NON-SPEECH AUDIO-SEMIOTICS
A REVIEW AND REVISION OF AUDITORY ICON AND EARCON THEORY

David Oswald

Design in Business Communication Management
HTW Berlin University of Applied Science,
10313 Berlin, Germany
oswald@htw-berlin.de

ABSTRACT

The aim of this paper is to develop a theory and taxonomy of auditory signs, based on semiotics. For more than two decades, the discourse on non-speech audio interfaces has been dominated by a dichotomy between auditory icons, which are based on everyday hearing, and earcons, which are based on musical hearing. The corresponding theory behind these concepts has to be revised for several reasons. First, the authors of these theories partly use semiotic concepts and terminology, but not always in a correct way. Second, the classification of auditory icons as "iconic", and earcons as "abstract" is too simple and based on the questionable premise that everyday sounds are per se iconic and musical motives are per se abstract and symbolic. Third, this widespread idea ignores the crucial role of the user in the process of perception. In addition, the users' perception of visual and auditory signs in computer interfaces is fundamentally different today, from how it was in the early years of graphical user interfaces — the time when the first auditory interfaces and the corresponding theories were developed.

1. INTRODUCTION

Computers operate with several layers of symbolic code ranging from binary machine code to high level programming languages. Therefore, strictly speaking, all signs in human computer interfaces are symbolic — at least on a technical level. Iconic signs have been introduced to human interfaces by a metaphoric transfer from the actual world to the computer model world. Visual icons have served as a model for both auditory icons and earcons [1], [2]. The related theory construction drew parallels between auditory and visual icons. Literature on both, auditory icons and earcons, has employed semiotic terms and definitions, but in some cases in a rather unorthodox way. The most common fallacies are the confusion over indexical and iconic signs, thus confusing *causality* with *similarity* [3], and the notion of earcons being purely conventional and symbolic [1].

In order to outline a semiotics-based theory of non-speech audio in human computer interfaces, the first necessary step is to correct these misbelieves. Not as an end in itself — a revised semiotic theory of auditory signs will also shed a different light on stereotype attributions concerning advantages and disadvantages of auditory icon and earcon use. It can be expected, that a better understanding of the semiotic processes will improve decision-making during the design process.

In a second step, the theory needs to be amended with respect to today's users who have grown up with digital media, the so-called *digital native*. The concepts of auditory icons and earcons were developed in the 1980s — at a time when graphical user interfaces and the desktop metaphor where still new and unfamiliar to the users. Today, many users have internalised the model world of the graphical interface to an extent that makes menus, icons, and windows actually feel "natural". This habituation effect strongly influences also the perception of auditory signs, and hence, changes the semantic relation between the auditory sign and its meaning.

After a brief introduction to some basic terms of semiotic theory, these two steps of review and revision will be made. Based on semiotic definitions, a taxonomy of auditory signs in human machine interfaces will be suggested.

2. RELATED WORK

Semiotics for non-speech audio has been adressed systematically only in recent years. Pirhonen et. al. [4] and a related article of Murphy et. al. [5] have rightly adressed the fact that a sign's interpretation is influenced by its semiotic context (syntagma). However, they adhere to the distinction betweeen real-world auditory icons and earcons that are "symbolic in nature". Petocz et.al. [6] have clearly described the listener's essential role in the sign process. Nevertheless, their re-interpretation of auditory icons as "conventional indicators" can be questioned. Last, Nam and Kim [7] provide a (too) simple one-to-one mapping of sign classes to auditory cues. Whereas they use Peirce's refined "ten principal classes of signs" on the semiotic part of the equation, they use only a rather undifferentiated classification of auditory signals.

3. SEMIOTICS

Semiotics, the study of signs and sign processes, is rooted in philosophy and linguistics. Due to the modern semiotics' tradition of more than a century, the various semiotic schools and their respective terminology cannot be discussed here in detail. However, in order to discuss a semiotic theory of auditory signs, it is necessary to introduce to a minimum of semiotic terminology beforehand. In this article, the semiotic terminology will follow that of Charles Sanders Peirce, who introduced the triadic concept of the sign, which emphasises the role of the perceiving person in the sign process [8].

3.1. The three aspects of the sign

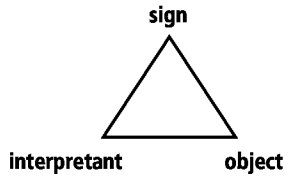

Figure 1: The three aspects of a sign, following Peirce [8].

1. Sign: the sign-carrier, the perceptible signal.
2. Object: the thing or the concept the sign refers to.
3. Interpretant: the interpretation in the mind of the perceiver.

It is seems somewhat confusing that one of the sign's parts is again called the "sign". In Peirce's terminology, it denotes the physically existing sign, which can be auditory, visual, haptic or olfactory. Some scholars refer to it as the "sign-carrier", "sign-vehicle", or the "signal". Eventually, mostly the more simple term "sign" is used.

In addition, the term "object" might be misleading. The object can be a physical object or thing, like a car or a trashcan, but it does not have to be physical. The object can also be an abstract concept like "democracy", or an action like "erase" [9].

