Overview of Auditory Representations in Human-Machine Interfaces

ÁDÁM CSAPÓ, Budapest University of Tech. and Economics / MTA SZTAKI
GYÖRGY WERSÉNYI, Széchenyi István University

In recent years, a large number of research projects have focused on the use of auditory representations in a broadened scope of application scenarios. Results in such projects have shown that auditory elements can effectively complement other modalities not only in the traditional desktop computer environment but also in virtual and augmented reality, mobile platforms, and other kinds of novel computing environments. The successful use of auditory representations in this growing number of application scenarios has in turn prompted researchers to rediscover the more basic auditory representations and extend them in various directions. The goal of this article is to survey both classical auditory representations (e.g., auditory icons and earcons) and those auditory representations that have been created as extensions to earlier approaches, including speech-based sounds (e.g., spearcons and spindex representations), emotionally grounded sounds (e.g., auditory emoticons and spemoticons), and various other sound types used to provide sonifications in practical scenarios. The article concludes by outlining the latest trends in auditory interface design and providing examples of these trends.

Categories and Subject Descriptors: H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems

General Terms: Design, Human Factors, Auditory Display

Additional Key Words and Phrases: Auditory icon, earcon, sonification

ACM Reference Format:
Csapó, Á. and Wersényi, G. 2013. Overview of auditory representations in human-machine interfaces. ACM Comput. Surv. 46, 2, Article 19 (November 2013), 23 pages.
DOI: http://dx.doi.org/10.1145/2543581.2543586

1. INTRODUCTION

Recently, a large number of research projects have focused on the use of auditory representations in a broadened scope of application scenarios. This is in part due to the fact that modern technological developments are creating applications in which the use of sound is expected to have a more central role than before. However, the novelty of many application scenarios is also forcing researchers and developers to take into consideration theoretical aspects that had not previously formed an integral part of auditory interface design theory.

For example, the growing prevalence of virtual reality environments, which were much less accessible when the theoretical foundations of auditory interfaces were first

---

This work is supported by the Universitas Győr Alapitvány in the framework TÁMOP 4.1.1./A-10/1/KONV-2010-0005. This research was realized in the frames of TÁMOP 4.2.4. A/1-11-1-2012-0001 "National Excellence Program—Elaborating and operating an inland student and researcher personal support system." The project was subsidized by the European Union and cofinanced by the European Social Fund.

Authors' addresses: Á. Csapó, Budapest University of Technology, H-1117, Budapest, Magyar tudósok krt 2, Hungary; G. Wersényi, Széchenyi István University, H-9026, Győr, Egyetem tér 1, Hungary.

DOI: http://dx.doi.org/10.1145/2543581.2543586

developed, has prompted the use of Virtual Audio Displays (VADs) to identify auditory scenes (or soundscapes in a wider sense) and to present the user with sound objects from the scenes by means of a playback system. This has been achieved by using either loudspeaker systems (e.g., through multichannel systems or loudspeaker arrays) or, in more typical cases, headphone-based playback. As will be described later in the article, all of these solutions introduce a number of new challenges, which are related to the perceptual specificities of sound localization.

The realization that audio will have an increasingly important role in computing environments has, on the one hand, sparked renewed interest in the basic auditory representations. On the other hand, given the fact that novel application scenarios are also creating new kinds of requirements and constraints with respect to the auditory interfaces that are being developed, researchers have begun to investigate how the basic auditory representations may be complemented by other approaches in order to meet the requirements of today's interfaces.

Basic auditory representations can be defined as unaltered or only slightly altered versions of human speech, music, or environmental sounds (both natural and human made). For example, sped-up human speech recordings, volume-adjusted or cut versions of instrumental music, natural recordings of everyday and environmental sounds (thunder, birds chirping, doors slamming, cars braking, etc.), or electronically created warning signals are examples of basic auditory representations. Newly designed versions of auditory representations can be derived from these representations by mixing, joining, hard editing, and compressing them, or by extending them with speech-based and emotional content according to various methods and paradigms.

In this article, we first provide an overview of existing methods for auditory representations and summarize some of the sound types that have been proposed more recently to cover previously unaddressed dimensions of nonspeech sound (i.e., how nonspeech sounds can be enhanced with speech-like and emotional characteristics, and how they can be used to provide navigation information, alert information, and other sonifications). This will be followed by an overview of two major use paradigms that have emerged in the design of auditory displays (we refer to these as the conceptual and interactive paradigms). Although the distinction between the two paradigms is not always clear-cut, we argue that they can be considered as two extremes in the application-oriented use of any combination of the more basic auditory representations. We also argue that an important advantage of adopting this point of view is that the rigid distinction between iconic (e.g., auditory icons) and abstract (e.g., earcons) representations becomes less relevant, allowing researchers to focus on unified approaches in the design of auditory interfaces. Finally, in a section on design approaches in applications, we focus on recent practical scenarios and trace the solutions in these use cases back to the theoretical sections of the article.

2. BASIC AND COMPOUND AUDITORY REPRESENTATIONS

The basic auditory representations can be derived from human speech, music, and environmental sounds. High-quality synthesized or prerecorded human speech can be used in some cases, but such sounds are language dependent and are relatively slow in communicating information, even if sped-up versions are used (e.g., this is often preferred by users with visual impairments). Further, it is difficult to modify the parameters of speech sounds without risking the loss of information. Although there certainly are applications where speech cannot be avoided (e.g., if a text document has to be read), narrative descriptions of the screen can be tedious. Speech is also not optimal for orientation, navigation, and manipulation tasks. Audio books and their revised versions (so-called audio films) may include sound samples and environmental sounds instead of speech (e.g., instead of saying "there is a clock ticking in the corner," the sound of the clock ticking might be embedded into the story) [Lopez and Pauletto 2009]. Other, more sophisticated solutions to using speech and/or human voice in iconic ways (e.g., as in spearcons and spindex representations) have emerged recently.

Both music-based and environmental sounds can be entertaining and can include dramatized and emotional content. However, in their simplest forms, neither of these sound types are well-suited to convey much information at the same time. Further, in the case of musical sounds, the listener has to learn the abstract mappings between sound and information; however, in the case of environmental sounds, the existence of a priori connotations often restricts the designer's ability to create free associations. Musical effects, short musical events, environmental sound events, or simple instruments can be used for iconic representations, but only with limitation.

Based on the preceding information, it was soon realized that good auditory representations mostly result from the structured combination of environmental sounds with musical and speech-based aspects. In this section, we provide a brief overview of the basic auditory representations and their evolution.

2.1. Classical Auditory Representations

2.1.1. Auditory Icons. Historically, auditory icons were the first basic auditory representations to appear in the literature and were introduced by Gaver [1986, 1988], followed by others [Blattner et al. 1989]. Auditory icons are short, icon-like sound events that have semantic connections to the physical events they represent. Auditory icons are easy to interpret and easy to learn. Users may connect and map (visual) events with the associated sound events after being exposed to an auditory icon for the first time. A typical example is the sound of a dot-matrix printer that is intuitively connected with the action of printing. Gaver [1989] provided many examples of easily learned auditory icons and also conducted pioneering research on the organization of auditory icons into "icon families" by considering the natural relationship between physical properties of objects and the events and processes that they generate. Based on this concept, applications were created that used auditory icons not only to label events and processes but also to reflect various physical properties of natural phenomena in the context of the virtual interface entities to which they are mapped (e.g., the sizes of files, the shapes of icons, the weight of interface elements as they are dragged) [Gaver 1989, 1991; Carello et al. 1998; Kunkler-Peck and Turvey 2000].

Environmental sounds provide a useful basis for auditory icons, because they are easily identifiable, learnable, and have a semantic-nomic connection to (visual) events. There are numerous factors that affect the usability of environmental sounds as auditory icons (a brief overview was provided in Gygi [2004]; Gygi et al. [2004]; Gygi and Shafiro [2009]). Among these are the effects of filtering on various types of environmental sounds. Some sounds are resistant against filtering and some completely lose their typical properties depending on their spectral content. Furthermore, some sounds are only identifiable after a longer period of time, and thus it is disadvantageous to use them as auditory icons. Ballas [1993] argued that a time period of about 200 to 600 ms is necessary for the proper recognition of a sound and would be a good start for creating an auditory icon. Last but not least, context contributes to recognition: logical, expected sounds will be recognized better than unexpected ones [Gygi and Shafiro 2009]. On the other hand, unexpected sounds do not have to be as loud to elicit attention and can thus be more appropriate for alerts. Realistic sounds are sometimes inferior to their less realistic but more familiar approximate versions. Thus, cartoonification, for example, may help to increase the applicability of a sound event (e.g., a gunshot sounds differently in movies than in real life) [Heller and Wolf 2002; Fernstrom and Brazil 2004]. For a more complete review of the state of the art in auditory icons, see Gygi and Shafiro [2009].

2.1.2. Earcons. Earcons were first introduced by Blattner, Sumikawa, and Greenberg [1989] as “non-verbal audio messages in the user-computer interface.” In contrast to auditory icons, earcons are message-like sounds (i.e., in the simplest case comprised of a serial succession of notes) that gain meaning through abstract relationships between signifier and signified. Because the relationship between sound and meaning is not obvious based on environmental experience, users are required to explicitly learn how earcons are linked with events in a system. This requirement induces a learning curve, which must be surmounted by the user.

