A review-based conceptual analysis of auditory signs and their design

Manne-Sakari Mustonen

Department of Computer Science and Information Systems,
P.o. Box 35, FI-40014 University of Jyväskylä, Finland
msmuston@jyu.fi

ABSTRACT

The research frames of auditory display have traditionally mainly focused on the evaluation of different applications and devices, whereas the theoretical development has had a minor role. In order to reach the goal of functional and intuitive auditory signs, the theoretical basis must be on a robust basis. User interface sound types have been traditionally divided into two exclusionary sound types: earcons and auditory icons. However, when approaching the issues from the viewpoints of for example human communication or semiotics, one can see that the current definitions and practices in auditory display as a scientific discipline are not pragmatic. It is recommended to define auditory signs to include different levels of meaning, as was originally proposed. Following current theoretical concepts leaves the full potential of auditory signs unexposed. In this paper, I introduce important viewpoints and approaches for more practical theoretical approaches for the design of auditory signs in order to develop a theoretical basis for usable syntax.

1. INTRODUCTION

Due to the changing and actively developing interaction styles with devices, the visual display cannot be seen as the one and only modality in interfaces. When devices go mobile and ubiquitous, growing interest for using supporting modalities for the traditional visual becomes evident. Besides mainstream applications, it is important to develop auditory support for the use of interfaces for visually impaired users. Ubiquitous computing in everyday life yields several challenges for multimodal interface design.

Auditory displays (AD) have been studied for a relatively long time. The concept of audio interfaces has been scientifically tested and evaluated in various applications, such as auditory graphs, hierarchical auditory menus and navigation for visually impaired users, alarm sounds, control room environments (such as cockpits and monitoring in factories), audio only games and several other interactive interfaces [1].

Providing nonverbal cues is seen to have an important role in supporting interaction with interfaces. In some cases (such as hospital patient monitoring systems etc.), sound conveyed meanings should be as accurate as possible, as false messages/alarms or annoying sounds can possibly hinder important tasks [2]. Commonly used strategy in AD research are using either symbolic or iconic cues to provide the necessary information.

Most of the AD studies conducted are case studies. Typically, in the research frame the symbolic and iconic types of auditory signs are seen as contradictory "either-or" approaches – very few attempts have been made to evaluate the most common generalizable pros and cons of different designs in pragmatic use but the focus is on comparison of these different approaches [3]. Furthermore, current auditory sign definitions and theories leave some important issues uncovered, which is discussed in this paper. General scientific design principles and guidelines are not adapted into practical sound design outside the scientific frame. The gap between scientific practices and sound design practices seems to be immense. The scientific AD studies have not widely discussed the pragmatic aspects of the everyday perception of sounds. This paper discusses the theoretical approaches to pragmatic design issues and reviews some main characteristics of different design approaches in AD, in order to illustrate the full potential that currently remains unexposed.

2. AUDITORY SIGNS

Currently, the AD research field divides the different sound types into *earcons* or *auditory icons*. Earcons are currently defined *"…as abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of an interface*" [4]. The symbolic relationship between the sound and its intended meaning can be considered useful and beneficial, as sounds do not have to correspond to what they represent. This provides the possibility to express combined complex messages (such as hierarchical menus) and events/objects that are impossible to sonify in a representational manner. Meanings are arbitrarily coded and, therefore, learning specific codes is required prior to effective understanding. The current mainstream of auditory sign research usually develops the abstract earcons according to the syntax of western tonal music [5, 6].

The contemporary definition of earcons excludes their iconic sign-signifier relations. The other type of auditory signs acknowledged is auditory icons, which are characteristic sounds based on a principle of direct mapping or metaphorical association of sound and source. When there is an available analogy between the event/object in the interface and some real world sound, it is possible to use auditory icons. Nevertheless, the number of analogies is limited and this type of direct mappings of meaning is not always applicable. In the terms used by William Gaver [7], the mapping of the object can be nomic (relying on the direct mapping of the properties between sign and object, e.g. photo-person) or metaphorical (relying on the similarities and structure-mapping of properties).

