Auditory Displays in Human–Machine Interfaces

GUNNAR JOHANNSEN, FELLOW, IEEE

Invited Paper

Auditory displays are described for several application domains: transportation, industrial processes, health care, operation theaters, and service sectors. Several types of auditory displays are compared, such as warning, state, and intent displays. Also, the importance for blind people in a visualized world is considered with suitable approaches. The service robot domain has been chosen as an example for the future use of auditory displays within multimedia process supervision and control applications in industrial, transportation, and medical systems. The design of directional sounds and of additional sounds for robot states, as well as the design of more complicated robot sound tracks, are explained. Basic musical elements and robot movement sounds have been combined. Two exploratory experimental studies, one on the understandability of the directional sounds and the robot state sounds as well as another on the auditory perception of intended robot trajectories in a simulated supermarket scenario, are described. Subjective evaluations of sound characteristics such as urgency, expressiveness, and annoyance have been carried out by nonmusicians and musicians. These experimental results are briefly compared with time-frequency analyses.

Keywords*—Auditory display, earcon, human–machine interface (HMI), mobile robot, music, sound design, user-centered development, wavelet.*

I. Introduction

Human–machine interaction comprises all aspects of interaction and communication between human users and their machine via a human–machine interface. This human interaction with the machine, i.e., with an industrial plant or any other dynamic technical system, has nowadays been recognized as being essential for process safety, quality, and efficiency. The whole system of human users, the human–machine interface (HMI), and the machine is the so-called human–machine system (HMS) [1]–[3]. Different human user classes may be involved, namely, operators, engineers, maintenance personnel, and managers. They have different but overlapping information needs. The term "machine" relates to many diverse application domains. It indicates any kind of dynamic technical system (or real-time application), including its automation and decision support equipment and software. The automation components of the technical system are denoted as supervision and control systems. They interact directly with the pure technical (production) process. Examples of such processes are a power generation process, a chemical or a discrete-parts production process, an aircraft, a telemanipulator, a medical measurement system for human life functions, or a real-time software application. The decision support systems are more advanced, knowledge-based functionalities of the machine which provide advice for the human users, e.g., in fault diagnosis tasks.

All displays in HMIs for supervision and control of dynamic technical processes and systems have to consider the very stringent real-time aspect. This means that appropriate dynamic display elements have continually to vary in their attributes in order to indicate the changes of a large number of dynamic process variables as well as the changes of component and system states under normal and failure conditions. Thereby, provisions are made for enabling the human users to successfully accomplish their supervisory control tasks based on the displayed information. Human supervisory control is regarded as guidance of computerized and automated technological systems [1] as well as of human cooperative systems, e.g., in music performance [4].

Auditory and visual displays are applied with different emphasis in the HMIs for supervision and control of industrial, transportation, and medical systems as well as many service domains. Displayed information can be either direct (natural views and sounds in real environments) or mediated via computer graphics, video, audio, and music technologies. In the early days of supervision and control, operators walked around in the machine system, manually controlled the process, and had, in addition to visual instruments, the sense of sound, vibration, and smell to inform them about the process. Present-day processes are usually too large, complex, or dangerous to walk around in the plant, or they exist

Manuscript received June 13, 2003; revised November 12, 2003. This work was supported in part by the VW-Stiftung (Volkswagen Foundation) and in part by the University of Kassel, Kassel, Germany.

G. Johannsen is with the Systems Engineering and Human–Machine Systems Laboratory (IMAT-MMS), Institute for Measurement and Automation (IMAT), Department of Mechanical Engineering, University of Kassel, Kassel D-34109, Germany (e-mail: g.johannsen@uni-kassel.de).

Digital Object Identifier 10.1109/JPROC.2004.825905

only as computer model applications. The centralized control rooms of large industrial plants have separated people from the processes they should control. In these supervisory control situations, the operators and other human user classes have to deal with a considerable amount of information through layers of human–computer/machine interaction. Perception is restricted mainly to the visual sense by using visual presentation media. Only telephone or radio links provide voice communication with maintenance personnel in the plant. Sometimes, some auditory warning displays exist.