A set of design principles for earcons was first established by Brewster [1994]. In a way analogous to the concept of (auditory) “icon families,” it was demonstrated that when designed well, even the musically uninitiated could recall up to 25 distinct earcons, if they were structured into a small number of conceptually and structurally distinct “earcon families” [Leplatre and Brewster 1998]. Research was also conducted to show how more than a single stream of earcons could be used in audio interfaces in order to provide more information to the user at the same time [Brewster et al. 1995; McGookin and Brewster 2004].

Despite the difficulties in learning abstract relationships between earcons and their meaning, there can be advantages to using earcons instead of auditory icons in certain cases. Because the relationship between earcons and the concepts they represent is abstract, earcons have a wider range of applicability than auditory icons, which may cause confusion when arbitrarily mapped to computational events. Further, Brewster [1998] and Edworthy and Hards [1999] have demonstrated in independent studies that the method used by users to learn earcons can have significant impact on recall accuracy. As a result, earcons are often considered as a viable alternative in developing audio interfaces.

2.1.3. Comparison of Auditory Icons and Earcons. Because the two sound types of auditory icons and earcons appeared in the research literature almost simultaneously (with earcons soon following the creation of auditory icons), and because they were proposed by researchers from widely different fields (i.e., Gaver is a cognitive psychologist and Blattner and her colleagues are computer scientists), the usability and effectiveness of earcons and auditory icons were initially considered separately for the most part.

More recently, several studies were published in which the applicability of auditory icons and earcons was compared based on several dimensions, such as learnability, memorability, and pleasantness. However, due to the large number of different application scenarios and evaluation methods, and difficulties in generalizing from specific sound designs, such results are often contradictory and lack generality. Thus, whereas there are results suggesting that auditory icons can be easier to learn and retain [Edworthy and Hards 1999; Bonebright and Nees 2007; Fagerlonn and Alm 2010], and that user reaction times to auditory icons can be shorter than to earcons [Graham 1999; Bussemakers and De Haan 2000], other researchers have demonstrated that earcons can be more pleasant in certain cases [Sikora et al. 1995], and that in any case, the users’ ability to retain different kinds of sounds depends heavily on the individual sound (not the sound type) as well as the learning method that is used [Brewster 1998; Edworthy and Hards 1999]. What is certain is that auditory icons and earcons seem to evoke different kinds of cognitive capabilities; therefore, at the very least, auditory icons can be associated with iconic entities more easily [Bussemakers and De Haan 2000; Bonebright and Nees 2007]. However, in other cases—where natural associations to physical entities are not as available, or when the users are allowed to create their own set of associations—earcons can be equally useful [Brewster 1998; Edworthy and Hards 1999].

With the large body of research on auditory icons and earcons showing no clear qualitative superiority of either sound type over the other, several researchers have argued that a strict discrimination between environmental and abstract sounds may be limiting for real-life applications, and that it may be useful to create interfaces in which this distinction becomes less sharp, or at least in which both sound types are used together [Gaver 1997; Hearst 1997; Mustonen 2008; Csapó and Baranyi 2011]. Thus, some researchers have suggested that auditory icons, earcons, and sonification techniques should be used in conjunction in real-world applications [Gaver 1997; Hearst 1997]. Others have suggested that auditory icons and earcons should be considered as theoretical extremes along a continuum of semiabstract nonspeech sounds [Mustonen 2008]. Still others have proposed a compositional relationship in which earcons are messages built up of auditory icons [Hermann and Ritter 1999; Singh 2010; Csapó and Baranyi 2011].

The first results that can be considered as pointing toward this change in viewpoint are the various studies performed on the applicability of auditory icons and earcons in different multimodal scenarios—at first, primarily scenarios in which vision and audition is necessary for interaction (for a review on the subject, see [Absar and Guastavino 2008]). The investigation of these scenarios was logical from a research historical point of view, because the sense of vision is used in an overwhelming majority of computer applications, and thus it was natural to investigate the effects of auditory icons and earcons in such settings.

Through such multimodal investigations, many researchers have come to the realization that sounds used in real-world scenarios demand different (i.e., more complex) design considerations than auditory icons and earcons used in isolation. As a result, several new sound types and research directions have emerged in research on auditory displays.

2.2. Auditory Representations with Speech-Based and Emotional Content

In this subsection, we provide a brief overview of the speech-based and emotional aspects of nonspeech audio interfaces. The auditory representations reviewed in this section can be viewed as in-between sound types between the two theoretical extremes represented by earcons and auditory icons.

2.2.1. Spearcons. Spearcons were first proposed by Walker et al. [2006], as sounds obtained by speeding up speech sounds—usually to the point where they are no longer recognizable as speech—while conserving their original pitch. Since they are usually unrecognizable as speech, spearcons are not simply sped-up sounds, but much rather are acoustic representations of spoken words (thus, spearcons are often described using a fingerprint analogy) [Jeon and Walker 2011].

Since the time they were originally proposed, it has been demonstrated through many experimental variations that spearcons are particularly well suited for audio-based navigation of menu structures in graphical user interfaces (GUIs). In the first study on the subject, Walker et al. [2006] compared earcons, auditory icons, and spearcons in menu navigation tasks and showed that the use of spearcons resulted in a statistically significant decrease in time-to-target performance measurements. Shortly after, it was demonstrated that learning rates for auditory menus were significantly shorter when using spearcons instead of auditory icons [Palladino and Walker 2007] and earcons [Dingler et al. 2008]. The learnability of these and other sound types were also compared in scenarios other than auditory menus, and spearcons were found to be equally superior to auditory icons and earcons in these applications [Dingler et al. 2008]. An interesting secondary conclusion of the work is that although hybrid icon-earcon sounds are subprime when compared to spearcons, they are better than auditory icons

AAAAABBBBB

Aaaaa Bbbbb

AaaaaBbbbb

A.....B.....

Fig. 1. Visual representation of normal, decreased, attenuated, and minimal spindex variations, in order from top to bottom. (Source: Jeon and Walker [2009].)

and earcons alone and can be even more useful in interfaces where localized audio is necessary [Dingler et al. 2008]. Although any such comparisons should be treated with caution (given the many different kinds of application scenarios and difficulties in generalizing design solutions, as mentioned earlier), these results demonstrate that spearcons can serve as an important component of auditory interface design.

In the practical sense, spearcons are time-compressed speech samples that are often names, words, or simple phrases. Regarding their generation, it was shown by Wersényi [2008] that useful quality measures and optimal compression ratios can be arrived at using spectral analysis. After presenting spearcons to visually impaired users, Wersényi concluded that the accessibility of spearcons can also be improved by changing parameters such as loudness and speed. Language independence was also investigated in the work by creating spearcons for the same events in different languages and accents [Wersényi 2008].

2.2.2. Spindices. Since their original appearance in the literature, several enhancements have been proposed to spearcons. These enhancements are mostly related to the spindex concept [Jeon and Walker 2009]. The term *spindex* stands for speech index, and the key idea behind the concept is to use accelerated initial sounds to provide direct information on the starting letter of the menu item currently navigated by the user (naturally, this idea is applicable when the user has to find items in alphabetically ordered lists). It was shown early that the benefit of using spindices is significant, especially in lists that are longer (the effect on lists with 150 items were considerable) [Jeon and Walker 2009].

Several variations of the spindex concept have been elaborated and tested [Jeon and Walker 2011]. Such variations include the attenuated spindex, in which the first occurrence of each spindex is louder than all the rest; the decreased spindex, in which the succession of the same spindex becomes gradually softer as the user traverses the list; and the minimal spindex, in which only the first occurrence of a spindex can be heard (for a graphical representation of these variations, see Figure 1).

Results showed that with very little practice, users were able to use all spindex types to their benefit, and that only the minimal spindex type caused perceived discomfort on the users' part [Jeon and Walker 2011].

2.2.3. Auditory Emoticons. Auditory emoticons are a vocally based analogy to graphical emoticons (i.e., "smileys"), generated based on auditory by-products of human emotional expression such as laughter, chuckling, or crying. To our knowledge, the concept of auditory emoticons first appeared in the literature in an article by Frohlich and Hammer [2004]. The authors used auditory emoticons in an audio-based automated e-mail reading application, but their tests regarding the usefulness of auditory emoticons were inconclusive, as only about half of their test subjects preferred auditory emoticons over abstract musical signals [Frohlich and Hammer 2004].

Table I. Visual Representations and Short Descriptions of Sound Events of the Most Important Emoticons Represented by User-Selected Male and Female Sounds

Auditory Emoticon	Visual Representation
Laughter (sound of laughter with open mouth)	:-D
Smile (sound of laughter with closed mouth)	:-), :)
Wink (smiling sound prefixed by a sparkling sound)	;-)
Mock ("beeeeh" sound of tongue pointed out in a mocking gesture)	:-P
Surprise (sound of a short "oh!")	:-o
Perplexed, distracted (longer female sound of "wheee?" and a short male version of "hmmm?")	:-S,
Sadness, sorry (female sound of "hmmm" and a longer male version of "oooooh")	:-(, :(,
Kiss (sound of a loud kiss)	:-*,
Disappointment (female sound of a longer "oh" and a short male version of "hm")	:-I,

(Source: Wersényi [2010].)

In recent investigations, Wersényi [2009a, 2010] conducted a more general set of experiments using auditory emoticons for both blind and sighted users, and results revealed that auditory emoticons were well received. Emoticons are widely used in e-mails, chat and messenger programs, forum posts, and the like. These different smileys and abbreviations (such as "brb," "rotfl," and "imho") are used so often that users suggested that there may be added value in representing emoticons using sound events as well. Auditory emoticons, then, are nonspeech human voice-based sounds, sometimes extended and combined with other sounds in the background. They can be compared to auditory icons, with the difference that they use human nonverbal voice samples with emotional load instead of a broader scope of environmental sounds.