The current definitions of auditory signs differ fundamentally from the original definition of earcons by Blattner et al. [8]. Furthermore, the current definitions are not compatible with the sign descriptions of e.g. semiotic science. The original definition of earcons includes all levels of meaning, thus being essentially more pragmatic. The definition by Blattner et al. includes auditory icons in earcons as representational earcons, and earcons are: *"non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction"* This first definition conceives earcons to include **abstract semi-abstract** and **representational earcons**. In their article, Blattner et al. suggest that; *"Gaver investigated representational earcons, which he called auditory icons".* Furthermore, Blattner et al. discussed that the earcons are aural counterparts of the visual icons. Dividing visual icons into abstract or representational types is pragmatically only a

theoretical distinction, as most of the signs we encounter in our everyday life (visual, haptic or auditory signs) fall more to the category of semi-abstract signs – meaning that they consist of both iconic and symbolic elements [9]. The "arbitrary" should not mean that any random sign would be suitable – the choice should fit our cognition and support the meaning.

When it comes to the excluded semi-abstract auditory signs, rather than inventing new words or concepts, it is more suitable to use the term *auditory sign* to *cover all non-verbal interface sounds*. The current division of sign types is to some extent unsuitable for the purpose of designing functional, intuitive auditory interfaces. The exclusionary division of signs creates pragmatic and theoretical problems – some auditory signs (e.g. speech-derived *spearcons* [10]) that have great potential must be defined with completely new terms – and due to the fact that most sounds do not fit into any definition, the full potential of auditory signs unfortunately remains scientifically undiscovered. For instance, the term *intuitive earcon* is to some extent paradoxical from the start, according to the current definition of earcons. Pragmatically, a sign can be considered as intuitive when it has some evident forms of suggesting its purpose – and when the sign has this kind of meaning coded in it, it is no more purely abstract.

2.1. Overlapping sign types

As an analogy to visual icons, the current dichotomy of earcons and auditory icons as two exclusionary sound types can be compared to visual interfaces where only either simple geometrical shapes with primary colours or photos exist in the interface at a time. However, when observing the everyday environment, it can be easily noted that this is not the case, which is discussed in this section. As a brief inspection of the most renowned theories of signs from semioticians Charles Peirce [11] and Ferdinand de Saussure [12] demonstrates, *symbolic* and *iconic* levels of coding should rather be considered as a theoretical concept than as a strict guideline.

De Saussure introduced the concept of arbitrary, artificial signs, which correspond to the Peircean term symbolic. Both argued that the symbolic/arbitrary sign-object relations are defined by known conventions, habits or codes – such as language. Both find the iconic sign-object relation similar: the sign is similar to the object. Peirce also defined a sign type called index. This refers to a sign that has a direct connection to its object. For example, smoke is an index of fire and a feedback sound can be considered an index of a pressed button.

However, semiotics acknowledges the difficulty of defining symbolic or iconic levels even when language is used as an example of symbolic/arbitrary codes. We use words that have an iconic relation to the object. According to the principles of onomatopoeia (e.g. wolfs howl, doors squeak), we use the sounds caused by the actual phenomenon as the basis for its corresponding word [9].

From this viewpoint, rather than discussing the categorization of auditory sign types, it is more suitable to understand the complex nature of auditory signs in a more pragmatic way. Motivation and restriction are terms from Roland Barthes which refer to how much the signified control the signifier (the object controls the forms the sign can be presented in). A sign encountered in everyday life rarely relies on purely symbolic or iconic mapping. Most of signs include levels of both (as in signs representing humans in figure 1.) Letters are a good example of purely symbolic signs, but for instance traffic signs include levels of different types of signification. The levels of signification should rather be considered as a continuum, where the labelling of an individual sign is supple.

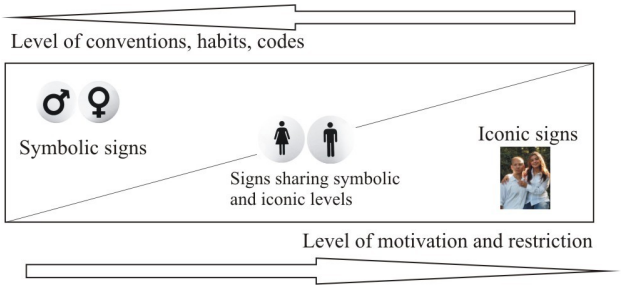

Figure 1. *Rather than either-or categorization, signs should be portrayed on a sliding scale.*

It is impossible to measure the level of motivation of a sign analytically. Extremes may be found, but everything between is impossible to place on a indisputable ordinal scale. Sometimes it is difficult to distinguish whether a sign is motivated by habit or whether it is iconic. Some instruments are considered to have "sad" sound [13], but are they considered sad merely due to sad spectral dimensions, or based on habits; is it just because we have heard that kind of instrumental timbre in a sad context?