Finally, the "interpretant" should not be confused with the interpreter, i.e. the interpreting person. It is rather the interpreter's mental conception of the sign's meaning. In other words, like the sign-carrier, the interpretant is a representation of the object. But whereas the sign exists physically in an auditory, visual, haptic or olfactory form, the interpretant exists "in one's head only". [8]:

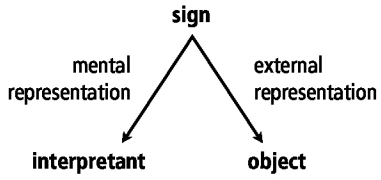

Figure 2: The sign refers to an external object and evokes a mental representation. Illustration by the author.

3.2. The three dimensions of semiosis

The sign process (semiosis) is subdivided into three dimensions that describe the relations between sign, object and interpretant [10]:

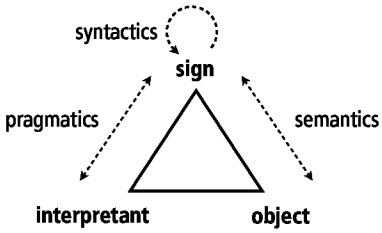

Figure 3: The dimensions of the sign process. Own illustration, partly based on Morris [11].

1. Syntactics: The relation between sign and other signs, rules for the formal structure of signs.
2. Semantics: The relation between sign and its object, its meaning.
3. Pragmatics: The relation between the sign and its interpretant, the effect the sign has on the perceiver.

3.3. The three types of relation between sign and object

Semantics are not only about the meaning of signs, but also about the principles behind the construction and encoding of their meaning. Semiotic theory differs between three types of signs, based on distinct relations between the sign and the referred object.

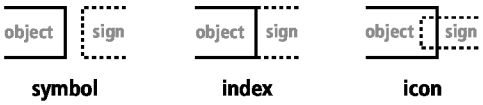

Figure 4: Schemes of the relations between signs and their object. Own illustration following Bense [12].

1. Symbol: based on *convention*, no factual link between sign and object
2. Index: based on *causality*, physical link between sign and object
3. Icon: based on *similarity* between sign and object

4. TYPOLOGY OF VISUAL AND AUDITORY SIGNS

The scientific discourse in the auditory display community has been utilizing some semiotic concepts and terminology, but — as will be discussed in chapter 3.1. — not in a consequent or consistent way. On the other hand, the semiotics community has hardly discussed the domain of non-speech audio.

Morris was the first to systematically apply semiotics to the visual domain [13] and to teach semiotics in a design context [14]. Today, semiotics is an integral part of the curricula of numerous graphic design study programs, but in auditory communications semiotics remain regrettably unutilized.

This blind spot of the semiotic discourse has its origin in the discipline's strong tradition in linguistics. Even in Nöth's extensive "handbook of semiotics" [15] only a small chapter on semiotics of music can be found, but the term "sound" is simply non-existent in the subject index. In musicology, there is also no great tradition in semantic analysis of music. The meaning of music, in the sense that it refers to extra-musical phenomena, is not in the focus of traditional art music theory. In most cases, musical analysis is mainly based on the syntactical and self-referential inner structures of music. Exceptions to this are the semiotic driven works of Tarasti [16], Nattiez [17], and Cummings [18], and Clarke's approach to musical meaning based on ecological perception [19]. In contrast to everyday sounds, music does not have an unambiguous meaning. If a piece of music has extra-musical meaning, it is often based on a complex, multi-layered, and interwoven symbolic (cultural) coding [16]. Hence, music is a form of communication with a great power of evoking associations and moods, but it is usually not used in a strictly functional context, that is to communicate well-defined

meanings effectively without ambiguity. However, below these multiple cultural layers there are also musical universals, which are independent of cultural context. For instance, the sense for tempo, and what is considered fast or slow, is similar across all cultural backgrounds. Musical universals can be used to design music-based signs that are *not* arbitrary and symbolic, and therefore are as easy to learn as natural everyday sounds. This will be discussed further in chapter 3.2.

4.1. Index

The most frequently used example for an indexical sign is smoke as a sign for fire. Smoke indicates a fire, and it does so by merely pointing to it, without being similar to the fire and without cultural conventions behind it [8]. The index sign is linked to its object simply by the laws of nature — it is a symptom. The auditory index sign for "fire" would be the fire's typical crackling sound. The fire physically causes this sound, it is the auditory effect of physical and chemical processes that we call "fire". The index sign "crackling" and its object "fire" are linked so closely, that one could argue that "smoke" and "crackling" are both integral parts of the perceiver's conception of "fire". Everyday listening is mainly based on these indexical sign processes. Gaver also points to the direct and effortless perception of physical everyday sounds:

Our normal mode of hearing is to listen to sounds to identify the events that cause them. From this perspective, sound provides information about materials interacting at a location in an environment. [2]

4.2. Icon

Most definitions of the iconic sign use the term "similar" to characterize it. Thereafter, the icon is based on a similarity between the sign and what it stands for [8]. In order to be more precise, Morris circumscribes the concept of similarity with "shared attributes between sign and object" [11]. The iconic principle of similarity is widely used in visual communications. For instance, a silhouette drawing of an animal on a traffic sign becomes understandable by the depiction's similarity to the animal. Sign and object share some attributes of shape. Iconic auditory signs in this sense would be sounds that *sound* similar like other sounds. Foley artists often use iconic sounds, for instance when using coconut shells to imitate horses, or when using a snare drum as an exaggerated illustration of a punch in the face.