In much the same way that smileys try to encapsule emotions in simple but limited (graphical) forms, auditory emoticons aim to achieve the same using brief sounds. In summary, it can be said that auditory emoticons:

(1) reflect the emotional status of the speaker,
(2) are always represented using nonverbal and language independent of human sounds,
(3) can be enhanced using sounds or noises that are outside of the scope of human speech-based emotional expressions in order to help give the user a deeper understanding of the intended meaning.

Although there is no scientific evidence that specific emotions can be represented better by a female voice than by a male voice, it was observed that subjects preferred female versions for smiling, winking, mocking, crying, and kissing. Table I shows the currently accepted visual representations of graphical emoticons and a description of their auditory emoticon counterparts. This dataset can also be downloaded with all the other auditory icons and earcons used in the research of Wersényi [2010, 2012].

All of the sound events in Table I are intended for use in auditory displays both for sighted and blind users to provide feedback information on a process or activity, as well as to help the user find a specific button, icon, menu item, or any other component of the graphical interface. Additionally, auditory emoticons can help the user gain a better understanding of the emotional content of the information provided in a conversation and can also provide an extra source of amusement.

2.2.4. Spemoticons. Spemoticons were first defined by Németh et al. [2011] as text-to-speech (TTS) technology–based auditory representations for emotional and intentional states. Although spemoticons are obtained using TTS technologies, they are just as unintelligible in a linguistic sense as spearcons. However, whereas spearcons are created based on existing words, spemoticons are acoustic events synthesized based on meaningless vocalized expressions that do not occur in real life. Spemoticons are obtained by modifying the intensity, pitch, and temporal structure (by inserting breaks of various lengths) of TTS synthesized phrases [Németh et al. 2011]. The advantages of spemoticons include the possibility of real-time generation and the directness with which they map emotions to sounds. A disadvantage is that the interpretation of spemoticons can be culturally dependent.

In their seminal work on spemoticons, Németh et al. [2011] proposed an intuitive, parameter-based design methodology for the creation of spemoticons. The methodology involves the use of an interactive interface to a professional TTS system in which a number of parameters, such as the lengths of pauses, word-level pitches, positions in sound, durations, and amplitudes can be specified independently by the user.

Based on the proposed design approach, Németh et al. [2011] explored the viability of spemoticons by asking test subjects to categorize 44 different sounds into 7 categories. Each category reflected an imaginary message and an associated emotion (e.g., “Continue, I like this.” or “I hate you. This bothers me.”). Those categories to which only a few sound samples were attributed were regarded as valid emotional categories, and the associated sound samples were declared by the authors as good examples of spemoticons.

2.3. Sonification: Alerts, Reminders and Navigation

Sonification was defined by Kramer as “the use of non-speech audio to convey information” [Kramer 1993; Hermann et al. 2011, p. 149]. Although the purpose of sonification in general is to transform any kind of data into sound, we focus in particular on those representations that have been used to provide alert information, reminders, and navigation information. In this subsection, a brief overview is provided of the various auditory representations that have been used for sonification in general, followed by a brief overview of example applications.

2.3.1. Musicons. Musicons have been recently defined as extremely brief samples of well-known music [McGee-Lennon et al. 2011]. Musicons were proposed as audio signals that can be used to provide reminders in a large variety of scenarios, both at home (private space) or at work (public space). From the perspective of these goals, musicons can be conceived as being conceptually situated between auditory icons and earcons: they build on familiarity more than earcons but less than auditory icons; on the other hand, they are more private than auditory icons but less private than earcons (earcons are only understood by those who have learned the mapping from concepts to sounds, whereas auditory icons can be easily understood by anybody).

Musicons are usually generated by sampling a well-known musical piece or song and then cutting out short sections that are a few hundred milliseconds long. Recent investigations have shown that effective musicons can be created simply by sampling user-selected parts of music [McLachlan et al. 2012]. Thus, it is suggested that musicons can be created by the user from his or her own musical database. In a set of experiments, users were asked to select 5-second-long parts of their favorite songs, which were then labeled as “favorite part” and “most representative part” for the actual song. After sampling (cutting) these selections into 0.2- and 0.5-second-long bits, recognition rates of 69 to 78% and 84 to 94% were measured for the two lengths, respectively. Based on the evaluation of the structure, timbre, and melodic and rhythmical patterns of the selections, it was found that usually the first part of the chorus/verse, main riff, or solo can be recognized the best. It has also been suggested that following the idea behind spearcons, using some kind of compression instead of just sampling the original audio track might lead to further useful musicons. Investigations on music other than Western popular music, as well as comparisons between earcons, auditory icons, and musicons, remain to be conducted.

2.3.2. Morphocons. Morphocons, or morphological earcons, have recently been proposed as a novel solution to the problem of creating earcons and earcon families that can be customized to users' preferences [Parseihian and Katz 2012]. It is noted in the study that whereas performance and efficiency of solutions in auditory display are often analyzed, the satisfaction of users is rarely addressed, despite the fact that those who are in the greatest need of sonification (e.g., the visually impaired) often have very specific—and also largely variable—requirements.

To resolve issues of user satisfaction, morphocons have been proposed as a special kind of earcon in which various dynamical properties of sound (which are allowed to change any number of times even within a single morphocon) are used as parameters to generate a wide range of sound palettes. For example, the same dynamical parameters related to, for example, the envelope of the sound, the harmonic properties of the sound, and the rhythmic properties of the sound can be mapped onto natural sounds, musical sounds, synthesized sounds, and so forth. In this way, although the general properties of sound remain constrained, the specifics remain to be determined based on the individual taste of the user.

Morphocons have been used as components in an audio-based navigation system in order to help the visually impaired learn and recognize points on their paths related to the itinerary and obstacles, as well as landmarks and other points of interest [Parseihian and Katz 2012].

2.3.3. Alerts and Warning Signals. The problem of urgency encoding is concerned with the task of mapping warning signals of different levels of urgency onto sound. Although warning signals have been used to alert the operators of computerized systems for as long as such systems have been in existence, the importance of auditory warning signals in particular, as well as the necessity to design them in structured ways, was realized only in the 1980s.

The first set of guidelines for the design of auditory warnings, originally in the context of aviation and later in a more general context, were laid down by Patterson et al. [Patterson 1982, 1989; Patterson and Mayfield 1990]. The guidelines include rules on the optimal loudness of warning sounds relative to levels of noise of different frequency domains and the optimal spectral distribution of warning sounds, as well as temporal characteristics such as pulse-repetition rates and rhythms. In the more general case—unrelated to aviation or any application in particular—Patterson et al. demonstrated that good auditory warning signals consist of a series of bursts—with each burst comprised of a number of repetitive pulses—and that the starting points and intensities of these pulses should be different in each burst.

Although Patterson et al. also suggested basic rules for encoding relative urgency within warning sounds, detailed and conclusive analysis was carried out only later. In subsequent studies, Edworthy, Hellier, and others have shown that a number of parameters, such as fundamental frequency, harmonic composition, envelope shape, and delayed harmonics, have significant effects on the perceived urgency of warning sounds [Edworthy et al. 1991; Hellier et al. 1993]. Further, it was shown that some parameters contribute more to perceived urgency than others [Hellier et al. 1993]. For example, it was found that the speed and repetition of pulses (i.e., within the bursts defined by Patterson) contributed more significantly than other parameters to the perceived urgency of sounds [Haas and Edworthy 1999; Brock et al. 2005].

2.3.4. Sonification for Navigation Purposes. Sonification is often used in assistive technologies for users with visual impairments, ranging from temporary loss of vision to long-term, serious visual impairments. Technologies of this kind aim to address the basic level of everyday problems and needs in satisfactory ways.

Electronic Travel Aids (ETAs), sometimes also referred as O&M (Orientation and Mobility) devices, include mobile and wearable devices that increase navigation safety and mobility using auditory display technology [Dobrucki et al. 2010; Loomis et al. 2005; Hersh and Johnson 2008; Dakopoulos and Bourbakis 2010; Kay 1984; Ebling 2009; Ventura and Fernandes 2011; Jameson and Manduchi 2010; Lahav et al. 2008]. The most important requirements for these devices include unencumbered interaction (e.g., interaction with free hands and uncovered ears during operation), small size, light weight, cheap price, and ease of use.

Devices such as the Mowat Sensor [Pressey 1977; Morrissette et al. 1981], SonicGuide [Kay 1973], Navbelt [Bourbakis 2008], LaserCane [Murphy 1971], and Tyflos System [Bourbakis 2008] operate using ultrasound and/or laser devices in order to help users avoid collision with obstacles while providing feedback on the surroundings through vibrations and/or sounds [Ando and Graziani 2009; Ando 2008]. Because lasers are quite expensive and ultrasound detectors cannot be used in case of too much reflection, alternative solutions use amplitude-modulated radiofrequencies with different carriers to detect obstacles [Debnath et al. 2001]. For example, the Smart Cane is an RFID-enabled navigation cane [Kahol et al. 2004] that incorporates an ultrasonic sensor that works in conjunction with navigation and senses RFID tags mounted on flags placed on the ground (e.g., along a sidewalk). Besides using such sensors to map distance onto various frequency and amplitude ranges, extended information provided by complementary environmental actuators (e.g., vibrotactile equipment) allows users to obtain enhanced sensations of the 2D/3D space.