The theoretical approaches from the film sound design support this view. David Sonnenschein [14] defines the sound types used in audiovisual narration as **concrete sounds**, **musical sounds**, **music** and **voice**. Common features are similar to the AD approach, with the difference that film sound design explicitly recognizes that *"…this categorization is flexible, as one particular sound may fit into more than one possible division or may fall between the cracks. Your choice should be based on what helps you create overall design… / … since ultimately the audience will have an undivided sonic experience with no labeling whatsoever."*

The approaches of film sound design and interface sound design differ pragmatically speaking very little from each other. Whereas, roughly speaking, film sound design aims at affecting emotions and interpretations and working as a narrative backbone, interface sound design strives to create accurate, precise meanings. However, despite the different goals, there are several similarities in these fields. In the end, the user's expectations and judgments are similar; the type of sign does not matter as long as it suits the context and has a meaningful function.

Due to often overlapping sign types, the categorization of sounds is not a practical task. In order to understand the design issues of auditory signs better, it is necessary to consider sound design from a viewpoint other than the dichotomy of the sounds themselves. In fact, the categorization seems to be implicitly more related to the creation structures of meaning. Discrimination between sounds themselves should be considered only as a theoretical concept.

3.THE DESIGN PARADIGMS OF AUDITORY SIGNS

To fully understand the potential and differences between different types of auditory sign design, one should consider the different design paradigm approaches for the intended use and desired effect, rather than iconicity or symbolicity of the sounds.

Current mainstream abstract earcon design guidelines consider sound by emphasizing psychoacoustic phenomena in how sound may be masked or how sound streams can be segregated (judgments on timbre, register, rhythm, concurrent sound segregation etc.) [16].

These channel-oriented perspectives and considerations of channel-noise factors (e.g. masking) further emphasize the information theory based view of communication. However, keeping in mind that sound-evoked meanings are not trivial and

that the meaning construction process is multifaceted (explained more detailed in chapter 4.1), sound designers cannot rely on truly random selection of the arbitrary sounds. If the connotations, clichés, genre sounds, habits and conventions have not been considered in the design process, abstract auditory signs may become *auditory distracters* [3, 17]. Users construct meaning from the whole use situation. Even though a meaningless beep may become meaningful in a reasonable chain of actions, undesired effects can occur as well if the sound does not correspond to the situation.

On the other hand, the auditory icon design paradigm relies on iconicity and the ecological perspective of auditory perception. Relations between sound and its meaning are based on connotations and similarities with familiar aspects of our everyday environment. They can be denotations of sound sources or attributes of sound that point to some physical properties of a sound-causing event. When listening to interface elements we intuitively recognize familiar parts from the sound and construct the meaning from their relation to the situation. Attributes such as size, material, force, sound source etc. are used to convey the information of the application to the user. This kind of mapping of the data can be used in e.g. data sonification techniques and overall control related interfaces such as in ARKola [18].

The most important difference of the earcon paradigm is that the design in auditory icon paradigm is more focused on how the sound itself, through similarities and metaphors motivates the meaning creation process. This approach is effectively used in film sound design and game audio design. The sounds are motivated by their meanings.

Nevertheless, even the distinction between design paradigms is not as clear. In most sound design the paradigms can easily exist simultaneously – an optimally designed AD can also enhance the intuitiveness with e.g. iconic and affective levels of meaning with communicational cues to some familiar qualities or habits from the surrounding environment. Due to the complex nature of audio-evoked meanings, interface sound designers should consider the whole: what is the use situation and what must be presented to the user to create as intuitive use situation as possible, and how the sound qualities can support the communicational purpose? In some situations, one approach is more fruitful than another. It is not possible to rank one design paradigm over another universally as a communicative element. Just as the distinction between sound types, also the distinction between design paradigms should be considered only as theoretical, overlapping and supplementary