A recording of a sound is, when played back, an icon for the original sound. Digital photo cameras use pre-recorded mechanical shutter sounds to indicate an otherwise silent digital process. When originally produced by a mechanical camera, this sound is a physically caused index sound for the shutter release. Everyone who is familiar with analogue photo cameras understands this indication intuitively. Therefore, when a digital camera reproduces a shutter sound, the imitated sound is interpreted due to its similarity with the original sound. It is an auditory icon. But what about younger users, who are not familiar with vintage photo gear? For them the meaning of the same sound is pure convention — a symbol. [20]

4.3. Symbol

The well-known error beep is a typical example for an auditory symbol. Symbols are based on mere convention, neither laws of nature nor perceivable similarity link a symbol to its meaning [8]. The sign's shape or sound has no factual connection with what it refers to, which is why the symbolic sign often is referred to as being *arbitrary*. The traditional error beep is in fact arbitrary, in the sense that its timbre, pitch and duration do not contribute anything to its meaning. A higher or lower pitch or a different waveform would do the same job just as well. Pure waveforms, like sine waves, lack physical indexical meaning because they are hardly heard in everyday interaction with the environment. They can only obtain a meaning by declaration and convention [20]. But what about using real world sounds, like frog's croak or glass bottle sounds, as a sign for a computer error? In relation to their actual meaning, these are just as arbitrary as a sine wave. Originally, they are index sounds, which indicate for instance the presence of a frog. Transferred to a different context the indexical meaning retreats to the background and gets overlaid by the new symbolic meaning. It is only a matter of repetition and training until the second meaning becomes dominant [6].

Multilayered meanings are not restricted to digital technology, for instance the sound of a church bell is initially only an indexical sign for a clapper hitting a metal bell-shaped vessel. Still, the predominant meaning of this sound is the appeal to attend church service, or the profane indication of the current time. Both of the latter codes work on a symbolic level, based on initially arbitrary cultural conventions — other cultures use different sounds for these purposes. Even this arbitrary coding can be perfectly internalized in a way that it will be understood just as fast and intuitively as natural indexical sounds [21].

4.4. Iconicity

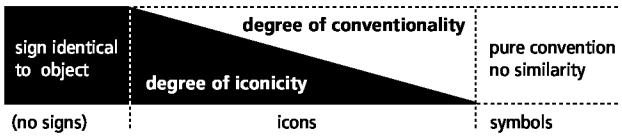

Figure 5: Gradual transition of icon to symbol, from high iconicity to high conventionality [22], [23].

In order to discuss the typology of auditory signs further, it is necessary to have a closer look at *similarity*. In the visual domain, it seems to be obvious when a sign is similar to its object. The silhouette drawing of a cow on a traffic sign is said to be similar to a real, living cow — at least in some aspects. Here similarity is based on proportional scaling, reduction to two dimensions, elimination of materiality and colour, and reduction of details in shape.

However, similarity is not restricted to analogue transfers like scaling or reduction of detail. A merely diagrammatic similarity is also considered to be iconic [8]. Even if a subway map is not drawn to scale, or a circuit diagram does not represent the spatial arrangement on the circuit board, both are still iconic representations based on structural similarity.

In order to describe different levels of similarity between sign and object Morris introduced the term "iconicity" [22]. In this sense, the attraction of Madame Tussauds' wax figures is based on a very high iconicity, whereas a subway map is based

on low iconicity. The upper end of the iconicity scale is delimited by a sign that is identical to its object, and therefore would not be a sign anymore. Below the lower end of the iconicity scale's is a sign that has no (more) similarity with its object — a symbol [23]. This delimitation is not defined by objective properties of the sign, but solely by the perception of the interpreter. If a low level similarity is recognised or not, depends strongly on the perceiver's previous knowledge, cultural background and frequency of use [24].

The concept of iconicity as a degree of similarity is easily understood when dealing with visual icons. Similarity of auditory icons is harder to define, since natural sounds are signs for *events*, they are *time-based*, whereas visual icons represent *things*, they are *spatial*. In the following chapters, the question of iconicity of auditory signs and to what they actually are similar will be addressed.

4.5. Using index, icon and symbol

Taking a superficial view, an index sign seems to be the most intuitive sign to be understood, because it is "natural". The second choice would be the icon, because it bears the potential to be understood by resemblance. The symbol would be coming in last, as "arbitrary" usually is considered almost synonymous with "inapprehensible". While it is undoubted that different sign-object relations exist, it must also be clear that in terms of understandability the different types of sign are only good for a head start effect. All described advantages can and will be overridden by the effects of repetitive use. Moreover, in fact index signs are not more intuitive because they are "natural" — they have become intuitive only because we have been exposed to them for a longer time.

The given description of the three types of signs has been simplified in order to be clear and concise. In fact, also the interpretation of indices and icons are to a certain extent subject to cultural differences and context. For a discussion on the cultural influence on the perception of "direct physical experience" (i.e. index signs), see Lakoff and Johnson [25]. For a discussion on the conventionality of icons and on perception of similarities as a cultural technique, see Eco [26].