Camera-based systems have also been used to transfer images onto auditory dimensions (e.g., vOICe [Meijer 1992] and Prosthesis Substituting Vision with Audition—PSVA [Capelle et al. 1998]). In these systems, the horizontal and vertical dimensions of images are continuously mapped onto frequency-based and temporal aspects of sound.

In the domain of wearable computers, the System for Wearable Audio Navigation (SWAN) was developed to serve as a safe pedestrian navigation and orientation aid for persons with temporary or permanent visual impairments [Walker and Lindsay 2006; Wilson et al. 2007]. SWAN consists of audio-only output and tactile input via a task-specific handheld interface device. Emphasis is placed on representing pertinent data with nonspeech sounds through a process of sonification. Once the user's location and head direction are determined, SWAN guides the user along the required path using a set of beacon sounds, while at the same time indicating the location of features in the environment that may be of interest to the user. More specifically, the sounds used by SWAN include navigation beacons (earcon-like sounds), object sounds (through spatially localized auditory icons), surface transitions, and location information and announcements (brief prerecorded speech samples).

3. CONTRASTING USE PARADIGMS IN THE DISPLAY OF NONSPEECH SOUNDS

In this section, we describe two major use paradigms that have emerged since the original appearance of auditory icons and earcons in the literature. The first of these two, which we refer to as the conceptual paradigm, makes use of the contextual and task-related background of the use-case scenario in order to provide users with rich auditory information through the auditory display. The second, which we refer to as the interactive paradigm, focuses on natural forms of interaction with the sound-producing model rather than on the use-case scenario of an application. After a brief description of the two paradigms, we provide several application examples. As we will aim to demonstrate, the difference between the conceptual and the interactive paradigms comes from the way they view the concept of information. Because of this difference in viewpoint, the two paradigms arrive at largely different kinds of systems, which can complement each other well in specific applications.

Before we introduce the two paradigms, it is important to note that other researchers have also defined taxonomies of design methodologies in auditory display. In a recent example, Frauenberger and Stockman [2009] have listed as many as seven different approaches to the design of auditory displays. The two major paradigms that we highlight in the following sections are both a unification and an extension of these categories (namely, the conceptual paradigm can be considered as a unification of various aspects of contextual design, task-driven design, and semiotics, whereas the various aspects of what we refer to as the interactive paradigm are included only implicitly in earlier taxonomies).

3.1. The Conceptual Paradigm

The information concept at the heart of the conceptual paradigm views information as an entity that provides *an answer to a question* [Bertin 1981; Barrass 1998]. The conceptual paradigm views the world as a place where questions that demand answers arise frequently and naturally, and thus the goal of the audio interface is to play messages that provide answers to a these questions. The questions that can be asked in systems developed according to the conceptual paradigm are in general fixed a priori, and the goal of the system is to convey this information to the user as clearly and concisely as possible.

Approaches in the conceptual paradigm generally emphasize the use-case scenario in which the auditory display will be used, and the task that must be performed by the user with the help of the auditory display. Indeed, such aspects form an important part of the categories of contextual design, task-driven design, and semiotic design, as pointed out in Frauenberger and Stockman [2009]. Another common trait among approaches in the paradigm is that they tend to use both iconic (e.g., auditory icons) and abstract (e.g., earcons) sounds together and according to various hybrid models. Later in the section on examples of design approaches, we highlight some of the ways in which such hybrid platforms have been created.

3.2. The Interactive Paradigm

The interactive paradigm—as opposed to the conceptual paradigm—views information as an entity that exists irrespective of any predefined set of questions that may be asked in relation to a task. In the case of the interactive paradigm, information is extracted from the system in less structured, less task oriented forms and in the end reflects the subjective actions of the user relative to the system.

A convincing argument supporting the need for the interactive paradigm is given in Hunt and Hermann [2004]. The key point of the argument is that during everyday multimodal interactions, the types of data we receive on physical processes and the order in which we are confronted with the data are rarely predetermined and objective. Instead, our real-life interactions follow the pattern of *awareness*, *interaction*, *multimodal rechecking*, and *confirmation*. Thus, the environment does not impose on us a notification on every event in a predetermined way, along with all of the statistical data describing the event. Instead, we discover events in an autonomous way based on the effects they produce (*awareness*). Once we are aware that something of interest has occurred, we manipulate the environment so as to tease out different effects to help us identify an underlying cause (*interaction* and *multimodal rechecking*). The elicitation of precise, objective measurement data occurs only during the final step (*confirmation*), in which we confirm that the underlying cause of what we perceive during interaction is indeed the event whose occurrence we initially conjectured [Hunt and Hermann 2004]. Thus, although in the end information in the interactive paradigm may also provide an answer to a question, its communication is evoked in a way that reflects the user's actions.

Approaches in the interactive paradigm generally create a basic sound model and emphasize the ways in which users can interact with the model in order to produce various kinds of sounds. This is in contrast to the conceptual paradigm, which begins the design process by making assumptions about the use-case scenario in which information will be provided as an answer to a question. Naturally, a system that is categorized as belonging to the interactive paradigm can also be used to provide answers to concrete questions (without this possibility, it would not qualify as an auditory display). However, the set of possible questions that the user might ask through the system would be contained in the system only implicitly, and answers to those questions would emerge only as a result of the subjective actions and discovery process of the user.

3.3. General Comparison

As described in the previous two subsections, the conceptual and interactive paradigms represent different approaches to the way in which information is viewed in an application. It is important to note that the two paradigms can be regarded more as two extremes than as two distinct categories. If an application views information as "an answer to a question" but still leaves room for open-ended interaction such that information is provided in ways that depend on how the user interacts with the system, then it is clear that the application represents a hybrid approach. An example of such a system would be a GUI in which objects (e.g., icons, windows, scrollbars) are modeled as physical entities with their own set of physical properties (e.g., sizes, weights, materials), such that various modes of interaction would produce sounds that highlight different aspects of those properties, in ways that vary depending on the interaction itself (hence, the interface would be designed with the goal of providing answers to a clear set of questions, such as "what kind of object is this?" whereas interaction with the interface would be exploratory in nature and could influence the way in which information is presented).

Nevertheless, it is useful to consider recent design approaches based on the distinction between conceptual and interactive for two reasons: first, such a distinction de-emphasizes the question of what kind of basic auditory representations should be used in an application (i.e., iconic or abstract, message like); second, as demonstrated later in Section 4, there certainly are applications that are characteristically conceptual and others that are characteristically interactive, and the two kinds of approaches represent different forms of interaction between the user and the system.

4. TASKS AND APPROACHES FOR AUDITORY DESIGN

This section highlights recent design approaches used to create auditory displays based on a combination of the basic auditory representations. Examples of approaches that can be categorized as belonging to the conceptual paradigm and the interactive paradigm are given in separate subsections. As the design of conceptual and interactive auditory displays can be motivated by widely different considerations, the examples given in this section also cover a wide range of application scenarios. However, all of the examples provided are common in that they describe or support the use of sound in human-machine interfaces.

4.1. Design Approaches in the Conceptual Paradigm

4.1.1. Evolution of GUIs toward VADs. Audio in GUI environments was first considered in the 1980s and 1990s, when the growing market for personal computers highlighted the importance of accessibility for those living with visual or other impairments. Since then, providing enhanced accessibility has become one of the most important areas in research on assistive technologies for computer sciences and for human-computer interaction. More generally, enhanced accessibility today concerns more than just basic and assistive technologies due to the growing popularity of computer-based entertainment (e.g., media and gaming). Especially the growing complexity of the gaming industry has motivated research on the enhanced accessibility of virtualized environments and augmented realities [Fish 1976; Harness et al. 1993; McKiel 1992]. Today, audio is being considered in a wide range of physical/virtual/augmented systems, and the following subsections provide a brief summary of traditional and more state-of-the-art solutions.

Audio in Traditional GUI Environments. GUIs provide the simplest means of communication with modern computing technologies. Personal computers, as well as mobile phones, television systems, and even motor vehicles, can be configured and controlled through hierarchical menu structures; spatially distributed and visually unique icons (through variations of form, size, and color); and simple control tools such as scrollbars, touch-screens, the mouse, or joystick. In these 2D graphical environments, users without visual impairments are capable of quickly and easily orienting themselves among large amounts of information; however, this is much more difficult for users with visual impairment. For this reason, investigations as early as in the 1990s tried to establish auditory interfaces and environments for the visually impaired using a wide variety of methods.

SonicFinder was an Apple program that aimed to integrate auditory icons into the operating system for file handling [Gaver 1989]. In the Mercator project, Mynatt [1997] presented a transformed hierarchical graphical interface, utilizing auditory icons, tactile extensions, a TTS module, and a simplified structure for navigation. The hierarchical structure was thought to best capture the underlying structure of a GUI. The Mercator project used filtering and frequency manipulations to portray screen events (e.g., the appearance of pop-up windows, the selection of items, and the number of objects that appear on the screen).

The most important applications today and presumably in the near future are the so-called screen-readers or TTS applications that simply read the content that can be seen on the screen. Today's speaking modules offer good synthesized speech quality, but they are language dependent and are only optimal for reading textual information.

