# New Approaches to Intuitive Auditory User Interfaces

Dieter Spath, Matthias Peissner, Lorenz Hagenmeyer, and Brigitte Ringbauer

Fraunhofer Institute for Industrial Engineering (IAO),
Nobelstr. 12, 70569 Stuttgart, German
{dieter.spath,matthias.peissner,lorenz.hagenmeyer,
brigitte.ringbauer}@iao.fraunhofer.de

**Abstract.** This paper gives an overview of recent projects done at Fraunhofer IAO which take innovative approaches to the use of non-speech audio in human-computer interfaces. The examples illustrate the benefits that non-speech sound - alone and in combination with other modalities - can provide for supporting an effective interaction.

**Keywords:** Auditory user interface, sound design, acoustic feedback, navigation, vigilance.

## 1 Introduction

The use of non-speech audio in human-computer interfaces offers significant benefits. The most obvious advantage of sound as an interaction technique is that it can help to make computer interfaces more usable for vision impaired users. But sound is beneficial also for a broader user group, especially in situations and during tasks that do not allow a permanent visual scanning of a graphical user interface. This applies, for example, to many user tasks in industrial contexts, to vehicle driving, and to most situations of using a mobile device. Sound helps to reduce the load on the visual channel. It grabs the user's attention no matter where he is currently looking at. Therefore, it can be used effectively for alerts and reminders and in order to direct the user's visual attention to the graphical user interface for more detailed information. Another issue with mobile devices is their small visual displays. A reasonable complement of graphical and auditory elements for the presentation of information can help to avoid cluttered screens. Moreover, sound is capable of conveying information which can hardly be displayed on a screen, for instance, information about internal mechanisms of complex objects (Gaver, 1989). Finally, a combination of sound and vision can improve the usability, e.g. by "sonically enhanced widgets" (Brewster, 1998), and contribute to an enriched user experience.

Many of the above mentioned benefits also apply to speech output. But non-speech audio offers significant advantages over speech in some concerns and for some types of information, such as, for example, feedback on user actions, system states and transitions. Non-speech audio can transmit even complex information within very short messages. While speech is always presented in a sequential and therefore time-consuming mode, non-speech audio provides the possibility to encode multiple bits of information in a parallel manner, for example, by using the different attributes of

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2007, LNCS 4557, pp. 975–984, 2007.

sound, such as pitch, rhythm, loudness and timbre (cf. Blattner et al., 1989). As another advantage, non-speech sound is much less disruptive than speech. This is especially important for giving continuous status feedback about ongoing processes. While habituation is difficult to achieve with speech, continuous sound can fade into the background of consciousness and come to the foreground again only when it changes or stops (cf. Brewster, 2002).

In this paper, we give an overview of some selected projects in which we are investigating innovative approaches to the use of non-speech audio in human-computer interfaces. In the first contribution, we present a universal, multimodal human-machine interface (HMI) for hypovigilance management systems which feature different acoustic stimuli for warnings and vigilance maintenance. The second project deals with the design of intuitive sounds for generic user interface events such as affirmative and negative feedback, alerts and requests for clarification or confirmation. Following an empirical approach, we show how paraverbal utterances commonly used in human-to-human conversations, for instance “m-mh” and “mmh?”, can serve as an effective design metaphor. The third project describes the development of a concept for an orientation and navigation guide for blind people and visually impaired people using audio cues.

## 2 Development of a Universal, Multimodal HMI for Hypovigilance Management Systems

Today's "24 hour society" seems to pressure many individuals to sacrifice sleep in favor of other activities, without realizing the negative effects. In addition, as much as 10% of the general population suffers to some degree from sleep disorders, such as: hypersomnia, narcolepsy, and sleep apnea (Backhaus & Riemann, 1999) eventually leading to extreme sleepiness.

Sleepiness reduces alertness and with this, vigilance, and attention, so that the ability to perform is impaired. The speed at which information is processed and the quality of decision-making is also affected (Reissman, 1996). Sleepiness or hypovigilance, therefore, are among the key causes of serious accidents: According to NASA Aviation Safety Reporting System, approximately 21% of the aviation incidents are fatigue-related. In a similar way, most nuclear accidents (among them Chernobyl and Three-mile Island) have clear fatigue-related causes (Mitler et al, 1988). Moreover, more than 40% of automotive accidents on US-highways can be related to hypovigilance (Garder, 1998).