5. AUDITORY ICON AND EARCON THEORY

The terms "auditory icon" and "earcon" have been coined in the 1980s, when the discipline of auditory computer interface design emerged. In the early years, the discourse has been dominated by a methodological debate about which of the two concepts is more effective and easier to learn. Today *both* are standards in auditory display design. Both concepts have constituted the (still improvable) auditory environment of today's computer users. Browsing for instance sound folders of Apple's OS and Microsoft Windows, auditory icons and musical earcons can be found in peaceful coexistence. This is also reflected in scientific discourse: A cumulative word count through the ICAD proceedings of the past three years shows 490 hits for the term "earcon" and 356 hits for "auditory icon", with an average of six occurrences (!) per paper.

In order to reconceive auditory icon and earcon theory, it is necessary to once again have a look at classic publications, which coined and imprinted these terms, since some debatable attributions that originate from these early papers keep being repeated until today. The most problematic stereotypes in this context is the notion that auditory icons are per se iconic, and that earcons are generally abstract, i.e. symbolic.

5.1. Are auditory icons really iconic?

It is needless to say that Bill Gaver's work on auditory icons [2], [3] has been seminal for auditory display design. In his dissertation, Gaver transferred Gibson's approach of ecological perception [27] from the visual to the auditory domain [28]. He analyses how information can be obtained from everyday sound and discriminates it strictly from musical hearing. Due to their intuitive understanding, Gaver recommends the use of everyday sounds for auditory interface design. In his argumentation, he refers to Peircian semiotics, but obviously confuses index and icon when he claims that "iconic mappings are based on physical causation" and "its characteristics are causally related to the things it represents" [3]. This is true for indices, but not for icons, which are not based on causality but on similarity. This flaw has been noted before by Petocz et. al. who then conclude that auditory icons in fact are auditory *indices* [6]. However, the matter is even more complex.

As we have seen in chapter 2.1, everyday sounds are indexical. But what happens when these sounds are being detached from the event of their physical causation? A recorded and played back sound could be described as an index for a past event. Even with the best high fidelity equipment, the recorded sound will not be exactly the same like the original sound. Hence, a played back sound is, due to its similarity to the original sound, only a representation of the original sound and thereby also a representation of the original sound's meaning. Gaver's argument that auditory icons are iconic because they are based on physical causation is not correct. However, only the explanation was wrong. They are iconic, because they have been *copied and imitated*, as we have seen in the example of the camera shutter sound in chapter 2.2.

Admittedly, the camera example is different to most computer interface scenarios. In the shutter sound example, the original context and the new application context reside in the very same domain. In contrast, computer interfaces do not have mechanical predecessors that could serve as a source of physical sounds and established listening habits. Everything in a graphical user interface is based on metaphors. Using a trashcan to delete data on a computer seems almost natural today, but of course, it is based on a conceptual analogy between throwing away waste in real live, and marking hard disc space as unused. In real live, the accompanying sound when trashing something is an integral part of the perceptional pattern of "trashing". A visually similar representation of a trashcan, a similar interaction and a similar sound create a holistic multisensual analogy in the computer model world — and *iconic* sign-object relations in all of the three aspects: visually, auditory and interaction-wise. Such coherence in all aspects of a conceptional model is rare, because many processes in computers do not have an analogue equivalent in real live.

Some of Gaver's auditory icons in the "Sonic Finder" [3] were an extension of the visual desktop metaphor with what can be called an auditory "carpenter metaphor": Applications sounded like metal, like tools do. Files and folders — the material to be worked with — sounded like wood. Are these metaphoric signs also iconic? In what sense is a wooden sound similar to a digital file? What attributes do they share? These attributions do not seem to be built on similarity in the usual sense. Although, taking a look into the classic definition of "meta-

phor", we again come upon the concept of similarity. Following Aristotle, a metaphor can be a transfer based on the principles of *analogy*, which is in turn based on similarity or comparability [29].

However, a metaphor does not have to build on an already existing similarity. The similarity is rather created by the introduction of the metaphor [30]. Aristotle already pointed out that coming up with a good metaphor "implies an intuitive perception of the similarity in dissimilars" [29]. Thus, the sounds of files and folders in Gaver's "Sonic Finder" are iconic indeed. They are based on a conceptional similarity between metal/wood, tool/material and application/file. A similarity that came into life by the metaphorical transfer introduced by Gaver.

Already Peirce described three kinds of iconic similarity: A picture sharing basic qualities with its object, a diagram displaying relations only, and a metaphor where the similarity refers to yet another sign [8]. In Morris' terms, a metaphor is an iconic sign with low iconicity.

So far, we only considered Gaver's metaphoric mapping of file type to material and timbre. Based on this metaphor he also proposed mapping the file size to pitch, so that — analogous to real life experience — big objects would produce low pitch sounds and small objects would produce high pitch sounds. This mapping has been coded into the sound-producing algorithm, with file size as the parameter that determines the sound's pitch [31]. Thus, file size and pitch correlate in a fully predictable and reproducible way. This suggests a *causal* relation — which is untypical for icons, but constitutive of *index* signs. Here, causality is not based on the laws of physics, but rather on man-made rules written into a software algorithm. In this sense, parametrised sounds act on an indexical level.

In conclusion, signs that are based on everyday sounds are not necessarily auditory *icons*. When there is not even a metaphorical similarity between auditory signs and their meaning, for instance when a frog's croak is used as an alert sound, then even a natural everyday sound is simply arbitrary and symbolic. More complex are parametrised auditory icons that have at least two semantic layers in which meaning is encoded concurrently; the metaphoric icon with low-iconicity where timbre denotes the file type, and a second indexical layer where for instance pitch has an algorithmic, causal relation with file size. If these layers are both perceived equally, or if one layer becomes dominant, is eventually depending on the listener.