More recent developments show a certain trend toward combining several presentation media in the HMIs, which supports the reintegration of the most important human sensory modalities. The main human modalities are vision, hearing, touch, and smell. The combination of several of them allows multimodal perceptions. Advanced visualization and multimedia approaches including auditory displays have a high potential for considerably improving human–computer/machine interaction with all kinds of systems. The idea of multimedia communication with multimodal displays is to combine and to integrate different visual, auditive, and other media for the display of information about tasks to be performed with a machine or a computer. Particularly, the following media are combined with each other: computer-generated visual displays, video recordings, and real-time video, and auditive information such as recordings of sounds, noises, musical elements, and synthetic speech. Different media need not be synchronized but when they are they often are very powerful. Three-dimensional (3-D) stereoscopic video scenes and stereoscopic computer graphics, spatially distributed audio, teleconferencing with video, audio, and face-to-face communication, as well as haptic information and vibrations may also be used within a broad framework of multimedia and multimodal communication.

The main objective of the multimedia approach with multimodal displays for process control is to make the best use also of other human sensory modalities, in addition to the visual one. This will avoid the visual overload of human operators. Also, it can bring the operator closer back to the process where audio signals, such as noises from process components, and vibrations can be exploited for transmitting important information. Thus, many drawbacks and restrictions of overemphasizing the visual channel can be corrected by the additional appropriate use of auditory information. This addresses the hearing sense, which is regarded at least as the second most important sensory modality. Hearing is probably more basic than vision—which is why it usually takes precedence, e.g., via alarms. In any case, for alarm information, auditory displays are often superior to visual displays. Furthermore, several task situations of some application areas require auditory displays when the human user should not be distracted by visual displays from viewing his/her field of work. There is, of course, the great importance of auditory information for blind people for whom new possibilities of participating in human–machine interaction can be opened.

Many questions still exist with respect to the application of multimedia communication in industrial process control, medical, and transportation tasks. The general problem of the HMI design remains: which information is needed by the human user, when, in which form, and why? This problem becomes more severe with a larger number of technological options which increase further with the multimedia domain. Therefore, the information needs of the real human end users have to be investigated by means of task analyses in early design stages. This requirement becomes even more important in the cases of cooperation among several human users in cooperative work situations [5].

Before developing the next generation of multimedia HMIs for process supervision and control applications, more knowledge on suitable auditory displays has still to be collected. Even the investigation of auditory warning displays needs to be further intensified [6]. An important foundation of the field of auditory display has been provided with the Proceedings volume of the first International Conference on Auditory Display (ICAD) [7], in a series of conferences which are now held annually. The scope of these conferences covers a broad range of scientific approaches, with a more diverse application spectrum compared to the paper presented here. This paper has a particular focus on supervisory control and auditory communication in HMSs.

The following section of this paper explains all functionalities of HMIs and, thereby, the environment of auditory displays for supervisory control. In Section III, three types of classifications are outlined, namely, application-oriented, user-oriented, and sound-oriented classifications of auditory displays, hereby discussing matters beyond the HMSs domain. The user-centered development of auditory displays is suggested in Section IV.

The design of and experiments with auditory displays for mobile systems are described in more detail as an example application domain in Sections V and VI respectively. More ambitious auditory displays which are able to communicate system states and intentions by means of semantic nonspeech sound symbols and sound tracks have been investigated and developed in this research. The domain of autonomous mobile service robots in a human–machine environment was chosen as an example of the future use of auditory displays within all kinds of multimedia process supervision and control applications. The idea of the auditory displays is to combine relevant noise signals of the system with basic musical elements to form intelligible auditory symbols and sound tracks.

II. Functionalities of Human–Machine Interfaces

The HMI provides the information links between one or several human users and the machine (i.e., the technical system or application) [8]. This is shown in Fig. 1. Therein, the four main classes of human users are exemplified, namely, operators, engineers, maintenance personnel, and managers. The machine consists of its technical process, as well as its supervision and control systems for automation,