Obviously, the support of technical systems, so called Hypovigilance Management Systems or HVMSs, in states of hypovigilance is desirable. In principles, HVMS measure the state of vigilance of the respective user and take according measures if a critical state of vigilance with respect to the work task is reached: The user is warned and informed about his/her state of vigilance, and the system might try to keep the user awake for a restricted time (if at all possible) in order to enable the user to finish dangerous actions (e.g. to safely drive to the next highway exit). However, the systems known so far are mostly in a research state of development. They mainly base on single-sensor approaches and lack reliability. Moreover, most of these systems focus on the sensor system but do not feature an effective HMI. Finally, the systems present are restricted to single use contexts such as the car. As hypovigilance is a problem in many different use contexts, a universal system is needed.

Therefore, Fraunhofer IAO developed a new universal, multimodal HMI for HVMSs. It includes both the communication strategy (i.e. the warning strategy) and its implementation, the actual physical HMI-elements. In order to create a universally applicable interface, first, a representative set of application scenarios with respect to the dynamic and the mobility of the work system was chosen. The car driver, for example, is part of a work system of high system dynamic as he/she needs to quickly react to his/her surrounding but is immobile with respect to the work place, i.e., the car.

The requirements for the construction of the HMI were gathered by means of a thorough review of the relevant human factors literature as well as a questionnaire survey in the pilot sites chosen (Löher & Hagenmeyer, 2006). A cluster analysis was performed on these requirements with the aim to identify a universally applicable HMI-configuration. As a result, three functional modes were introduced:

Upon activation, the system starts in normal mode in which a status indicator informs the user about the hypovigilance state monitored and indicates that the system is working properly. Such continuous information should be presented in a way that it is not disturbing the user but should be recallable at any time; a traffic light like status indicator was chosen for this purpose.

The warning mode is activated when the user reaches a critical level of hypovigilance. Within this mode, depending on the current hypovigilance state of the user, either a cautionary or an imminent warning is displayed with the aim to stop the user conducting the work task and, by doing so, avoid accidents.

According to Wickens et al. (1998), an effective warning must first draw attention and then inform about the hazard, the potential consequences and give recommendations on how to react. Optionally, a feedback might be required from the addressee in order to confirm the reception of the information. Therefore, for cautionary warnings, the attention of the user is drawn by a complex sound according to DIN ISO 7731 (2002) in combination with a haptic vibration pulse and a bright flash. Then, more complex information about the hazard, its potential consequences and according countermeasures are presented by means of speech messages. These were preferred vs. text displays because of their higher compatibility with most work tasks in respect of the mobility of the user and interference with the work task itself. Finally, feedback from the user is demanded which is to be given via a push-button. In case no feedback is given, the warning is repeated. If still no feedback is received or if the user reaches a more critical hypovigilance level, an imminent alarm is given which follows the same structure. However, the stimuli to draw attention are intensified by an increased volume, higher pitch and vibration frequency. Moreover, the speech messages are adapted, emphasizing the urgency to quickly react to the warning.

After an imminent warning was given and confirmed, the system switches into the vigilance maintaining mode aiming to support the user safely stopping his/her work task, e.g., safely driving to the next exit. The user is informed about the activation of the vigilance maintaining system (VMS) as well as its restricted, short term usefulness by a speech message. The VMS then stimulates the user by means of so called "Landström Sounds": Landström et al. (1998) showed that specific disharmonious sequences of sounds, each lasting around 4 seconds, that are presented in irregular intervals of up to 5 minutes significantly enhance wakefulness. Moreover, a high level of acceptance by the participants could be observed.

The above warning strategy was implemented in SWITCHBOARD, a software package developed at Fraunhofer IAO with various I/O-capabilities. The status indicator, visual master alarm and feedback button were integrated in a wrist worn device (compare to fig. 1; for presentation purposes, all LEDs are activated). Haptic stimuli were presented by a belt mounted vibration device and acoustic stimuli were presented by a standard ear phone. All of the HMI elements were connected to the central computer system via a body area/local area network enabling the user to move freely.

**Fig. 1.** Wrist worn HMI-device

## 3 Intuitive Auditory Cues for Basic Interface Events

According to Buxton et al. (1994), non-speech audio mainly fulfils one of the following three basic functions: (1) alarming and warning the user, (2) providing status feedback and information about ongoing tasks, and (3) encoding numerical data. Whereas the latter often applies to highly specialized tasks, for example, monitoring continuous biomedical data, the other two functions play a major role in nearly all types of user interfaces. These two basic functions are focused also in the project here described.