5.2. Are earcons really abstract, i.e. symbolic?

The counterpart to Gaver's first publications on auditory icons was the paper "Earcons and Icons: Their Structure and Common Design Principles" by Blattner et al. [1]. In this paper the authors coined the term "earcons" and defined them as auditory signs based on musical principles — short micro-compositions of only a few notes length.

Even if very short, earcons do share their design parameters with music: tempo and rhythm, melodic gestalt, timbre, dynamics, harmonics. Nevertheless, the authors mostly address parallels between earcons and visual icons as well as methods to create modular earcon families. Surprisingly, a discussion of how musical parameters can be used to convey meaning — the semantic impact of musical parameters — has been left aside completely. For instance, tempo and melodic gestalt obviously evoke strong associations, which can and should be utilized when designing earcons. Instead, the authors are content with the notion that earcons, in contrast to auditory icons, are abstract and symbolic and therefore simply have to be learned [1]. For Blattner et. al. the only way to facilitate earcon learning is a systematic and hierarchical earcon design. To speak in semiotic terms, it is a completely syntactic approach, ignoring semantic aspects of music. This compares to describing principles for writing readable text, while only focusing on grammar and spelling.

The concept of earcons as basically arbitrary compositions leads to a problematic negligence towards the actual composition of the earcon. Compared to everyday sounds, which are always indices for their causing events, it is much more difficult to describe the meaning of music. Music is widely considered being self-referential, bare of any extra-musical meaning. This may be true in some cases for "pure" art music. Programme music and especially functional music, like film music, show impressively how music is able to transport not only moods, but also information that can hardly be transmitted visually, such as the existence of monsters under a bed, or a protagonist's hidden feelings. These denotations are in most cases coded in multiple layers of cultural conventions, but there are also aspects in music that are directly understood, independent of musical training and across cultural differences. These so-called musical universals are based on biological and physiological structures [32], or rooted in human perception [19]. The sense of tempo correlates perfectly with both heartbeat and walking; 120 beats per minute are considered a fast tempo in music, a fast heartbeat rate, and also a fast walking pace. Universal music related patterns are also found across different spoken languages. An excited speaker will speak louder and faster, in a higher pitch, using greater intervals — features that are also used to describe excitement in musical theory [32].

The terms "high" and "low" pitch suggest a correlation between pitches and physical space. Indeed, most people associate a change in pitch with motion in an imaginary space. If this association is based on a physiological effect, is still being debated [33]. However, ecological approaches to the perception of musical meaning regard the association of motion as directly rooted in human perception [19]. Even if the effect was only culturally acquired, it is anchored into our listening habits so deeply that it is impossible to ignore when designing earcons. In Microsoft Windows, simple two-tone motives indicate when hardware has been added or removed. In fact, there is no objective reason for assigning an ascending interval to "adding" and an descending interval to "removing", but to match "in" with "up" and "out" with "down" fits listening habits and therefore feels intuitively right.

Longer motives can create a more complex contour or *gestalt*. Gestalt theory, originally developed in cognitive psychology in order to explain phenomena in visual perception, has also been applied to describe the perception of melodic patterns [34]. Tempo and melodic gestalt are just two examples to illustrate the non- arbitrariness of earcons. Rhythm, dynamics, and timbre also carry connotations that can and should be utilized in earcon design. The concept of gestalt had already been addressed during the very first ICAD conference in 1992 [35]. However, it did not lead to doubts about the concept of earcons as completely abstract and symbolic signs.

In conclusion, earcons can be completely arbitrary and symbolic, but they do not have to be. When a simple synthetic beep represents a system error, the beep is an arbitrary symbol. More complex earcons can also be arbitrary, for instance when

the famous four-note motiv of Beethoven's 5th symphony would be used to indicate "added hardware". However, there is a plethora of associations evoked by musical universals that can be utilized in earcon design in order to serve a communicative goal. Already a sequence of only two tones produces a notion of tempo, a directed motion, and a melodic gestalt with qualities like fast or slow, flowing or hesitant, up or down, and calm or volatile. In this case, meaning is based on similarity between patterns of musical perception on one side, and analogous perceptional patterns of extra-musical phenomena on the other. A musical tempo may have similarity with familiar timing patterns of strolling, walking, or running. These similarities are mainly metaphorical, since they cross domains like pitch and physical space. In this case, attributes from an original domain (i.e. physical space) are used to denote attributes in an alien domain (i.e. pitch). In consequence, well-designed earcons that build on musical universals are in fact *iconic* signs with low iconicity, for they make use of metaphorical similarity.

6. SIGN METAMORPHOSIS

The relation between the sign and its object does not exist objectively. It is not a fixed property of the sign. Whether a sign is perceived as indexical, iconic or symbolic does not solely depend on the quality and the characteristics of the sign, in fact it depends on the sign process as a whole. In which way a sign is interpreted by a perceiver is strongly depending on their previous knowledge and the present context. The same sign may be understood on a similarity basis by one perceiver and simply by habit and convention by another. Still, in large groups of perceivers, there are predominant patterns of interpretation. However, these predominant patterns of interpretation may change over time. In his theory of sign metamorphosis, Keller has described the shifting semantic relations between signs and their objects, and the changing ways of how a perceiver derives meaning from a signal [24].