Audio in Directionally Informed GUI Environments. Later, in the GUIB project (Graphical User Interface for Blind persons), a multimodal interface was proposed using tactile keyboards (Braille) and spatially distributed sound (at first, loudspeaker playback was provided using the so-called sound-screen, but later this was substituted with headphone playback and virtual simulation via Head-Related Transfer Functions (HRTFs) [Crispien and Petrie 1993; Petrie and Morley 1998; Wersényi 2003, 2007a, 2007b, 2009b; Liard and Beghdadi 2001]. Research showed that spatially distributed sound events could be used in special window arrangements and in different resolutions according to the users' experience and routine. The first tests, performed on blind children, supported the conclusion that audio can help users gain an understanding of 2D and 3D structures [Lahav et al. 2008].

Expansion toward 3D Interfaces and VADs. The graphical elements used in everyday interfaces have recently begun to change. Although simple mobile devices and cell phones still make extensive use of hierarchical menus in which the menu items are related to each other in a parent-child structure, a wider range of possibilities are being offered by more advanced—especially 3D—GUIs. As a result, it is again becoming difficult for the visually impaired to access computing technologies at more than a very basic level [Boyd et al. 1990; Nees and Walker 2007]. In addition, the growth in demand for heavily audio-oriented (online or other) games, audio-films (e.g., both for entertainment and education), and wearable and other mobile applications is highlighting the need for the effective replacement of visual information by audio. This is increasingly necessary not only to enhance accessibility but also to allow users to perform virtual or semivirtual localization tasks to find out how virtual environments and the real-life environment can interact with each other [Loomis et al. 2005; Hersh and Johnson 2008; Kay 1973; Cardin et al. 2007].

A rapidly evolving solution in this direction is to use VADs to identify auditory scenes (or soundscapes in a wider sense) and present the user with sound objects from the scenes by means of a playback system. This has been achieved by using either loudspeaker systems (e.g., through multichannel systems or loudspeaker arrays), or in more typical cases headphone-based playback. Many, but not all, VADs use directional information to provide the user with an understanding of where objects are located in the auditory scene. Directional information can be reproduced more or less accurately by allocating different components of the sound to spatially distributed locations of the loudspeakers, or by making use of interaural differences and/or directional filtering in headphone-based systems [Wenzel 1992; Wenzel et al. 1994]. Although several sources of error—such as in-the-head localization, front-back reversals, and elevation shift [Moller 1992; Begault et al. 2001]—exist in the latter case due to the difficulty in predicting the user's head movements, successful applications have been appearing steadily. HRTFs in particular are used extensively to implement directional filtering in headphone-based systems [Blauert 1983; Moller et al. 1995; Cheng and Wakefield 2001], as well as to help achieve increased vertical localization in Wave-Field Synthesis (WFS) applications designed for loudspeaker systems [López et al. 2010].

4.1.2. Assistive Technologies.

Design approaches in assistive technologies can be categorized as belonging to the conceptual paradigm, as they generally view information as "an answer to a question." Any audio-based solution that focuses on the rehabilitation or support of cognitive capabilities in general can be cited here.

As described in Section 2.3.4, sonification is often used for navigation purposes in ETAs based on any of a number of approaches. Many solutions provide a direct mapping of echoes to audible sound (using ultrasonic echoes as in the SonicGuide system [Kay 1973] and infrared signal–based echoes as in the LaserCane System [Murphy 1971]). Other systems translate horizontal and vertical dimensions of the image into frequency-based and temporal aspects of sound [Meijer 1992; Kim and Zatorre 2008; Capelle et al. 1998]. Navigation systems such as SWAN make use of a combination of auditory icons, earcons, and short prerecorded bits of speech to provide users with beacon sounds as well as contextual information on the environment [Walker and Lindsay 2006; Wilson et al. 2007].

Sonifications have also been used to provide feedback on contact forces in telemanipulation tasks [Massimino 1992, 1995]. In these solutions, a direct mapping was used between contact force magnitude and loudness, although the latter was modeled to high precision based on the subjective experience of sound intensity. Further studies have focused on modeling various contact surfaces and transferring them to audio representations [Gaver 1993; O'Brien et al. 2002; Van den Doel et al. 2001].

Various design methods have been used to help rehabilitation patients in learning the correct limb and joint movements necessary for everyday life. Sonification has been used, for instance, to substitute proprioceptive feedback—normally obtained from muscles and joints—using sound. Solutions include the mapping of vertical distance from a target to pitch and horizontal distance from a target to amplitude, as well as mapping of joint movement direction to the directionality of a melody (rising and falling) and the required temporal structure of movements to metrical beats [Ghez et al. 2000]. Further studies on audio-based feedback have been conducted related to rehabilitation with prosthetic hands [González et al. 2010], as well as vestibular rehabilitation to help stroke patients and other patients with disabilities [Dozza et al. 2004; Basta et al. 2008].

Finally, sonification has been used in the context of elite sports to help guide athletes away from inefficient movements toward the correct, efficient ones. For example, the acceleration-time trace of rowing boats was translated to sound—by mapping acceleration to pitch—in order to provide rowers with a sense of their performance at a higher temporal resolution than pure vision would have allowed [Schaffert et al. 2010; Schaffert et al. 2011; Schaffert et al. 2012]. Similar studies have focused on other sports with repetitive movements (e.g., ice skating, dance, and aerobics) to provide users with a sense of how well they are doing compared to a reference model [Godbout and Boyd 2010; Jylha and Erkut 2011; Hermann and Zehe 2011].

4.1.3. Task-Oriented Design Frameworks in the Conceptual Paradigm. TaDa! and EarBenders are two complementary approaches to conceptual auditory interface design proposed by Barrass [1998, 1996a, 1996b]. In broad terms, TaDa!—which stands for task-oriented and data-sensitive auditory information design—provides a structured way of finding the meeting point between information requirements of a task and the information representation used to achieve the designer's goals. EarBenders, in turn, is a framework that builds on the TaDa! approach in order to enable users and designers of auditory interfaces to share, through informal stories, their knowledge and experience regarding real-life examples in which environmental and other sounds helped them achieve their goals.

The use of EarBenders consists of the four steps of situation description, situation analysis, example lookup, and design synthesis. The first two steps can broadly be equated with the TaDa! approach. In general terms, application design begins with a functional description of the use-case (in terms of questions that can arise in the use-case and possible sets of answers to those questions) and leads to a fully functional design synthesis based on an analysis of the use-case and comparison based on that analysis with existing solutions.

It is clear based on the earlier discussion that the TaDa! approach and the EarBenders framework are prime examples of the conceptual paradigm, as they both focus on the analysis of information requirements and the structured design of auditory interfaces such that the conveyed information provides an answer to a question. Further, these two approaches represent a well-developed framework in which a number of different kinds of sounds (e.g., speech, nonspeech, environmental, and abstract sounds) can be used parallel to each other.

4.2. Design Approaches in the Interactive Paradigm

4.2.1. Sound Authoring Tool. The Sound Authoring Tool, developed by Bezzi, De Poli, and Rocchesso, is an example of a sound design tool that encapsulates the viewpoint of the interactive paradigm [Bezzi et al. 1999; Rocchesso et al. 2003]. The concept of sound objects, which serves as the core of the framework, was first proposed by Schaeffer [1966] and further developed by Schafer [1977]. In Schaeffer's definition, sound objects are musical objects perceived as a single, abstract entity and can be categorized through their differences in one or more perceptual properties. Despite the fact that even today we have only available a few properties of sound that can be characterized unequivocally and with certainty (e.g., loudness, pitch, and brightness), Bezzi, DePoli, and Rocchesso have shown that the concept of sound objects can be useful nevertheless if similarity is defined based on properties of the models that are used to generate sounds—in a way similar to how the same model can be used to generate families of auditory icons, earcons, morphocons, and the like ("sound control can be more straightforward if we generate sounds with [...] techniques that give access to control parameters directly connected to sound source characteristics" [Rocchesso et al. 2003, p. 46]).

The model proposed by Bezzi et al. [1999] involves three layers: a physical layer that specifies a physical model used to generate the sounds (i.e., the identity of the sound), a signal layer that depends on the physical layer and specifies a number of sound parameters that may be altered by the user (i.e., it allows the user to change the quality of the sound), and finally a geometric layer that defines acoustic aspects of the space in which the sound exists (i.e., its reverberation, spatialization, etc.).

The overall goal of the Sound Authoring Tool is to support the creation of nonspeech sounds that convey dynamic and multidimensional information. While working toward this goal, the user might perform various different kinds of operations in order to gain an intuitive understanding of the sound-generating model. For example, the physical layer—which is essentially a black box model—has a nontrivial influence across all possible combinations of parameters in the signal and geometric layers, and the characteristics of the layer must be elicited by the user through a series of open-ended, exploratory interactions, as described by Hunt and Hermann [2004]. The possibility for these interactions is contained within the Sound Authoring Tool; however, the specific ways in which information is extracted from the model are not specified explicitly.

4.2.2. The Spiral Discovery Method. A more generic model was proposed recently by Csapó and Baranyi [2011, 2012]. Although the model does not incorporate physical metaphors as does the Sound Authoring Tool, it enables users to interactively explore parametric sound spaces based on the perceptual qualities of the generated sounds. The goal of the Spiral Discovery Method (SDM) is to provide the user with a cognitive artifact that can be used to create sequences of sounds that are perceptually orderable, irrespective of whether these sounds are auditory icons, earcons, or other sound types. As the authors argue, creating such a sequence may be simple in certain cases (e.g., when primitive parameters such as pitch or loudness are taken as a basis for perceptual continuity), but in most cases the structure of the parameter space used to generate the sounds is high dimensional and nonlinear. In the latter case, it can be difficult for the user to understand the relationship between a set of parameter values and the perceptual qualities of the resulting sound without the use of structured methods that rely on a set of simplifying assumptions.