The project aims at developing empirically founded design guidelines for intuitive sounds that can be adopted for generic interaction events in different application contexts. In order to achieve guidelines that can be applied to diverse situations and domains, the events to be sonified are addressed on a general and abstract level. In our project, we were focusing the following nine events:

1. Affirmative feedback when the user successfully triggers a function
2. Affirmative feedback when the system has successfully completed a given task
3. Negative feedback when the user tries to trigger a function currently not available
4. Negative feedback when the system has failed to complete a given task
5. Request for confirmation before executing an irreversible system action
6. Request for clarification in case of ambiguous or underspecified user inputs
7. Critical system state – medium priority
8. Critical system state – high priority
9. Alert/reminder

The above listed events resemble situations that often arise also in human-to-human conversations. In human dialogs, some of these or similar messages are simply conveyed by a short paraverbal utterance such as “m-mh” or “mmh?”. This led us to the idea that paraverbal human utterances might be an effective design metaphor for auditory cues in human-computer interfaces. This approach corresponds with Bill Gaver’s notion of *auditory icons* (Gaver, 1989), that rely on conveying information about computer events by analogy with every-day sound producing events.

### 3.1 Recording and Evaluating Paraverbal Utterances

As a first step, we recorded paraverbal human utterances which were systematically stimulated in an experimental setting[1]. This should help us to identify the needed pattern later used as a model for sound design. Eight students of speech training courses at the State University of Music and Performing Arts Stuttgart participated in our study. We had translated the above listed events to every-day situations that served as a setting for the recordings. In a scenario-based role play, the participants were instructed to react vocally – but without using speech - to certain events, and actions or questions by the experimenter. The participants' utterances were recorded. As the result of this first study we collected a total of 72 human voice recordings, eight for each of the nine everyday-life situations that were derived from the generic user interface events above.

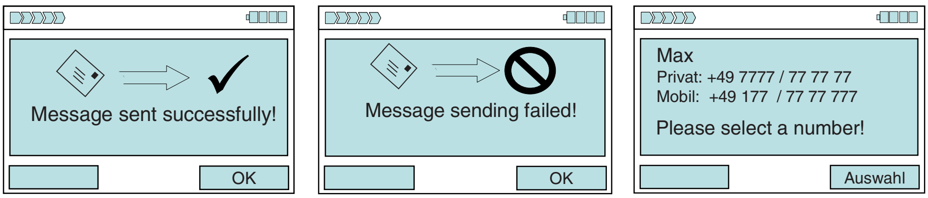


**Fig. 2.** Schematic screen views (examples) used for empirical evaluations

We conducted a second study in order to evaluate how effective the recorded sounds were in communicating a clear message to a naive listener. For this study, we translated the generic interface events to schematic screen views as possibly displayed on a mobile phone or a washing machine (see figure 2). 48 participants were instructed to assign each of the 72 recordings to the one of the concrete screen views which they felt suited best. The participants were randomly and in equal shares assigned to one of the two conditions of mobile phone screens and washing machine screens. For none of the presented sounds, the results from both test conditions differed significantly. This allows for a conjoint analysis of the total data set, and suggests that the findings might be generalized also to other similar applications. For 65 of the 72 recorded sounds, the participants' matching behavior differed significantly from an equal distribution over the nine events. It seems that most of the recorded sounds convey information that is related somehow to the presented interface situations.

For each interface event, we selected the recorded sounds, with which the event was most frequently associated. We found that in many cases, the listeners of a sound were not able to discriminate between two semantically similar events, such as the two affirmative feedback events (1 and 2 in the list above), the two negative feedback events (3 and 4), the two requests (5 and 6), and the two critical states (7 and 8). Sounds which were strongly associated with one event of such a pair often ranked high also for the other event of the same pair. For this reason, we decided to consider

[1] Sounds downloadable at http://www.hci.iao.fraunhofer.de/de/projekte/sounddesign

classification rates on basis of semantic pairs rather than single events. Doing so, we obtained highly accumulated classification patterns with peaks at 60 to 80%. One single exception is the reminder event. No recorded sound was clearly associated with the reminder screen. Figure 3 (a) illustrates the classification distributions over the paired events for some example sounds. In summary, it shows that the selected recordings can be effectively used for the discussed interface events. However, the chosen design approach seems to be applicable only to a restricted set of messages.