4. MEANING CREATION PROCESS

For a long time HCI as a scientific discipline has implicitly considered human cognition as information processing systems. Usually computers are used as a metaphor to describe human cognition. However, designers must bear in mind that humans are not computers, and interaction cannot be considered solely in the dualistic terms of coder/decoder or subjective/objective. In the pragmatic interaction situation, the "polarities" are complementary and inseparable, and in parallel with the environment. A fundamental problem arises when auditory signs under study are separated from the physical and social world. The dualist subjective-objective dichotomy is not realistic for describing the relation between subject (user) and environment (device, interface and context) reasonably (for in-depth discussion, see e.g. [19]). The process of meaning creation is rich and complex; the meaning structures of sender and receiver are not symmetrical or necessarily similar [20]. However, discussing this question exhaustively is not the purpose of this paper. Therefore, to simplify the issue, the listening process and the meaning creation process of sounds are covered separately in the following section.

4.1. Meaning structures of interpretation

In the research field of auditory signs in user interfaces, there has been only little discussion concerning sound-evoked meaning construction. In 1994, the workshop report of CHI'94 [1] stated; "*For example footsteps or door knocks can be used as auditory icons in variety of ways. Yet there can be deeper more expressive level to sounds as well; different footsteps or knocks can subtly indicate size, importance and host of other feelings. Current user interfaces have not yet addressed this deeper expressive level in their use of sounds. Perhaps this is partly attributable to the fact that people can create this kind of sounds for movies are Foley 'artist' not 'scientist' or 'engineers'.*" Despite this early recognition of the issue of sound-evoked meaning, it has not been widely considered within the research paradigms of UI-sounds.

Listening is an active process, and it is influenced by the listener's attention, orientation and focus – including low-level and high-level cognitive processes. The information theory based psychoacoustic approach to the *hearing* process is insufficient when considering the listener's experience of sound in a situation. Users do not consider sound by its abstract qualities. Rather, we reflect the sound, source and potentially useful information and the communicational purpose it conveys within the particular context. It is important to consider the psychoacoustic low-level processes, such as masking, but this is not sufficient when considering functional and intuitive AD. A listener can perceive several different aspects of a single sound. In the sound design process, designers should focus on how sounds can be interpreted. The latest theoretical contribution to the distinction of listening modes [21] distinguishes seven different levels: *reflexive*, *connotative*, *causal, empathetic, functional*, *semantic, critical* and *reduced*.

- Reflexive mode refers to pre-conscious, automatic reactions to sounds – e.g. startling or instant attention grabbing – usually caused by sudden or surprising sounds.
- Connotative mode is another involuntary pre-conscious mode, which signifies freely formed associations of a sound that are evoked without further reasoning. These include connotations of the physical properties of the source or cultural associations evoked beyond logical reasoning. Some of the connotations can be difficult to explain why they are perceived, and they are not indisputable, such as the connotations of power perceived in the sound of a car engine.
- Causal mode emphasizes the denotation of the source. What causes the sound, what is the source? When one hears footsteps, the cause is recognized as footsteps caused by a human – the outcomes of the causal mode are often easily and indisputably explained as what was the source of the sound just heard, like footsteps, door slamming, telephone ringing, woman speaking etc..
- Empathetic mode is closely related to causal and connotative modes. In empathetic mode, focus is on perceived mental states of the source. Not necessarily concentrated on sounds made by humans or animals, but also on situations such as sound of the slammed door can be perceived as aggressive.
- Functional mode refers to the fact that listening is not separable from the context – in this mode our focus on the purpose and function of the sound heard i.e. what communicational action the sound indicates.
- Semantic mode focuses on the reflection of sound-conveyed conventional, intended meanings. In the critical

mode, focus is on the suitability of the sound – what is the reflective judgment of the sound, is it suitable for the situation (aesthetically or semantically etc.), and is it understood correctly? These three context-oriented modes are very closely related to each other.

- Reduced mode refers to listening to the qualities of the sound, which requires a conscious resistance of denotations and connotations of cause and meaning. This kind of listening should be considered more as a tool, "opening ears" for sound designers, as it requires a very high level of cognitive abstraction.