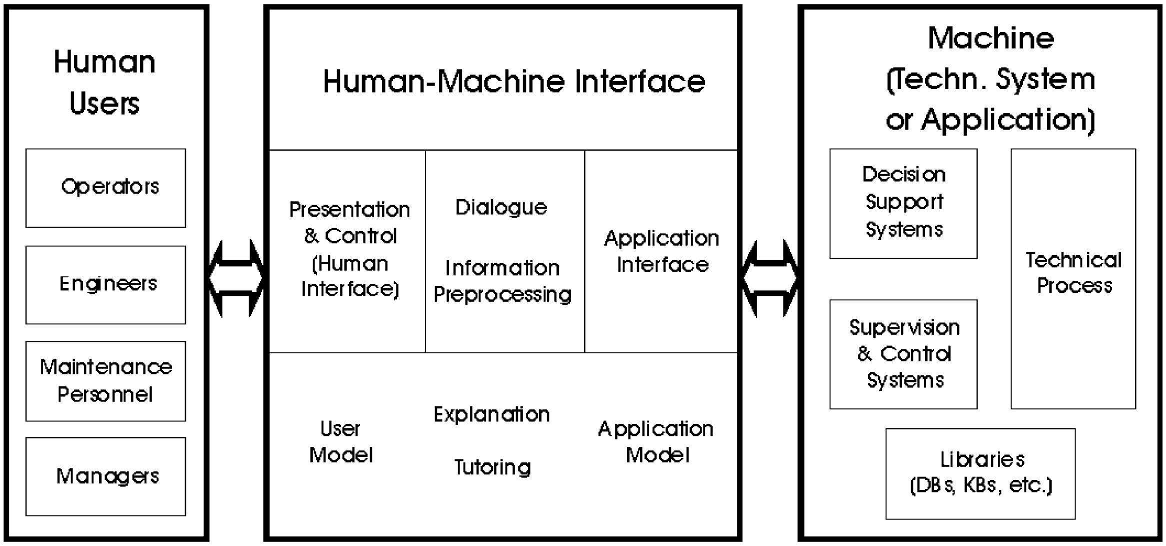

Fig. 1. Main functionalities of the HMI.

its decision support systems for knowledge-based advice, and its data and knowledge libraries.

As also outlined in Fig. 1, the HMI is subdivided into components for presentation and control (as the human interface), dialogue and information preprocessing, and the application interface as the more traditional functionalities, and into user model, explanation and tutoring, and application model components as the more advanced functionalities. The latter include also new approaches toward improved information preprocessing.

The more traditional levels of HMIs, such as presentation and dialogue, are generally separable from each other [8], [9]. The presentation and control level is concerned with the problems of how to present information to the human users, and how to transform their control inputs to the machine. The dialogue level deals with the information flows regarding such problems as what information to handle when, i.e., the handling of the suitable information at the appropriate time, between the human users and all other components of the whole HMS. Among other subfunctionalities, this includes the resolution of possible conflicts between different components with respect to dialogue requests to and from the human users. Information preprocessing is more and more often applied in order to provide improved information context and, hence, to facilitate the information processing activities of the human users.

The presentation and control level can also be viewed as the human interface in the narrower sense, on the human side. It is contrasted by the pure technological application interface on the machine side. Generally, visual, auditory (nonspeech sound and speech), gestural, mimic (including face-to-face), and haptic (sense of touch) information as well as vibrations can be used as forms of display for presentation. Combinations of these lead to multimedia and multimodal displays [10]. Currently, the main mode of presentation is still visualization, and the graphical user interface (GUI) is, and will be for some time, predominant. In former times, visual displays were implemented with electromechanical instruments. Today, they are replaced by computer graphics systems which contain a lot of graphical and textual dynamic pictures. These pictures are created with dynamic graphical editors [11]. Functional displays have been derived from the more traditional visual displays. They include new information preprocessing modules for a better consideration of user and goal orientations. Thereby, these displays realize more advanced functionalities [12].

Both the presentation and the dialogue levels can explicitly depend in their functionalities on the goals of the system as well as on knowledge-based technical systems or application models, user or operator models, and task models [8]. The more explicit representation of such models in the HMI leads to more advanced paradigms. An application model contains the knowledge about the goals, the structure, and the functions of a particular application. The functionality of this application or technical systems model internally supports all the other functionalities of the HMI with the respective context.

A user model functionality is needed if a certain adaptability to human user classes or single users is to be achieved. This is the major ingredient for adaptive interfaces. A more elaborate user model will always include a technical systems model in order to represent the user's view with respect to the technical system. In addition, knowledge of human information processing behavior and cognitive strategies has to be represented in a user model by means of algorithms, rules, and, possibly, active inference mechanisms.