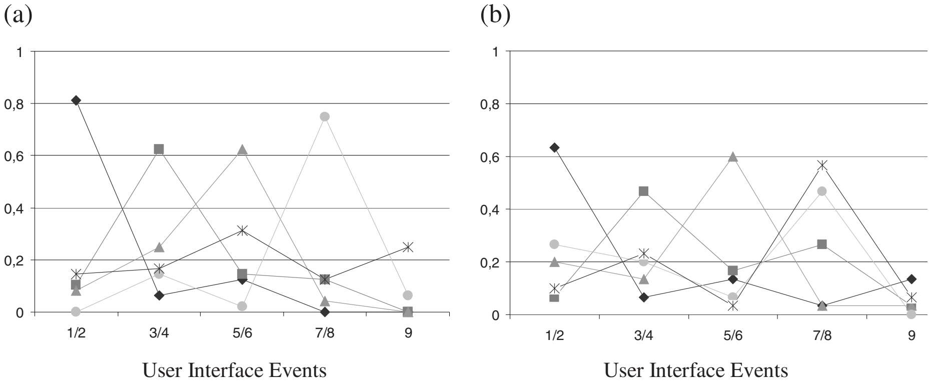


**Fig. 3.** Assigning sounds to user interface events: empirical distributions (relative frequencies) for selected paraverbal utterances (a) and respective synthesized sounds (b)

### 3.2 Generating and Evaluating Synthesized Sounds

Despite the evident effectiveness of the recorded human utterance, they will hardly be acceptable as interface sounds for a concrete product, as they do not meet essential requirements: The auditory user interface plays an important role for emphasizing certain product features, for example, valuable materials, and for contributing to an aesthetic and coherent over-all impression. Moreover, sound is more and more recognized as an important arena for branding.

Therefore, we need basic design guidelines which concentrate only on the sound features that are most substantial for conveying the intended message and leave undefined all the other sound features that can then be used for differentiation. Another related requirement of the envisioned guidelines is their succinctness that shall make the synthetic sound reproduction as economic and easy as possible.

In a first step, we generated synthesized sounds that imitate the selected recordings in a reduced and simplified manner[2]. (1) Rhythm was copied unchanged, i.e. the number and duration of the sound segments of the original sound were directly adopted. (2) We conserved the dynamics of the recorded sounds. The intensity of each sound segment was kept constant at the average intensity of the respective segment in the original recording. Variations within segments were eliminated. (3) Pitch variations in the recorded sounds that can be interpreted as involuntary were

[2] Sounds downloadable at http://www.hci.iao.fraunhofer.de/de/projekte/sounddesign

eliminated by setting the pitch of each segment to the average fundamental frequency ($f_0$) of the respective segment. We identified also obviously intended pitch variations within segments. Most of them were upward slides of about an octave with a doubling of frequency from the beginning to the end point of a segment. These segments were modelled by linear octave slides with the same average $f_0$ than the original segment. (4) Timbre was eliminated by using sinus waves.

The synthesized sounds were evaluated with 30 participants in the same fashion as reported for the recorded ones. The results are illustrated in figure 3 (b). Although some sounds obtained a significantly different assignment distribution for their synthesized versions than for the recorded ones, the results suggest that most of the relevant information was kept during the described process of synthesizing and simplifying the sounds.

To sum up, the described study supports the effectiveness of paraverbal utterances as a metaphor for the design of auditory interface cues. Further investigations will be needed in order to yield a deeper understanding of the sound features relevant for communicating the intended message. The presented procedure resulted in concrete design guidelines which can be written down in musical notation and easily adopted for reproduction. In future studies, we will investigate the effect of diverse timbres on the comprehensibility of the sounds.

## 4 Pedestrian Navigation for Blind and Visually Impaired People

### 4.1 Motivation and Related Work

The wayfinding requirements for blind people differ significantly from the ones using their visual senses to orientate. Blind pedestrians can not see obstacles on the pavement or spot the church in the distance for orientation. For micro-navigation, i.e. navigation through the immediate environment, the long cane, a guiding dog and additional electronic mobility aids help navigation. For macro-navigation, i.e. navigation through the more distant environment like the church example, mobility support is much more complicated (Petrie 1995).