These modes of listening should be considered as theoretical concepts – they are useful when we are discussing all the different meanings sound can convey or whether there is a possibility of misunderstanding. However, in real-life situations, we concentrate only on some level of information that is useful and meaningful in the current situation. In addition, it is important also to consider the possible disturbing and distracting level of information. The different elements of information we pick up is reflected to the context, situation, overall aesthetics, conventions and other elements of information. The same sound can be listened to with different outcomes in different situations and orientations. One can use (often subconsciously) different strategies in listening. Different modes usually work together, influencing each other.

What kinds of information can be perceived from sounds is important for the sound designer in order to design functional sounds and avoid undesired effects. Sounds used in a situation should be considered from various perceptual perspectives – including the viewpoint of the use situation and the functional concept. Sound design is on a robust basis only when connotations, perception of emotional information, suitability, functions of the sound and the context is considered in detail from the viewpoint of the listener.

4.2. Meaning structures of representation

For sound designers, rather than re-inventing the wheel, it is sensible to use sound as we know it. It was noted in the CHI´94 report that, in order to use sounds effectively, one can: *"... mimic the ways we constantly use sound in our natural environments..."* This remedy is still relevant. The use of deeper expressive levels in the sign can be challenging – usually it relies on the professional skill of the sound designer, and a purely analytic approach might be difficult. Despite the challenges, this is an issue that cannot be ignored. In parallel with the listener's multifaceted meaning perception process, designers should consider how they can affect those structures.

The seemingly arbitrary sounds can and should have minor cues and motivational levels to make the sign easily interpretable. In addition, Gaver describes symbolic mapping: *"Symbolic mappings are essentially arbitrary, relying on social conventions for their meanings."* The dependence on conventions relates to the fact that symbolic mappings should not be merely random. The meaning of the sound is inseparable from the function it serves in each use context.

Despite the demonstrated and speculated strengths and abilities of the symbolic syntax-approach in earcon paradigms as forming complex, compound messages, current research frames do not provide any syntax. Current design guidelines focus more on the psychoacoustic aspects of hearing and masking effects than creating any reusable syntax for meaning formation. Currently in the earcon-paradigm, the sounds used are usually selected randomly. Every design case seems to follow a different syntax.

As the Wittgenstein's concept of "language game" demonstrates, "*meaning is use*" [22]. This refers to the fact that for example words are not defined by their references to objects, but rather by their use in pointing to certain things – the meaning derives from the conventions. The meaning of the sign is defined by what it is used for, what is the function. In communication we do not express signs, but rather *communicational acts* which have desired effect and function in the context. The successful expression of communicational acts is not dependent of the sign, the same communicational act can be expressed with different discourses or modalities. Habits and conventions are inseparable from the meaning creation processes. In addition, a single arbitrary sign may become iconic through convention (e.g. ringing telephone, alarm sounds, car horn etc.).

Considering the strong nature of habits and conventions, it is not suitable to just invent new syntax or grammar for every singular application or research frame. Instead, we can find effective, already learned intuitive means of meaning construction from our surrounding environment and by observing our everyday communicational acts that can be applied to the design of auditory signs. Human communicational acts, our everyday perception of environment and cultural knowledge offer a vast array of exploitable means to code intuitive meanings into non-verbal interface sounds.

4.2.1. Human communication

Humans are very accurate in expressing and interpreting very detailed and even slight nonverbal content on the fly [23]. Effective use of this content may be the key to intuitive auditory sign design, as habits of human nonverbal communication are also key functions in the construction of musical meaning. Effective use of timbre, rhythm, intervals, silences and other means correspond to expression of affective structures, human nonverbal communication and expression of emotions, intentions and other meanings. Music often mimics the characteristics of human behavior – expression of emotions and other meanings in music are traceable to emotions and intentions expressed in non-verbal communication.

For example, in mother-child communication words are in a minor role – most of the affective content of speech lies in non-verbal prosodic information. The emotional and affective content of an interface elements is not an irrelevant consideration – e.g. in feedback sounds affective content of the sound can imply whether the object was accomplished or not. Especially warning sounds and monitoring interfaces can possibly exploit this content very fruitfully – slight emotional and affective cues in the sound can easily imply, for instance, the desirability of progress and the urgency of the task. Some studies argue for the important role of evoked emotions in decision making and their guiding effects on future behavior [24].