6.1. From index to icon

When an index is imitated, it becomes an icon. To illustrate this effect, Keller uses the example of a simulated yawn. A real yawn is an index for a shortage of oxygen. Like index signs in general, yawning is usually not used for intentional communication. However, a simulated yawn can serve as an effective iconic sign for letting someone know how bored the listeners are. It is understood because it is similar to the real yawn. The same rule applies to auditory signs. As seen in chapter 2.2, a camera shutter sound becomes an iconic sign by imitation — it is then interpreted by an associative inference, based on the similarity between the original and the recorded sound. [24]

6.2. From icon to symbol

Whereas an icon becomes meaningful by an association that is triggered by perceived similarity, a symbol obtains its meaning by conventions, i.e. written or unwritten rules. Keller points out that the associative way in which iconic similarity is interpreted, is a creative process without normativity. It is always possible that the interpreter has an association different from the intended goal. This procedure of association can be compared to solving riddles. Confronted with the same riddle for several times, one does not have to associate and guess anymore. Therefore, by repetitive use, an icon will not be interpreted by similarity any more but based on a habit, a rule. The similarity actually is still there, but now remains unnoticed. The similarity has become useless. In consequence, iconic signs that are used frequently over long periods of time will lose more and more of their iconicity by simplification and abstraction. A visual example is the metamorphosis of the iconic cipher III to the symbolic 3, which developed over the centuries by cursive handwriting and rotation by 90°. In everyday conception, the cipher 3 is a symbol for most people, until they learn about the relation to its iconic predecessor III and start to see the visual similarity. Then the cipher 3 has again become an icon — for just as long as the similarity remains conscious. [24]

6.3. From any sign to index

In Keller's linguistic perspective the described sign metamorphosis is a one-way street where signs start as indices or icons, and become symbols at the end [24]. This may be true for spoken language, but is not necessarily the case in digital interactive systems. When we interact with interfaces, we continuously interpret visual and auditory signs emitted by the system. These signs follow the logic that has been encoded into the system by the system's designer and are meant to be either indexical, iconic, or symbolic. Whereas repetitive use of iconic signs in the analogue world often leads to a symbolification of these signs, in digital interactive systems it leads to *indexicality*. Whatever sound is played back, when for instance a file is dragged to the trash, if it is only repeated often enough, it will become an *index* for the event of successfully putting a file into the trashcan. This can work even with the most arbitrary auditory cue; it will require only more repetition.

In the perception of a frequent computer user, it does not make a difference if a sound is determined by physical parameters when interacting with the real world, or if a sound is triggered by the user's interaction with the virtual world and determined by man-made algorithms. The only required condition that leads to an indexical sensation is *perceived* causality. When I *always* hear the same sound when trashing something, and when I *never* hear it when I missed the trash, then the sound becomes quickly an indicator for trashing — independent of the sound's features and qualities. In the user's perception, his or her activity in the computer model world *causes* this sound.

6.4. Polysemy

Usually the term polysemy is used to describe ambiguous or multiple meanings of a sign. Thereafter a "beetle" can denote either an animal or a car, and in spoken language, it could also denominate John, Paul, George, or Ringo. An outline drawing of a man may represent a man — or a bathroom in a different context. In the latter example not only the meaning changes, but also the sign-object relation. In the first case, the relation is iconic, for the drawing visually resembles a man. In the second case, the relation is symbolic, because the depicted manikin does not share any visual attributes with the signified bathroom.

So there is obviously also a second meaning of polysemy, which does not deal with multiple meanings but with multiple types of sign-object relation. In Keller's theory of sign metamorphosis we saw that these relations change, from index to icon by imitation, and from icon to symbol by frequent use.

	indices	icons		symbols			
everyday sounds »auditory icons«	parameterised sounds	played back index sound orignal and new context within the same domain	played back index sound crossing related domains			arbitrary index sounds from arbitrary domains	interface sound
	causal relation by algorithm	strong similarity of physical sound attributes	metaphoric similarity an index sound's original meaning is a symbol for the intended meaning			pure convention, no factual relation, original meaning irrelevant/misleading	sign-object -relation
	pitch ≙ file size	real trashcan's *crash* sound ≙ virtual trashcan's sound	baby's *scream* ≙ error alert			frog *croak* ≙ error alert	example
musical motives »earcons«	any sound can become indexical by habituation		motives based on musical universals	motives based on cultural connotation of music	single abstract tones	arbitrary musical motives	interface sound
	perceived causal relation by algorithm		metaphoric similarity, patterns in musical perception similar to analogous perceptive patterns	multilayered symbolic meanings, reduced arbitrariness by tradition and cultural imprinting	pure convention no factual relation, no similarity, no causation	pure convention no factual relation, musical meaning irrelevant/misleading	sign-object -relation
			ascending interval ≙ add hardware	fanfare ≙ king ≙ glory ≙ success	simple *beep* ≙ error alert	5th symphony ≙ add hardware	example
			5th symphony ≙ error alert				

Figure 6: Proposed taxonomy of auditory signs.