There have been several projects that tried to meet this challenge. The MoBIC project (Strothotte 1996; Petrie 1997) used a preparation system to plan a walk in an urban area combining touch tablet maps with speech output for guiding exploration. The selected route can transferred to the MoBIC outdoor system (MoODS) that guides the user outdoor on his way. The MoODS takes the journey plan and assists travellers finding their way in by matching GIS information against travellers' positions (GPS). Directions, warnings, and other information is given in speech or Braille.

Loomis et al. 1994 aimed at developing a portable, self-contained system that allows visually impaired individuals to travel through familiar and unfamiliar environments without the assistance of guides. The system used GPS, a Geographic Information System comprising a database of the environment and functions for automatic route planning and for selecting the database information. The system provides a virtual acoustic environment for travellers, guiding them to buildings and other objects in the environment by sound signals which appear to come from the direction of the respective object.

There are also several navigation aid systems on the market (for an overview see http://www.tiresias.org/equipment/eb24.htm). An interesting recent product is Trekker (www.humanware.ca/web/en/p_DA_Trekker.asp). It is a navigation aid for blind and visually impaired people and uses GPS and digital maps to help blind persons find their way in urban and rural areas. It complements existing aids (white canes and guide dogs) and provides information through synthesized speech output.

### 4.2 Pedestrian Navigation for Blind in ASK-IT Project

ASK-IT[3] is a project that aims at supporting mobility impaired people in a mobile world. The prototype that is developed in this context will support blind people in wayfinding based on acoustic interface elements. It is researched in more detail how the provision of sound elements can be used to guide navigation.

Sound offers broad possibilities for unobtrusive information provision. This might help blind people to know that they are on the right way, without annoying spoken information that is currently not as important. Also privacy plays a major role here. The idea is to give blind people the power to choose what they attend to (Strothotte 1996, p. 3). The requirements analysis in the project revealed that it is especially important to give the control on the level of detail of guidance to the blind person. Some users require detailed guidance whereas others simply need reassurance that they are where they expect to be and are headed towards their destination.

We propose that acoustic icons and earcons for orientation-related and wayfinding information. A close connection to the prototypes on acoustic user interface elements and the speech user interface is seeked, e.g. to clarify ambiguous situations in wayfinding.

In December 2006 interviews with six severely visually impaired and blind people have been performed to learn about essential information to be presented by an electronic travel aid and nice-to-have information and events. The acceptance of the sound-based concept was very high although concerns on the challenges of sound design and concept development were raised as well. The sounds should e.g. pleasant and not annoying.

The concept can be divided in provision of orientation-related information and navigation-related information. The first class of information has the aim to orientate the person in the environment without a planned route – similar to the visual scanning of the physical surroundings of an able-bodied person. The second class of information deals with navigation-related information i.e. guide a person on a pre-planned route.

Those three categories of information need to be considered for **orientation**:

1. Information for orientation: Essential information like street names or directions,,detailed information like public buildings, shopping malls, and transport related information like stations, bus stops

---

[3] The project ASK-IT is funded by the EC in the 6th framework program, priority IST-2002-2.3.2.10 eInclusion. Contract no.: IST-2003-511298.

2. Safety-related information: Critical situations like street crossings and stairs and information on safe pathways
3. Temporary barriers like construction sites

For **navigation support** information like ”You are on the right / wrong way”, ”You have to turn right / left soon”, and “Now you have to turn right / left” needs to be given.

It turned out that the overall concept for navigation for blind people should not only contain sounds, but also some speech elements will be needed for convenient interaction. Examples include the commands: Where am I? Am I on the right way? What is the next orientation point? Additionally it will be researched, if some information (like “turn right now”) should be better given as a sound or in voice.

Currently the design of sounds is under development. Not only sounds have to be designed that are easy to learn and to distinguish, but also a concept for sound families needs to be developed to indicate the information categories and to allow the chunking of information. The user tests to evaluate the produced sounds are planned for April 2007. They will then be optimized and evaluated in the real world context in field tests two months later.

The developed sounds for orienting and guiding blind people will be also used to enhance graphical user interfaces for navigation. Even though orientation for visually able people is mainly supported by the visual sense, the developed sounds for navigation support will be useful: being on the move it is more comfortable to get a sound feedback if the selected path is still the right one than having to constantly monitor the PDA GUI to find out about wrong turns.

With this approach we do not only aim at developing a pedestrian navigation system especially suited to the needs of visually impaired people. We also aim to collect some important knowledge on sound families, knowledge on how complex and rich an audio based information system without visual feedback can be and get an idea about the potential of an audio-based display in real-world mobile situations.