Therefore, the recommended usage of "emotionally dead" midi synthesizers to create functional, intuitive auditory signs may hold back the full potential of AD. When potential affective structures are taken into account during design and production, users may intuitively know the meaning of the sound heard. Arbitrary coding should not mean simply randomly selecting tones and instruments.

4.2.2.Communicational cues

Donald Norman introduced the concept of constraints to aid the design process [25, 26]. These constraints can and should be used as a basis for the design of auditory signs as well. The constraints are cultural, physical and logical, and they can be adapted to sound design when we consider the communicational cues of the sound used to imply a meaning. The conventions and habits we are familiar with should have a major role in

interface design processes. Cultural conventions are similar what Norman addressed as cultural constraints. When designing something new, the designer should bear in mind similar use cases. It is easy to learn new concepts if these are similar to ones we already know and understand, and if the designer bypasses conventions intending to create something new, design might end up simply confusing to users.

One approach to the design of cultural cues is the sound-evoked meanings we learn from music, game audio and movies – in addition to the existing interfaces we already use. Designers or scientists cannot ignore the strong influence of our cultural environment (e.g. the sounds computers make in movies) on our perception of the everyday environment. Conventions and habits are not arbitrary – they evolve in use and require everyday practice. Furthermore, users do not contemplate whether a convention arises from the movies or "serious" use. In the end, the only important issue is that the convention is already familiar. These cultural cues are in an important role when designing abstract earcon syntax.

Norman introduced a supplementary concept of physical constraints. This approach is very closely related to the auditory icon paradigm discussed earlier. In addition, it is very closely related to the concept of affordances. Affordance is a term originally introduced by J.J. Gibson [27] and originates from the ecological approach to cognitive psychology. The term refers to the natural properties of objects, which offer and imply the possibilities of actions between the object and actor. An important characteristic of affordance is complementarity of interaction between organisms and the environment. Physical properties of sound can imply the meaning (source and its state and intentions) and suggest behavior for the user in the interaction situation. Affordances and physical cues are very important in the meaning creation process. We perceive and interpret our everyday environment accurately, and using these cues is a robust basis for meaning creation in auditory signs. The concepts of physical cues and affordances are, to some extent, inseparable from the cultural cues. For example, some conventions in music are usually developed based on our everyday soundscape qualities – musicians often mimic the environment for effect.

Logical cues are results of reasoning. In AD this refers to the phenomenon where a seemingly meaningless beep becomes meaningful in a reasonable context and in a chain of reasonable events. If we press a button, we easily interpret the beep correctly as positive feedback of our action, even though the beep itself does not have any properties that could be interpreted as an indicator of successful button pressing. If the sound is heard alone, it is just a beep. Logical reasoning does not necessary depend on the action-obvious reaction relationship, but it is also applicable in interfaces where learning can occur in long-term use.

Each of these communicational cues can work individually in some use situations, but on the other hand most use situations might suggest they are used simultaneously. Designers should carefully consider the use situation and the required cues to imply the underlying communicational act, or their justified absence.

Even in the absence of musical syntax and obvious "instrumental qualities" simple beeps can be the most effective. Without pure iconic mappings, sounds can be used very effectively. With iconic sounds, due to cultural constraints, usually authentic sounds are not the most functional. Realistic sounds are not necessarily as dramatic or suitable as artificial sounds a professional Foley artist can produce.

As discussed in the CHI'94 workshop report, *"Audio designers from various domains (film, video, games, music) all have design knowledge which may be applicable to auditory interface design."* And due to the growing and active role of different interfaces in our lives, UI sound design should be well-aimed in order to achieve the desired effects and avoid undesired ones. The professional know-how of sound designers should not be excluded, even though approaching the use of different types of communicational cues in purely analytical terms might be impossible.

5. CONCLUSIONS AND DISCUSSION

In the AD field, several successful studies have been conducted in which different applications and domains of AD have been tested. However, the lack of pragmatic evaluation of the theoretical definitions has made its mark: the design principles and guidelines derived from these evaluations have not been adapted outside the scientific discipline of AD studies. The semantic gap between research paradigms and actual sound design can be considered enormous.