During the use of SDM, the user is allowed to perform the following actions in any order and any number of times:

(1) Assemble an arbitrary succession of (not necessarily perceptually continuous) sounds that are deemed to be increasing or decreasing along an arbitrary perceptual scale (examples of perceptual scales can include auditory roughness, auditory softness, harmonic complexity, timbre-based scales, etc.).
(2) Change the values of the parameters that are used to generate individual sounds in the arbitrary succession of sounds.
(3) Transform the mentioned tuning parameters into a single parameter that allows the user to explore the sound space in a simplified but structured way. The user then

is allowed to use the single parameter to travel along a hyperspiral in the high-dimensional parameter space—in much the same way as changing the distance along a 3D spiral could be used to travel through parts of a 3D space (hence the name Spiral Discovery Method).

SDM allows the user to trade off the complexity of the original system (the high-dimensional nonlinear parameter space) and the interpretability of a rank-reduced system (a parameter space consisting of only a single parameter). In much the same way as Bezzi, de Poli, and Rocchesso's Sound Authoring Tool, SDM provides the user with an interactive way to explore open-ended sound spaces. Whereas the Sound Authoring Tool keeps the user's actions between constraints by adopting a layered architecture in which each layer imposes increasing constraints on the sound, SDM keeps the user's actions between constraints by allowing the use of a reduced parameter space and hiding from the user the periodical changes of those components of the parameter space that are less important (i.e., the direction and radius of the hyperspiral). The two models are common in that the user is confronted with a black box model from which various sound qualities can be elicited depending on the user's actions.

SDM can be used in conjunction with sound-producing models to assist the user in finding relationships of order between various parameter configurations. However, the precise way in which this is done is not fixed by SDM and emerges as a result of the user's interactions with the model.

4.2.3. Interactive Model-Based Sonification. Interactive sonification was defined by Hermann and Hunt [2005] as "the use of sound within a tightly closed human-computer interface where the auditory signal provides information about data under analysis, or about the interaction itself, which is useful for refining the activity." Although the first aspect of the definition is true of many approaches, the second aspect—pertaining to gaining information about the interaction itself—is a distinguishing characteristic of interactive sonification.

Model-based sonification (MBS) is an increasingly relevant approach within interactive sonification [Hermann 2002; Hermann and Ritter 1999]. An MBS application consists of a setup of dynamical elements, a dynamics (or a set of *virtual physical laws*), and an interaction interface. When the goal is to sonify existing data, MBS works by allowing the user to excite the setup of dynamical elements through a set of possible interactions (e.g., by plucking, hitting, rubbing, or scratching the dynamical system) [Hermann 2002]. Through these interactions and an understanding of the virtual physical laws that govern the excitations, the user can obtain an intuitive feel for how the sonified multidimensional dataset is structured. Conversely, when the goal is to create sounds that can be utilized in an auditory display, the basic scheme of MBS can be applied in reverse direction so that the user is essentially required to search for datasets that provide interesting sounds.

Recently, a number of research directions have appeared based on the foundations of interactive sonification. For example, several authors have focused on providing emotional information and behavioral data using interactive sonification in augmented/virtual reality environments [Legroux et al. 2007; Kummer et al. 2012]. In these cases, the goal is to create a more realistic, immersive user experience. The work of Bovermann on tangible auditory interfaces is also a recent example of interactive sonification, which can in a certain sense be viewed as an extension to MBS in particular [Bovermann et al. 2006; Bovermann 2009]. Tangible auditory interfaces can be obtained by substituting the setup of dynamical elements and the virtual physical laws of an MBS system with physical objects and a set of interactions made possible by those objects. When using a tangible auditory interface, the user would typically hold and manipulate a physical object and generate sounds through these manipulations and the mediation of any number of physical sensors attached to the physical object. As in the case of other solutions belonging to the interactive paradigm, this is done without any specific data-oriented communication in mind between the user and the system.

5. CONCLUSIONS

In this article, our goal was to summarize the history of research on auditory representations in computing environments. We began by providing an overview of the elementary sound types such as auditory icons and earcons, as well as a number of more recently developed speech-, emotion-, and sonification-based sound types. This was followed by an overview of trends in the way in which the various basic sound types are used together. We distinguished between two major use paradigms: the conceptual paradigm, which views information as an entity that has a well-specified meaning that must be communicated to the user with a specific purpose in the context of a predefined use-case scenario, and the interactive paradigm, which lays more emphasis on natural forms of interaction and on gaining information through the user's exploratory actions with respect to a sound-producing model. In the final section on design approaches, a broad overview was given of a number of existing applications within the conceptual and interactive paradigms.

ELECTRONIC APPENDIX

The electronic appendix for this article can be accessed in the ACM Digital Library.

REFERENCES

ABSAR, R. AND GUASTAVINO, C. 2008. Usability of non-speech sounds in user interfaces. In *Proceedings of the International Conference on Auditory Display*.

ANDO, B. 2008. A smart multisensor approach to assist blind people in specific urban navigation tasks. *IEEE Trans. Neural Syst. Rehabil. Eng. 16*, 6, 592–594.

ANDO, B. AND GRAZIANI, S. 2009. Multisensor strategies to assist blind people: A clear-path indicator. *IEEE Trans. Instr. Measurement 58*, 8, 2488–2494.

BALLAS, J. A. 1993. Common factors in the identification of an assortment of brief everyday sounds. *J. Exp. Psychol. Human 19*, 2, 250–267.

BARRASS, S. 1996a. TaDa! Demonstrations of auditory information design. In *Proceedings of the International Conference on Auditory Display*, 6 pages.

BARRASS, S. 1996b. EarBenders: Using stories about listening to design auditory interfaces. In *Proceedings of the 1st Asia-Pacific Conference on Human-Computer Interaction (APCHI'96)*. 525–538.

BARRASS, S. 1998. Auditory information design. Ph.D. thesis. Australian National University.

BASTA, D., SINGBARTL, F., TODT, I., CLARKE, A., AND ERNST, A. 2008. Vestibular rehabilitation by auditory feedback in otolith disorders. *Gait Posture 28*, 3, 397–404.

BEGAULT, D. R., WENZEL, E., AND ANDERSON, M. 2001. Direct comparison of the impact of head tracking reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc. 49*, 10, 904–917.

BERTIN, J. 1981. *Graphics and Graphic Information Processing*. Walter de Gruyter and Co.

BEZZI, M., DEPOLI, G., AND ROCCHESSO, D. 1999. Sound authoring tools for future multimedia systems. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*. 512–517.

BLATTNER, M. M., SUMIKAWA, D. A., AND GREENBERG, R. M. 1989. Earcons and icons: Their structure and common design principles. *Hum.-Comput. Interact. 4*, 1, 11–44.

BLAUERT, J. 1983. *Spatial Hearing*. MIT Press, Cambridge, MA.

BONEBRIGHT, T. AND NEES, M. 2007. Memory for auditory icons and earcons with localization cues. In *Proceedings of the International Conference on Auditory Display*. 419–422.

BOURBAKIS, N. 2008. Sensing surrounding 3-D space for navigation of the blind. *IEEE Eng. Med. Biol. Mag. 27*, 1, 49–55.

BOVERMANN, T., HERMANN, T., AND RITTER, H. 2006. Tangible data scanning sonification model. In *Proceedings of the International Conference on Auditory Display*.

BOVERMANN, T. 2009. Tangible auditory interfaces: Combining auditory displays and tangible interfaces. Ph.D. thesis. University of Bielefeld.

BOYD, L. H., BOYD, W. L., AND VANDERHEIDEN, G. C. 1990. The graphical user interface: Crisis, danger and opportunity. *J. Vis. Impair. Blind. 48*, 10, 496–502.

BREWSTER, S. 1994. Providing a structured method for integrating non-speech audio into human-computer interfaces. Ph.D. thesis. University of York.

BREWSTER, S., WRIGHT, P., AND EDWARDS, A. 1995. Parallel earcons: Reducing the length of audio messages. *Int. J. Hum.-Comput. St. 43*, 2, 153–175.

BREWSTER, S. 1998. Using non-speech sounds to provide naviation cues. *ACM Trans. Comput.-Hum. Interact. 5*, 2, 224–259.

BROCK, D., BALLAS, J. A., AND MCFARLANE, D. 2005. Encoding urgency in legacy audio alerting systems. In *Proceedings of the 11th International Conference on Auditory Display.* Limerick, Ireland.

BUSSEMAKERS, M. P. AND DE HAAN, A. 2000. When it sounds like a duck and it looks like a dog: Auditory icons vs. earcons in multimedia environments. In *Proceedings of the International Conference on Auditory Display*. 184–189.

CAPELLE, C., TRULLEMANS, C., ARNO, P., AND VERAART, C. 1998. A real-time experimental prototype for enhancement of vision rehabilitation using auditory system. *IEEE Trans. Biomed. Eng. 45*, 10, 1279–1293.

CARDIN, S., THALMANN, D., AND VEXO, F. 2007. A wearable system for mobility improvement of visually impaired people. *Vis. Comput. 23*, 2, 109–118.

CARELLO, C., ANDERSON, K. L., AND KUNKLER-PECK, A. J. 1998. Perception of object length by sound. *Psych. Sci. 9*, 3, 211–214.