It is important to note that these change processes do not proceed simultaneously or in a regulated way for all users [24]. Thus, a sign can be interpreted on a similarity basis, and *at the same time,* someone else may interpret it based on mere habit or convention. For the first interpreter it is an icon, whereas it is a symbol for the second. Still, in spite of their different ways of making sense, both interpreters can derive the very same meaning at the end. Concluding this chapter, we can say that it is hard, or almost impossible, to predict in which *way* a perceiver will interpret an auditory sign. Nevertheless, it is comforting to see that the intended meaning can still come across, even if in different ways.

7. INDEXICALITY: DIGITAL NAÏVES, DIGITAL IMMIGRANTS AND DIGITAL NATIVES

When Gaver published his first article on auditory icons in 1986 — only two years after the introduction of the Apple Macintosh — graphical user interfaces (GUI) where still new and unfamiliar to most computer users. Computer users who were confronted with iconic representations of files, folders, printers and trashcans on a computer screen, rightly conceived these icons as *representations* of something. Icons were perceived consciously as signs that stand for digital, symbolic, and invisible code. Users were very aware that the desktop metaphor is a *metaphor*, and that it was designed to facilitate learning to use a computer.

As explained in the previous chapter, signs change the way they are conceived for instance by frequent use. A similarity-based associative inference will be superseded by a rule-based inference or mere habit. Like this, icons become symbols. The initial iconic sign process is completely contingent upon the interpreter's ability to recognize or construct similarity [24]. In the 1980s these interpreters were inexperienced GUI users — digital naïves. In their everyday life, files and folders were physical objects made of paper and cardboard. In contrast, today's young adult users grew up with computers. These digital natives do not conceive the computer model world as a representation of an office [36]. Depending on their age, they probably did not even know paper files and cardboard folders before they encountered the corresponding representations on the screen. Therefore, for digital natives, these representations never were perceived as representations. Due to the lack of knowledge about the originally depicted objects, they were unable to construct any similarity. For them, representative "icons" were just arbitrary symbols. Hence, a semiotic explanation of the digital native phenomenon can be subsumed as *a sign metamorphosis taking a shortcut from symbols directly to indices*. The same effects apply to auditory signs. Neonates need some time to learn symbolic sounds like doorbells or police sirens. However, natural sounds also have to be learned in the first place. For instance discerning sounds of bouncing and breaking glass does not have to be easier than internalising a symbolic "beep" as an index for error. Hence, some of the advantages of everyday sounds are simply based on longer learning time.

In addition, *digital immigrants*, who did not grow up with digital technology but have adopted it, also develop indexical perception by continuous use. When the computer model world behaves consistently over long periods of time, when user interaction triggers predictable and reproducible feedback, then every user will soon internalise feedback signs and consider them as *indicative* for his or her actions. Like this, signs that once were consciously conceived as representations of something, become quasi-natural index signs. In the actual world, sounds are created by the natural law of physics. In the computer world sounds are caused by the laws of man-made algorithms. Once these algorithms are implemented, the sounds are determined by the user's interaction and the algorithms — the sounds can be internalized just like natural sounds.

8. CONCLUSION

Auditory icons and earcons cannot be attributed with fixed types of signs. In everyday life, natural sounds are indexical signs, based on causality. In an interface they can as well be iconic, or even completely symbolic and arbitrary. Earcons do have a tendency towards a conventional and symbolic coding of meaning. However, if composed attentively, they can also become iconic and intuitively meaningful (see figure 6).

The question of *how* a sound communicates its meaning, if a user makes sense of it by causality, similarity or convention, does have an effect on its learnability. However, since the type of sign (index, icon, or symbol) is *not* directly depending on the type of sound (everyday or musical), there cannot be a well-

defined rule of which type of sound is easier to learn. Learnability does not depend primarily on the *type* of sound, but rather on the distinct sound that is used, its characteristics, its sound design, composition, cultural connotation, or original context. Hence, especially in its early years, auditory interface discourse put too much emphasis on the types of sounds (everyday or musical). Over the discussion of principles, the concrete design of proposed and tested sounds has often been neglected — especially in the case of earcons and their proper composition.

Of course, a scientific community has to generate generalisable knowledge. However, generalisation should not lead to over-simplification. Labeling earcons with "abstract" and auditory icons with "concrete", and the deduced cliché that music-based earcons have to be learned, whereas everyday sounds are intuitively understood, are over-simplifications in that sense. In contrast to the sciences, design does not have to produce general truth. Design usually aims for specific solutions for specific users in a specific context. A rule in the manner of "use sound type A for purpose B with user C" is per se too simple to be valid. Of course there can be patterns or rules of thumb like this, but the fate of auditory signs is decided by the adequacy of their original context (everyday sound), their composition (musical sounds) and the specific sound design details.

Last, there is the decisive influence of the user: Habituation of individual users on one hand, and different ways of perceiving interfaces of the digital naïve, the digital immigrants, and the digital native on the other hand. A deeper understanding of these factors will allow for a better focus on the relevant issues of sound design for auditory display.

9. REFERENCES

[1] M. Blattner, D. Sumikava, R. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles", in: *Human-Computer Interaction*, vol. 4, 1989, pp. 11-44.

[2] W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces", in *Human-Computer Interaction*, vol. 2 no. 2, June 1986, pp. 167-177.

[3] W. Gaver, "The SonicFinder, An Interface That Uses Auditory Icons", *in Human Computer Interaction*, 4, 1989, pp. 67-94.

