, T. (1991) Effective Sounds in Complex Systems: the ARKOLA simulation, In Proceedings of Human Factors in Computing Systems (CHI '91) ACM Press, New York: 85-90

Gaver, W. W. (1986) Auditory Icons: Using Sound in Computer Interfaces. Human-Computer Interaction, 2(2)

Gaver, W. W. (1989) The Sonic Finder: An Interface that Uses Auditory Icons. Human-Computer Interaction, 4(1): 67-94

Grey, J. M. (1977) Multidimensional Perceptual Scaling of Musical Timbres. Journal of Acoustical Society of America, 61: 1270-1277

Handel, S. (1989) Listening, An introduction to the Perception of Auditory Events, MIT Press

Hankinson, J. & Edwards, A. (1999) Designing Earcons with Musical Grammars. ACM SIGCAPH Computers and the Physically Handicapped: 16 - 20

Haykin, S. & Chen, Z. (2005) The Cocktail Party Problem. Neural Computation, 17(9): 1875 - 1902

Houser, N. (1992) In The Essential Peirce (Eds, Houser, N. and Kloesel, C.) Indiana University Press, Bloomington, Indiana

Hyrsykari, A. (2001) Lukuprosessin silmänliikkeisiin perustuvat mallit In Käyttöliittymäteoriat ja mallit (Ed, Raisamo, R.) Tampereen yliopisto, Tietojenkäsittelytieteiden laitos, Raportti B-2001-7, Tampere: 49-82

Iverson, P. (1995) Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. Journal of experimental Psychology: Human Perception and Performance, 21(4): 751-762

Jensen, K. (1999a) Hybrid Perception, Papers from the 1st Seminar on Auditory Models, Lyngby, Denmark

Jensen, K. (1999b) Timbre Models of Musical Sounds, Ph.D. dissertation, Department of Computer Science, University of Copenhagen, 1999 Report no. 99/7

Jensen, K. (2001) The Timbre Model In Workshop on current research directions in computer music: 174-186

Jensen, K. (2002) Perceptual and Physical Aspects of Musical Sounds. Journal of Sangeet Research Academy: 1-22

Juslin, P. N. & Sloboda, J. A. (2004) Music And Emotion, Theory and Reseach, Oxford University Press

Lemmens, P., Bussemakers M.& de Haan, A., The Effect of Earcons on Reaction Times and Error-Rates in a Dual-Task vs. a Single-Task Experiment, In *Proceedings ICAD 2000*, Atlanta: 177-183

Lucassen, T. (2006) Emotions of Musical Instruments In 4th Twente Student Conference on IT

McGookin, D. & Brewster, s. (2004) Understanding Concurrent Earcons: Applying auditory scene analysis principles to concurrent earcon recognition. ACM Transactions on Applied Perception (TAP), 1(2): 130-155

McGookin, D. K. (2004) Understanding and Improving the Identification of Concurrently Presented Earcons , PhD, University of Glasgow, 2004

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., Glover, G. H. & McAdams, S. (2002) Neural Correlates of timbre change in Harmonic sounds In Neuroimage, Vol. 17: 1742-175

Moore, B. C. J. (2001) An Introduction to the Psychology of Hearing, Academic Press, California

Murphy, E., Pirhonen, A., McAllister, G. & Yu, W. (2006) A Semiotic Approach to the Design of Non-Speech Sounds In HAID: 121-132

Nielsen, J. (1993) Usability engineering, Academic Press, California, US

Nielzén, S. & Cesarec, Z. (1981) On the perception of emotional meaning in music. Psychology of Music, 9: 17-31

Norman, D. A. (1988) The Psychology of Everyday Things, Basic Books, New York

Norman, D.A. (2001) The Design of Everyday Things, MIT Press, U.S.

Oxford Advanced Learner's Dictionary (2005) Oxford University Press

Padova, A., Bianchini, L., Lupone, M. & Belardinelli, M. i. (2003) Influence of Specific Spectral Variations Of Musical Timbre on Emotions in the Listener, In 5th Triennial ESCOM Conference Hanover, Germany

Padova, A., Santoboni, R. & Belardinelli, M. (2005) Influence of Timbre on Emotions and Recognition Memory for Music In Proceedings of the Conference on Interdiscplinary Musicology (CIM05) Montreal (Quebec) Canada

Pirhonen, A. (2005), To simulate or to stimulate? In search of the power of metaphor in design In Future Interaction Design (Eds, Pirhonen, A., Saariluoma, P., Isomäki, H. and Roast, C.), Springer London: 105-123

Pirhonen, A.Murphy, E., McAllister, G. & Yu W., (2006) Non-Speech Sounds as Elements of a Use Scenario: A Semiotic Perspective In Proceedings of the 12th International Conference on Auditory Display, 2006 London, UK: 121-132

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieilleard, S. & McAdams, S. (2004) Effects of a Change in Instrumentation on the Recognition of Musical Materials. Music Perception, 22 (2): 239-263

Preece, J. (1994) Human Computer Interaction, Addison-Wesley, Wokingham

Radvansky, G. A. & Potter, J. S. (2000) Source cueing: A case involving memory for melodies. Memory & Cognition, 28: 693-699

Richards, A. M. (1976) Basic Experimentation in Psychoacoustics, University Park Press, Baltimore

Saussure, F. (1983) Course in general linguistics, Duckworth, London

Schafer, M. R. (1994) Our Sonic Environment and The Soundscape, the Tuning of the World, Destiny Books, Rochester, Vermont

Sloboda, J. (2005) Exploring the Musical Mind: Cognition, Emotion, Ability, Function, Oxford University Press, Oxford ; New York

Tarasti, E, (1996), Johdatusta Semiotiikkaan, Esseitä taiteen ja kulttuurin merkkijärjestelmistä, Oy Gaudeamus Ab, Helsinki

Thayer, R. (1974) The Effect of the Attack Transient on Aural Recognition of Instrumental Timbres. Psychology of Music, 2

Walker, B. N., Nance, A. G 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

Warren, R. M. (1999) Auditory Perception, Cambridge University Press, New York.

Yankelovich, N., Levow, G.-A. & Marx, M. (1995) Designing SpeechActs: Issues in Speech User Interfaces In CHI ‘95 Conference on Human Factors in Computing Systems, Denver, CO: 369-376

Yost, W. (1997) In Binaural and spatial hearing in real and virtual environments (Eds, Gilkey, R. and Anderson, T.) Erlbaum, Ahwah: 329-347

Zagorski-Thomas, S. Shouting quietly: Changing Musical Meaning by Changing Timbre with Recording technology (2005) In Second Conference on Interdisciplinary Musicology CIM05 Montreal, Canada

Zwicker, E. & Fastl, H. (1990) Psychoacoustics, Springer-Verlag, Berlin-Heidelberg-New York