CHENG, C. I. AND WAKEFIELD, G. H. 2001. Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. *J. Audio Eng. Soc. 49*, 4, 231–249.

CRISPIEN, K. AND PETRIE, H. 1993. Providing access to GUI's using multimedia system—based on spatial audio representation. *J. Audio Eng. Soc. 95th Convention Preprint.*

CSAPÓ, A. AND BARANYI, P. 2011. Perceptual interpolation and open-ended exploration of auditory icons and earcons. In *Proceedings of the International Conference on Auditory Display*.

CSAPÓ, A. AND BARANYI, P. 2012. The spiral discovery method: An interpretable tuning model for CogInfoCom channels. *JACIII 16*, 2, 358–367.

DAKOPOULOS, D. AND BOURBAKIS, N. G. 2010. Wearable obstacle avoidance electronic travel aids for blind: A survey. *IEEE T. Syst. Man Cy. C 40*, 1, 25–35.

DEBNATH, N., HAILANI, Z. A., JAMALUDIN, S., AND ALJUNID, S. A. K. 2001. An electronically guided walking stick for the blind. In *Proceedings of the 23rd Annual EMBS International Conference*. 1377–1379.

DINGLER, T., LINDSAY, J., AND WALKER, B. N. 2008. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the International Conference on Auditory Display*.

DOBRUCKI, A., PLASKOTA, P., PRUCHNICKI, P., PEC, M., BUJACZ, M., AND STRUMILLO, P. 2010. Measurement system for personalized head-related transfer functions and its verification by virtual source localization trials with visually impaired and sighted individuals. *J. Audio Eng. Soc. 58*, 9, 724–738.

DOZZA, M., CHIARI, L., AND HORAK, F. B. 2004. A portable audio-biofeedback system to improve postural control. In *Proceedings of the IEEE Annual International Conference of the Engineering in Medicine and Biology Society*. 4799–4802.

EBLING, M. 2009. Virtual senses. *IEEE Pervasive Comput. 8*, 4, 4–5.

EDWORTHY, J. AND HARDS, R. 1999. Learning auditory warnings: The effects of sound type, verbal labelling and imagery on the identification of alarm sounds. *Int. J. Ind. Ergonom. 24*, 603–618.

EDWORTHY, J., LOXLEY, S., AND DENNIS, I. 1991. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Hum. Factors 33*, 2, 205–231.

FAGERLONN, J. AND ALM, H. 2010. Auditory signs to support traffic awareness. *IET Intell. Transp. Syst. 4*, 4, 262–269.

FERNSTROM, M. AND BRAZIL, E. 2004. Human-computer interaction design based on interactive sonification—hearing actions or instruments/agents. In *Proceedings of the International Workshop on Interactive Sonification*.

FISH, R. M. 1976. An audio display for the blind. *IEEE Trans. Biomed. Eng. 23*, 2, 144–154.

FRAUENBERGER, C. AND STOCKMAN, T. 2009. Auditory display design—an investigation of a design pattern approach. *Int. J. Hum.-Comput. St. 67*, 11, 907–922.

FROHLICH, P. AND HAMMER, F. 2004. Expressive text-to-speech: A user-centred approach to sound design in voice-enabled mobile applications. In *Proceedings of the 2nd Symposium on Sound Design*. 4 pages.

Gaver, W. W. 1986. Auditory icons: Using sound in computer interfaces. *Hum.-Comput. Interact. 2*, 2, 167–177.

Gaver, W. W. 1988. Everyday listening and auditory icons. Ph.D. thesis. University of California.

Gaver, W. W. 1989. The SonicFinder, a prototype interface that uses auditory icons. *Hum.-Comput. Interact. 4*, 67–94.

Gaver, W. W. 1993. Synthesizing auditory icons. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Series*. 228–235.

Gaver, W. W. 1991. Sound support for collaboration. In *Proceedings of ECSCW'91*, Amsterdam, The Netherlands, 293–308.

Gaver, W. W. 1997. Auditory interfaces. *Hum.-Comput. Interact. 4*, 67–94.

Ghez, C., Rikakis, T., du Bois, R. L., and Cook, P. R. 2000. An auditory display system for aiding interjoint coordination. In *Proceedings of the International Conference on Auditory Display.*

Godbout, A. and Boyd, J. E. 2010. Corrective sonic feedback for speed skating: A case study. In *Proceedings of the International Conference on Auditory Display*. 23–30.

González, J., Yu, W., and Arieta, A. H. 2010. Multichannel audio biofeedback for dynamical coupling between prosthetic hands and their users. *Ind. Robot 37*, 2, 148–156.

Graham, R. 1999. Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application. *Ergonomics 42*, 9, 1233–1248.

GW Micro. Window-Eyes. http://www.gwmicro.com/Window-Eyes/.

Gygi, B. 2004. Studying environmental sounds the Watson way. *J. Acoustical Soc. Am. 115*, 5, 2574.

Gygi, B. and Shafiro, V. 2009. From signal to substance and back: Insights from environmental sound research to auditory display design. In *Proceedings of the International Conference on Auditory Display.* 240–251.

Gygi, B., Kidd, G. R., and Watson, C. S. 2004. Spectral-temporal factors in the identification of environmental sounds. *J. Acoustical Soc. Am. 115*, 3, 1252–1265.

Haas, E. C. and Edworthy, J. 1999. The perceived urgency and detection time of multitone auditory signals. In N. A. Stanton and J. Edworthy, Eds., *Human Factors in Auditory Warnings*. Gower Technical Press, 129–149.

Harness, S. J., Pugh, K., Sherkat, N., and Whitrow, R. J. 1993. Enabling the use of Windows environment by the blind and partially sighted. In *Proceedings of the IEEE Colloquium on Information Access for People with Disability.* 1–3.

Hearst, M. 1997. Dissonance on audio interfaces. *IEEE Expert 12*, 5.

Heller, L. M. and Wolf, L. 2002. When sound effects are better than the real thing. *J. Acoustical Soc. Am. 111*, 5, 2339.

Hellier, E. J., Edworthy, J., and Dennis, I. 1993. Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Hum. Factors 35*, 4, 693–706.

Hermann, T. 2002. Sonification for exploratory data analysis. Ph.D. thesis. Bielefeld University.

Hermann, T. and Hunt, A. 2005. An introduction to interactive aonification. *IEEE Multimedia 12*, 2, 20–24.

Hermann, T. and Ritter, H. 1999. Listen to your data: Model based sonification for data analysis. In G. E. Lasker, Ed., *Advances in Intelligent Computing and Multimedia Systems*. 189–194.

Hermann, T. and Zehe, S. 2011. Sonified aerobics—interactive sonification of coordinated body movements. In *Proceedings of the International Conference on Auditory Display*.

Hermann, T., Hunt, A., and Neuhoff, J. G. (Eds.). 2011. *The Sonification Handbook*. Logos Publishing House, Berlin.

Hersh, M. and Johnson, M. (Eds.). 2008. *Assistive Technology for Visually Impaired and Blind People*. Springer, 51–58, 289–299.

Hunt, A. and Hermann, T. 2004. The importance of interaction in sonification. In *Proceedings of the International Conference on Auditory Display*.

Jameson, B. and Manduchi, R. 2010. Watch your head: A wearable collision warning system for the blind. In *Proceedings of IEEE Sensors*. 1922–1927.

Jeon, M. and Walker, B. 2009. Spindex: Accelerated initial speech sounds improve navigation performance in auditory menus. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet. 53*, 17, 1081–1085.

Jeon, M. and Walker, B. 2011. Spindex (speech index) improves auditory menu acceptance and navigation performance. *ACM Trans. Access. Comput. 3*, 3, 10.

Jylha, A. and Erkut, C. 2011. Auditory feedback in an interactive rhythmic tutoring system. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*. 109–115.

KAHOL, K., TRIPATHI, P., PANCHANATHAN, S., AND GOLDBERG, M. 2004. Formalizing cognitive and motor strategy of haptic exploratory movements of individuals who are blind. In *Proceedings of the 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications.* 25–30.

KAY, L. 1973. The design and evaluation of a sensory aid to enhance spatial perception of the blind. *Electrical Eng. Report 21*, Department of Electrical Engineering, University of Canterbury, New Zealand.

KAY, L. 1984. Electronic aids for blind persons: An interdisciplinary subject. *IEEE Proc. A, Phys. Sci. Meas. Instrum. Manage. Educ., Rev. 131*, 7, 559–576.

KIM, J. K. AND ZATORRE, R. J. 2008. Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Res. 1242*, 263–275.

KRAMER, G. 1993. *Auditory Display: Sonification, Audification and Auditory Interfaces*. Perseus Publishing.

KUMMER, N., KADISH, D., DULIC, A., AND NAJJARAN, H. 2012. The empathy machine. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC'12).* 2265–2271.

KUNKLER-PECK, A. J. AND TURVEY, M. T. 2000. Hearing shape. *J. Exp. Psychol. Hum. Percept. Perform. 26*, 1, 279–294.

LAHAV, O., SCHLOERB, D. W., KUMAR, S., AND SRINIVASAN, M. A. 2008. BlindAid: A learning environment for enabling people who are blind to explore and navigate through unknown real spaces. In *Proceedings of the Conference on Virtual Rehabilitation*. Vancouver, Canada, 193–197.