[4] A. Pirhonen, E. Murphy, G, McAllister, W. Yu "Non-speech sounds as elements of a use scenario: A semiotic perspective", *Proc. of ICAD 2006*, London, pp. 134-140.

[5] E. Murphy, A. Pirhonen, G. McAllister, W. Yu, "A Semiotic Approach to the Design of Non-speech Sounds", in *Proceedings of HAID 2006*, Glasgow, pp. 121-132.

[6] A. Petocz, P. Keller, C. Stevens, "Auditory warnings, signal-referent relations, and natural indicators: Re-thinking theory and application", *Journal of Experimental Psychology: Applied*, vol. 14, no. 2, Jun 2008, 165-178.

[7] Y. Nam, J. Kim, "A semiotic analysis of sounds in personal computers: Toward a semiotic model of human-computer interaction", in *Semiotica* 182, vol. 1 no. 4, pp. 269–284, 2010

[8] C. S. Peirce, "Logic as Semiotic: The Theory of Signs" in *The Philosophy of Peirce: Selected Writings*, London: Routledge & Kegan Paul, 1956, pp. 98-119.

[9] U. Eco, *A Theory of Semiotics*, Bloomington: Indiana University Press, 1979.

[10] C. Morris, "Foundations of the Theory of Signs", in *Writings on a General Theory of Signs*, The Hague: Mouton, 1971, pp. 17-71.

[11] C. Morris, "Esthetics and the Theory of Signs", in *The Journal of Unified Science (Erkenntnis)*, vol. 8, no. 1/3, The Hague, Holland, June 1939, pp. 131-150.

[12] M. Bense, *Zeichen und Design*, Baden-Baden: Agis, 1971.

[13] P. Betts, "New Bauhaus und School of Design, Chicago", in J. Fiedler, P. Feierabend, (ed.), *Bauhaus*, Cologne: Könemann, 1999, 66-73.

[14] C. Poisson, "De l'objet au sujet; pour une sémiotique du projet en design. Charles Morris et le New Bauhaus", in I. Rauch (ed.) *Semiotics around the world: synthesis in diversity*, Berlin: Mouton De Gruyter, 1997, pp. 859-862.

[15] W. Nöth, *Handbuch der Semiotik*, Stuttgart: Metzler, 2000.

[16] E. Tarasti, *Signs of Music – A Guide to Musical Semiotics*, Berlin: Mouton de Gruyter, 2002.

[17] J. Nattiez, *Music and Discourse – Toward a Semiology of Music*, Princeton University Press, 1990

[18] N. Cummings, *The Sonic Self: Musical Subjectivity and Signification*, Bloomington: Indiana Univ. Press, 2000.

[19] E. Clarke, *Ways of Listening – An Ecological Approach to the Perception of Musical Meaning*, New York: Oxford University Press, 2005.

[20] D. Oswald, *Sound in Computer Interfaces*, unpublished thesis, Cologne University of Applied Science, 1996.

[21] D. Oswald, "Semiotik auditiver Interfaces – Zur Geschichte von Gestaltung und Rezeption auditiver Zeichen in Computer-Interfaces" in A. Schoon, A. Volmar (eds.), *Das geschulte Ohr. Eine Kulturgeschichte der Sonifikation,* Bielefeld: Transcript, 2012, pp. 247–267.

[22] C. Morris, "Signs, Language, and Behavior", in *Writings on a General Theory of Signs*, The Hague: Mouton, 1971, pp. 79-397.

[23] C. Morris, *Signification and Significance: A Study of the Relation of Signs and Values*, Cambridge: MIT Press, 1964.

[24] R. Keller, *A Theory of Linguistic Signs*, New York: Oxford University Press, 1998.

[25] G. Lakoff, M. Johnson, *Metaphors We Live By*, Chicago: The University of Chicago Press, 1980.

[26] U. Eco, *Einführung in die Semiotik*, Munich: Fink, 1972.

[27] J. J. Gibson, *The ecological approach to visual perception*, Hillsdale: Lawrence Erlbaum Associates, 1979.

[28] W. Gaver, *Everyday Listening And Auditory Icons*, PhD Thesis, University of California, San Diego, 1988.

[29] Aristotle, *Poetics*

[30] M. Black, Metaphor, in: *Proceedings of the Aristotelian Society*, vol. 55, pp. 273–294, 1954

[31] W. Gaver, "Synthesizing Auditory Icons", in *Proceedings of INTERACT '93 and CHI '93*, New York, 1993.

[32] H. de la Motte-Haber, "Die Musik als Affektlaut", in *Welt auf tönernen Füßen*, Forum, vol. 2, Göttingen, 1994.

[33] P. Janata, "The Highs And Lows Of Being Tone Deaf", in *Nature Neuroscience* 10, 2007, pp. 810-812.

[34] M. Wertheimer, "Untersuchungen zur Lehre von der Gestalt", in *Psychologische Forschung*, 4, pp. 301-350, 1923.

[35] S. M. Williams, Perceptual Principles in Sound Grouping, in: G. Kramer (Ed.) *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Reading, 1994, pp. 95–125

[36] M. Prensky, "Digital Natives, Digital Immigrants", in *On the Horizon*, MCB Univ. Press, vol. 9 no. 5, October 2001.