Another knowledge-based functionality of a HMI is the explanation functionality. It is a kind of knowledge-based online help system which informs the human users on request about what the components of the technical system (and possibly also of the HMI) mean and how they function. A tutoring functionality supplies even more knowledge in an interactive way to novices and occasional users.

The presentation functionality is of main concern in this paper. It has gained new actuality with current research investigations and recent development solutions for advanced

Table 1
Application-Oriented Classification of Auditory Displays

Application Domains
vehicular guidance
- automotive guidance
natural sounds of vehicular motion
warnings and state information
sounds for product appeal
- aircraft guidance
natural sounds of system components
alarms and state information
speech communication
medicine
intensive-care units
operation theaters
monitoring of life supporting functions with medical appliances
cinema
sound tracks of films
comprise human voice, natural sounds, sound effects, and music
industrial plants
mobile phones

Functional Objectives
state and intent information
functioning of system components
intent communication
requests of human inputs
alarms and warnings
urgency
distinctiveness
arousal
appeal of products
desired emotional impression

Table 2
User-Oriented Classification of Auditory Displays

User Classes
different jobs, task allocations, and responsibilities
operators, engineers, managers, etc.
complementary sounds for team members

Communication with Individual Preferences and Needs
preferences or needs of sensory modalities depending on capabilities, impairments, tasks, and environments
sighted and blind users
transforming data and graphics into auditory interfaces

Individual User Capabilities
experts, novices, and occasional users
practice and quality of the hearing sense
non-musicians and musicians
human errors

visual, auditory, and multimedia displays. Particularly, auditory displays are investigated in this paper as means of presentation in HMIs. First, the following section provides a more general overview of auditory displays with three classifications based on application, user, and sound orientation.

III. Classifications of Auditory Displays

The three classifications of application orientation, user orientation, and sound orientation are summarized in Tables 1–3.

A. Application-Oriented Classification

Auditory displays have already been in use for some time as warnings in fault management situations [6]. These displays are especially designed in order to attract the human's attention. Thereafter, visual displays transmit the respective detailed information. Examples can be found in aircraft and in cars, in industrial plants, in health monitoring devices, and with mobile phones. Different nonspeech sounds of audio effects and of musical elements have been implemented.

The fault management activities of human operators in complex HMSs are very important in order to guarantee safe system operations under all possible conditions and, at the same time, to avoid complete shutdowns of the whole system as often as justifiable. Good designs of fault management support functionalities contribute to improved human–machine interaction in cases of technical system failures. "Alarm showers" may develop as a sequel of fault consequences in interdependent subsystems. They are a definite problem related to supervision and control system functioning and to operator actions in response to faults. Improved fault management support facilities have to aim at a reduced number of alarms, which need to become more intelligent, based on task-oriented transparency and understandability with respect to the context of a particular application domain. This requirement is true also for auditory displays of alarm information and, beyond, of any other sound information.

Sound is used, for example, in the application domains of vehicular guidance, medicine, and film with different intentions and objectives in mind (see Table 1), and also with different degrees of sophistication. The above requirement for improved sound information is still far from being fulfilled in many applications.

In vehicular guidance, sound is used in different forms, as well as in aircraft guidance and in automotive guidance.

Table 3
Sound-Oriented Classification of Auditory Displays

Speech		
artificial speech outputs recordings, speech synthesis		
Non-Speech Sound		
natural sounds (a) characteristic for components in real environment (b) re-used as sound symbols (metaphors)		
artificially created sounds (a) music generated with traditional music instruments, electronic means, and computers (b) sound effects transmit entertaining and emotional expressions created with computers and synthesizers (c) sound symbols (earcons) provide specific meanings and contents created with computers and synthesizers		
Parameters of Sound and Music		
rhythm	melody	attack
pitch	time duration	decay
timbre	harmony	modulation
dynamics	counterpoint	consonance
volume		dissonance
reverberation		
several effect parameters	spatial orientation	

First, we cannot ignore the inherent sounds caused by the mechanics of vehicles such as cars and aircraft. Thus, an important sound source arises from the natural sounds of vehicular motion and, particularly, of the drives. Such natural sounds are, for example, motor noise, engine noise, compressor noise, as well as noises which originate with lowering the landing gear, the motion of the landing flaps, the switching of the gears, etc. The usefulness of such natural sounds is not always clear for the human. Even a 1988 aircraft crash in the United States could be attributed to the wrong interpretation of sounds by the pilots. They had fatally mistaken a compressor noise for an engine noise [13].