LEGROUX, S., MANZOLLI, J., AND VERSCHURE, P. 2007. Interactive sonification of the spatial behavior of human and synthetic characters in a mixed-reality environment. In *Proceedings of the 10th Annual International Workshop on Presence*. 27–34.

LEPLATRE, G. AND BREWSTER, S. 1998. An investigation of using music to provide navigation cues. In *Proceedings of the International Conference on Auditory Display*. 10 pages.

LIARD, C. AND BEGHDADI, A. 2001. An audiodisplay tool for visually impaired people: The sound screen system. In *Proceedings of the 6th International Symposium on Signal Processing and Its Applications*. 198–201.

LOOMIS, J. M., MARSTON, J. R., GOLLEDGE, R. G., AND KLATZKY, R. L. 2005. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *J. Vis. Impair. Blind. 99*, 219–232.

LOPEZ, M. J. AND PAULETTO, S. 2009. The design of an audio film for the visually impaired. In *Proceedings of the International Conference on Auditory Display*. 210–216.

LÓPEZ, J. J., COBOS, M., AND PUEO, B. 2010. Elevation in wave-field synthesis using HRTF cues. *Acta Acust. United Ac. 96*, 2, 340–350.

MASSIMINO, M. 1992. Sensory substitution for force feedback in space teleoperation. Ph.D. thesis. MIT Department of Mechanical Engineering, Cambridge, MA.

MASSIMINO, M. 1995. Improved force perception through sensory substitution. *Control Eng. Pract. 3*, 2, 215–222.

MCGEE-LENNON, M. R. AND BREWSTER, S. 2011. Reminders that make sense: Designing multimodal notifications for the home. In *Proceedings of the 5th International Conference on Pervasive Computing Technologies for Healthcare*. 495–501.

MCGEE-LENNON, M. R., WOLTERS, M. K., MCLACHLAN, R., BREWSTER, S., AND HALL, C. 2011. Name that tune: Musicons as reminders in the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11).* 2803–2806.

MCGOOKIN, D. AND BREWSTER, S. 2004. Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM Trans. Appl. Percept. 1*, 2, 130–155.

MCKIEL, F. 1992. Audio-enabled graphical user interface for the blind or visually impaired. In *Proceedings of the Johns Hopkins National Search for Computing Applications to Assist Persons with Disabilities.* 185–187.

MCLACHLAN, R., MCGEE-LENNON, M., AND BREWSTER, S. 2012. The sound of musicons: Investigating the design of musically derived audio cues. In *Proceedings of the International Conference on Auditory Display*. 148–155.

MEIJER, P. 1992. An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng. 39*, 2, 112–121.

MOLLER, H. 1992. Fundamentals of binaural technology. *Appl. Acoust. 36*, 171–218.

MOLLER, H., SORENSEN, M. F., HAMMERSHOI, D., AND JENSEN, C. B. 1995. Head-related transfer functions of human subjects. *J. Audio Eng. Soc. 43*, 5, 300–321.

MORRISSETTE, D. L., GOODRICH, G. L., AND HENNESSEY, J. J. 1981. A follow-up study of the Mowat Sensor's applications, frequency of use, and maintenance reliability. *J. Vis. Impair. Blind. 75*, 211–247.

Murphy, E. F. 1971. The VA-Bionic laser cane for the blind. In National Research Council (Ed.), *Evaluation of Sensory Aids for the Visually Handicapped*. National Academy of Sciences.

Mustonen, M.-S. 2008. A review-based conceptual analysis of auditory signs and their design. In *Proceedings of the International Conference on Auditory Display*.

Mynatt, E. D. 1997. Transforming graphical interfaces into auditory interfaces for blind users. *Hum.-Comput. Interact. 12*, 1, 7–45.

Nees, M. A. and Walker, B. N. 2007. Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. In *Proceedings of the International Conference on Auditory Display*. 266–273.

Németh, G., Olaszy, G., and Csapó, T. G. 2011. Spemoticons: Text-to-speech based emotional auditory cues. *International Conference on Auditory Display (keynote lecture)*.

O'Brien, J. F., Shen, C., and Gatchalian, C. M. 2002. Synthesizing sounds from rigid-body simulations. In *Proceedings of the ACM Siggraph /Eurohaptics Symposium on Computer Animation*. 175–182.

Palladino, D. K. and Walker, B. N. 2007. Learning rates for auditory menus enhanced with spearcons versus earcons. In *Proceedings of the International Conference on Auditory Display*. 6 pages.

Parseihian, G. and Katz, B. F. G. 2012. Morphocons: A new sonification concept based on morphological earcons. *J. Audio Eng. Soc. 60*, 6, 409–418.

Patterson, R. D. 1982. Guidelines for auditory warning systems on civil aircraft. CAA paper 82017. London Civil Aviation Authority.

Patterson, R. D. 1989. Guidelines for the design of auditory warning sounds. *Proc. Inst. Acoust. 11*, 5, 17–24.

Patterson, R. D. and Mayfield, T. F. 1990. Auditory warning sounds in the work environment. *Phil. Trans. R. Soc. London 327*, 1241, 485–492.

Petrie, H. and Morley, S. 1998. The use of non-speech sounds in non-visual interfaces to the MS Windows GUI for blind computer users. In *Proceedings of the International Conference on Auditory Display*. 1–5.

Pressey, N. 1977. Mowat sensor. *Focus 3*, 35–39.

Rocchesso, D., Bresin, R., and Fernstrom, M. 2003. Sounding objects. *IEEE Multimedia 10*, 2, 42–52.

Schafer, R. M. 1977. *The Tuning of the World*. McClelland and Stewart Limited.

Schaeffer, P. 1966. Traite des objects musicaux. *Editions du Seuil*.

Schaffert, N., Gehret, R., and Mattes, K. 2012. Modeling the rowing stroke cycle acoustically. *J. Audio Eng. Soc. 60*, 7/8, 551–560.

Schaffert, N., Mattes, K., and Effenberg, A. O. 2010. A sound design for acoustic feedback in elite sports. In S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds., *Auditory Display* (CMMR/ICAD 2009 post proceedings ed.). Lecture Notes in Computer Science (LNCS) 5954, Springer Verlag, Berlin, 143–165.

Schaffert, N., Mattes, K., and Effenberg, A. O. 2011. An investigation of online acoustic information for elite rowers in on-water training conditions. *J. Hum. Sport Exerc*. *6*, 2, 392–405.

Sikora, C., Roberts, L., and Murray, L. T. 1995. Musical vs. real-world feedback signals. In *Proceedings of the Conference on Human Factors in Computing Systems*. 220–221.

Singh, D. 2010. Hybrid auditory based interaction framework for driver assistance system. *J. Comput. Sci. 6*, 12, 1499–1504.

Van den Doel, K., Kry, P. G., and Pai, D. K. 2001. FoleyAutomatic: Physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'01).* 537–544.

Ventura, L. C. L. and Fernandes, P. R. 2011. Remote guide for guiding the visually Impaired. In *Proceedings of the 2011 ISSNIP Biosignals and Biorobotics Conference (BRC'11).* 1–5.

Walker, B. N., Nance, A., and Lindsay, J. 2006. Spearcons: Speech-based earcons improve navigation performance in auditory menus. In *Proceedings of the International Conference on Auditory Display*. 63–68.

Walker, B. N. and Lindsay, J. 2006. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Hum. Factors 48*, 2, 265–278.

Wenzel, E. 1992. Localization in virtual acoustic displays. *Presence 1*, 1, 80–107.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman F. L. 1994. Localization using nonindividualized head-related transfer functions. *J. Acoustical Soc. Am. 94*, 1, 111–123.

Wersényi, G. 2003. Localization in a HRTF-based minimum audible angle listening test on a 2D sound screen for GUIB applications. Audio Engineering Society Convention Preprint Paper, No. 5902, New York.

Wersényi, G. 2007a. Localization in a HRTF-based minimum-audible-angle listening test for GUIB applications. *EJTA 1*, 1–16. Available at http://www.ejta.org.

WERSÉNYI, G. 2007b. Localization in a HRTF-based virtual audio synthesis using additional high-pass and low-pass filtering of sound sources. *J. Acoust. Sci. Technol. Jpn. 28*, 4, 244–250.

WERSÉNYI, G. 2008. Evaluation of user habits for creating auditory representations of different software applications for blind persons. In *Proceedings of the International Conference on Auditory Display*.

WERSÉNYI, G. 2009a. Evaluation of auditory representations for selected applications of a graphical user interface. In *Proceedings of the International Conference on Auditory Display*. 41–48.

WERSÉNYI, G. 2009b. Effect of emulated head-tracking for reducing localization errors in virtual audio simulation. *IEEE Audio, Speech, Language Process. 17*, 2, 247–252.

WERSÉNYI, G. 2010. Auditory representations of a graphical user interface for a better human-computer interaction. In S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds., *Auditory Display* (CMMR/ICAD 2009 post proceedings ed.). Lecture Notes in Computer Science (LNCS) 5954. Springer Verlag, Berlin, 80–102.

WERSÉNYI, G. 2012. Virtual localization by blind persons. *J. Audio Eng. Soc. 60*, 7/8, 568–579.

WILSON, J., WALKER, B. N., LINDSAY, J., CAMBIAS, C., AND DELLAERT, F. 2007. SWAN: System for Wearable Audio Navigation. In *Proceedings of the 11th International Symposium on Wearable Computers (ISWC'07)*.

W3: http://www.w3.org/.

Received February 2012; revised November 2012; accepted March 2013