## 5 Concluding Remarks

We expect an increasing need for effective auditory user interfaces in the near future. One the one hand, this will be due to the persistent growth in the use of mobile devices and their expanding functionalities. On the other hand, advances in the field of intelligent systems and environments will lead to a higher level of automation and to the need of keeping the users informed about current system actions and states. The vision of an “implicit human-computer interaction” (cf. Schmidt, 2000) will result in vanishing user interfaces. Human behavior and actions that are not explicitly directed towards a computer interface will be recognized and interpreted as a trigger of system reactions. Auditory cues and feedback will be needed in order to support user orientation and to enable a transparent and trustable human-computer interaction in smart environments.

## References


1. Backhaus, J., Riemann, D.: Schlafstörungen. Fortschritte in der Psychotherapie. Hogrefe, Göttingen (1999)
2. Blattner, M., Sumikawa, D., Greenberg, R.: Earcons and icons: Their structure and common design principles. Human Computer Interaction 4(1), 11–44 (1989)
3. Brewster, S.A.: The design of sonically-enhanced widgets. Interacting with Computers 11(2), 211–235 (1998)
4. Brewster, S.A.: Non-speech auditory output (Chap 12). In: Jacko, J.A., Sears, A. (eds.) Human-Computer Interaction Handbook, pp. 220–239. Lawrence Erlbaum Associates, Mahwah, NJ (2002)
5. Buxton, W., Gaver, W., Bly, S.: Auditory Interfaces: The Use of Non-Speech Audio at the Interface. Unfinished book (1994) (retrieved February 2007) manuscript at: http://www.billbuxton.com/Audio.TOC.html
6. DIN ISO 7731, Gefahrensignale für öffentliche Bereiche und Arbeitsstätten. Akustische Gefahrensignale (2002)
7. Garder, P.: Continuous Shoulder Rumble-Strips – A safety evaluation. In: CSRS Seminar, Linkoeping (August 1998)
8. Gaver, W.: SonicFinder: An interface that uses auditory icons. Human Computer Interaction 4(1), 67–94 (1989)
9. Landström, U., Englund, B., Nordström, B., Aström, A.: Laboratory Studies of a Sound System that Maintains Wakefulness. Perceptual and Motor Skills 86(1), 147–161 (1998)
10. Loomis, J.M., Golledge, R.G., Klatzky, R.L., Speigle, J.M., Tietz, J.: Personal guidance system for the visually impaired. In: Proceedings of the First Annual ACM Conference on Assistive Technologies. Assets '94, pp. 85–91. ACM Press, New York (1994)
11. Löher, L., Hagenmeyer, L.: Gestaltungsrichtlinien für die Nutzerschnittstelle von Hypovigilanz-Management-Systemen. In: GfA (Edt.): Innovationen für Arbeit und Organisation, pp. 39–42. GfA-Press, Dortmund (2006)
12. Petrie, H.: User requirements for a GPS-based travel aid for blind people. In: Gill, J.M., Petrie, H. (eds.) Proceedings of the conference on orientation and navigation systems for blind persons, Hatfield, UK, Royal National Institute for the Blind, London (1995)
13. Petrie, H., Johnson, V., Strothotte, T., Raab, A., Michel, R., Reichert, L., Schalt, A.: MoBIC: An Aid to Increase the Independent Mobility of Blind Travellers. The. British Journal of Visual Impairment 15(2), 63–66 (1997)
14. Reissman, C.J.: The Alert Driver. A Trucker's Guide to Sleep, Fatigue, and Rest in our 24-Hour Society. American Trucking Associations, Alexandria (1996)
15. Schmidt, A.: Implicit Human Computer Interaction Through Context. Personal Technologies, vol. 4(2&3), pp. 191–199. Springer, Heidelberg (2000)
16. Strothotte, T., Fritz, S., Michel, R., Raab, A., Petrie, H., Johnson, V., Reichert, L., Schalt, A.: Development of dialogue systems for a mobility aid for blind people: initial design and usability testing. In: Proceedings of the Second Annual ACM Conference on Assistive Technologies. Assets '96, pp. 139–144. ACM Press, New York (1996)
17. Wickens, C.D., Gordon, S.E., Liu, Y.: An introduction to human factors engineering. Addison Wesley Longman, New York (1998)