In view of practical use and everyday design issues, the previous theoretical concepts prove to be insufficient, and fundamentally different from the original concepts of auditory signs. There have been some evaluations of abstract sounds against iconic sounds. However, due to the different uses, purposes, users and the vast array of different applications, these evaluations cannot be generalized. In some situations with certain goals, iconic use of sounds is more effective than abstract, and vice versa. In some situations the abstract sounds might overcome iconic sounds with regard to performance time, but if abstract sounds are dissonant to the overall situation, they might even hinder the use compared to a situation without sounds.

There are certain risks in randomly selected abstract sounds. Michel Chion [28] discussed the same problem in film sound design. There is a vast array of sounds that suit a specific purpose, of which some are wholly conventional. On the other hand, there is also a vast array of sounds that are anything but suitable. The dichotomy of symbolic and iconic sounds excludes the majority of audio usage potential.

It is recommended that the definitions of different auditory sign types be considered as theoretical concepts only, and the original definition of earcons by Blattner et al. adopted: *"non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction"* This definition divides auditory signs into **abstract**, **semi-abstract** and **representational earcons**. However, due the established role of the term earcon as concept of compoundable messages, it is reasonable to widen the auditory user interface elements to be referred with term **auditory signs**, which covers all types of UI sounds. However, the distinction is only analytic, and analytic distinctions should not be confused with pragmatic ones.

Rather than discussing the symbolic/iconic sound types, which are extremely complicated and sometimes impossible to define, sound design research should consider the **earcon** and **auditory icon** approaches as different design paradigms and means of constructing meaning into sounds, without setting restrictive semiotic chains on design.

The metaphorical approach of traditional HCI to human cognition as a computer is not suitable for AD research. Audio-evoked meanings are discreet and formed in subtle ways. Rather than discussing *hearing*-related issues, designers and researchers should focus on the complex phenomenon of *listening* and on how we construct meanings form what we hear. Simply making sounds audible and easily recognizable is not the key to functional AD, even though it is an crucially important domain that cannot be bypassed.

Misinterpretation or dissonance in communication usually

arises from a lack of equivalence between sides of interaction. To avoid this, and to create as intuitive and easily interpretable AD's as possible, researchers should focus on how we can affect modes of listening in creating and designing sounds. What types of listening are potentially involved in the current situation and how can we affect them with sound? We can exploit ecological structures or cultural conventions, either from the domain of interfaces or e.g. music and film sound design, and non-verbal informational content of speech to convey meanings. Our surroundings provide a plethora of different exploitable means for non-verbal sound design.

Important future work involves categorically analyzing different functions, **communicational acts**, of sound in different user interfaces, and identifying the corresponding cues to improve intuitive recognition in interfaces. This meta-analysis of interfaces and communicational conventions would have an important role in developing reusable, rich syntax for auditory signs and it would narrow the semantic gap between scientific definitions and sound design.

One must consider the whole use situation (the user(s), applications, functions, goals, environment, socio-cultural habits and conventions) in order to define the required communicational nature of an interface. The iconicity and symbolicity can vary among different interface elements. Two polarities of interface sound are not exclusionary – bearing in mind that a single interface can consist of signs with different meaning creation strategies. The form of the interface and its elements should be determined by the functions, and the use of communicational cues should be formatted to fit the purpose.

6. ACKNOWLEDGMENTS

This work is funded by Finnish Funding Agency for Technology and Innovation (www.tekes.fi) and the following partners: Nokia Ltd., GE Healthcare Ltd., Sunit Ltd., Suunto Ltd. and Tampere city council.

7. REFERENCES

[1] B. Arons and E. Mynatt "The Future of speech and audio in the interface" In S*IGCHI Bulletin* Vol. 26, 1994, pp. 44-48

[2] Y Xiao, FJ Seagull "An Analysis of Problems with auditory alarms in process monitoring tasks" in *Proc. Human Factors Ergon. 43rd Annu. Meeting*, 1999 pp. 256-260

[3] P.M.C. Lemmens, M.P. Bussemakers and A. de Haan "The Effects of earcons on reaction times and error-rates in dual-task vs. single task experiments" In *Proceedings of the ICAD*, P.R Cook (ed.), International Community for Auditory Display, 2000, pp. 177-183

[4] S.A. Brewster, P.C. Wright and A.D.N. Edwards "Detailed investigation into the effectiveness of earcons" In G. Kramer (ed.) *Auditory Display*. Reading, Ma, Addison Wesley, 1994.

[5] J.C.K. Hankinson and A.D.N. Edwards, "Designing earcons with musical grammars" In *ACM SIGCAPH*, September Issue, 1999

[6] S.A Brewster, P.C. Wright and A.D.N. Edwards, "Experimentally Derived Guidelines for the Creation of Earcons" in *Proceedings Of BCS-HCI'95* Vol.2, 1995, pp. 155-159

[7] W.W. Gaver "Auditory Icons: Using Sound in Computer Interfaces" In *Human-Computer Interaction*, Vol. 2, 1986, pp. 167-177

[8] M.M. Blattner, D.A. Sumikawa, R.M. Greenberg "Earcons and Icons, Their Structure and common design principles" in *Human-Computer Interaction* Vol. 4. Issue 1. 1989, pp. 11-44

[9] J. Fiske "Merkkien Kieli – Johdatus viestinnän tutkimiseen" Vastapaino, Jyväskylä, Finland, 1994

[10] B.N Walker, A. Nance and J. Lindsay "Spearcons: Speech based earcons improve navigation performance in auditory menus" In *Proceedings of the ICAD*, London, UK., 2006, pp. 63-68

[11] C.S. Peirce "What is a Sign" In *Essential Peirce: Selected philosophical writings* Vol2., Bloomington, IN., Indiana University Press, 1998, pp. 4-10

[12] F. de Saussure "Course in general linguistics" Duckworth, London, 1983

[13] T. Lucassen "Emotions of Musical Instruments" In *4th Twente Student Conference on IT*, 2006

[14] D. Sonnenschein "Sound Design – The Expressive Power of Music, Voice and Sound Effects in Cinema" Studio City, CA, Michael Wiese Productions, 2001

[15] C. Shannon and W. Weaver "The Mathematical Theory of Communication" Urbana, ILL, University of Illinois Press, 1949

[16] D. McGookin and S.A. Brewster "Understanding Concurrent Earcons: Applying auditory scene analysis principles to concurrent earcon recognition" in *ACM Transactions on Applied Perception*, vol. 1(2), 2004, pp. 130-155

[17] M.P. Bussemakers and A. de Haan "When it sounds like a duck and looks like a dog… auditory icons vs. earcons in multimedia environments" in *Proceedings of the ICAD*, P.R. Cook (ed.) International Community for Auditory Display, 2000, pp.184-189

[18] W.W. Gaver, R.B Smith and T. O'Shea "Effective sounds in complex systems: the ARKola Simulation" In *Proceedings of CHI*, New Orleans, LA, April 28-May 2, ACM, New York, 1991, pp. 85-90

[19] K. B. Baerentsen, J. Trettvik "An Activity Theory Approach to Affordance" In *ACM Conference Proceedings Series*, Vol. 31, ACM, New York, NY, 2002, pp. 51-60

[20] S. Hall "Encoding/Decoding" In *Culture, Media, Language, Working Papers in Cultural Studies*, Hutchinson, London 1981, pp. 117-121

[21] K. Tuuri, M-S. Mustonen, A. Pirhonen "Same sound – different meanings: A Novel Scheme for Modes of Listening" in *Proceedings of Audio Mostly Conference*, Fraunhofer Institute for Digital Media Technology IDTM, Ilmenau Germany, 2007, pp. 27-28

[22] L. Wittgenstein, "Philosophical Investigations" MacMillan, new York, 1953

[23] P. N. Juslin and P. Laukka "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" In Psychological Bulletin, vol. 129, No. 5, 2003, pp. 770-814

[24] T. Ketelaar, A.S. Goodie "The Satisficing Role of Emotions in Decision Making" in Pyskhe Vol.7 No. 1. 1998, pp. 63-77

[25] D. Norman "The Design Of Everyday Things" MIT Press, London, 2001s

[26] D. Norman "Affordance, Conventions and Design" in Interactions, May Issue 1999, pp. 38-43

[27] J.J. Gibson "Ecological Approach to Visual Perception" Lawrence Erlbaum, Hillsdale, 1986

[28] M. Chion "Audio-Vision: sound on screen" New York, NY, Columbia University Press, 1994