Another class of sounds in aircraft guidance, but increasingly also in automotive guidance, concerns auditory alarm displays. These have the advantage, compared with visual alarm displays, that they arouse the attention of the human without his or her need to be looking in a particular direction. Unfortunately, considerable problems appear here too. The meaning of the alarm sounds is not always clear; and they can easily mask each other. A 1991 aircraft crash in Sweden made this especially evident. The ongoing investigations have shown that auditory alarm displays in all civil aircraft of the world need considerably to be improved [14]. This is particularly important for the occurrence of multiple faults. The intelligibility of natural and of alarm sounds can further be impaired through speech communication in the cockpit and with air traffic control. Similarly, impacts of interferences arise from traffic radio and infotainment systems in automotive vehicles.

The number of alarm sounds is sometimes even higher in medicine. They are used in intensive-care units and in operation theaters. A multitude of life supporting functions have to be monitored with medical appliances. These can only be displayed auditorily because the field of vision of physicians and nurses has to be focused on the patient. The problems seem to be even greater in this application domain. There are too many and confusing alarm sounds, the urgency of which is not clearly identifiable. Even the death of a patient due to the mixing up of alarm messages is reported [15].

In cinema, sound design is an artistic as well as a technical activity [16]. It determines the overall quality of a film to a large extent. An expressive sound track may be as equally multilayered and intricate as the sequence of pictures. The complete sound track comprises four essential parts: the human voice, natural sounds, sound effects, and music. They need to be mixed and balanced to give the desired impressions in the film. Sound effects can be applied synchronously or asynchronously with the picture events. It should be possible to learn from the immense professionality of the sound design in cinema, and from other areas of the entertainment industries, for human–machine communication in many application areas. This is true, for example, with respect to the expressiveness and the emotionality of sounds, but also concerning the courage of designers to use more complex, and yet short, appealing sounds.

Sound symbols (i.e., earcons, which will briefly be explained in the following section on sound-oriented classification) and sound effects can be designed in such a way that they are suitable for the audification of system information and system appeal. This is particularly true for system information which describe a system's states in failure situations. In many application areas, the visualization of system states is thereby complemented by the usage of earcons for the characterization of alarm states. In addition, other system states can be marked through easily remembered and expressive sounds. This refers to positions, speeds, and directions of spatial movements in mechanical systems or in fluids as well as to the labeling of unusual, abnormal system states or to the transmission of plans and intentions. As an example, the auditory display of aircraft deviations during approach and landing has recently been investigated [17].

The appeal of systems can be improved by using well-designed sounds, e.g., of closing doors in cars. The appeal should produce a desired emotional impression from perceiving the system's outward appearance and should, thereby, achieve a specific impact on the human, e.g., the sensation of high product quality and reliability. Hence, appeal of products receives a high place value in advertising and in marketing. Good products should also represent themselves as appearing to be good. Therefore, sound effects and sound symbols used for appeal must not sound cheap but have to be convincing.

B. User-Oriented Classification

Several dimensions of user orientation can generally be distinguished within user interfaces (see Table 2). One dimension concerns the different human user classes, which are determined by different jobs, task allocations, and responsibilities. Examples are operators, engineers, and managers. With visual displays, it is assumed that the cooperation between these different human user classes can be facilitated by the combination of several graphical representations on different abstraction levels [5]. The different information needs of these user classes are considered in the multihuman interface design by providing them with dedicated windows or screens. Engineers and managers may more often want to use goal-orientated hierarchies and functional presentations, whereas operators may use the less abstract presentations of ecological and topological displays [12]. However, free navigational access to all display options must be allowed for all user classes, based on the different focuses of their preferences. Concepts such as visual momentum [18] and cognitive layouts [19]—both representations which minimize cognitive load—should be implemented in such a way that they support the integrated view among team members of multiuser groups across all different forms of graphical representations.

Less experience exists with the design of auditory displays for different human user classes. However, it can be assumed that different kinds of auditory displays are equally feasible on different levels of abstraction for different user classes. This can be explained with the example of a power plant. The well-trained operator of such a plant is able to distinguish between the natural sounds of a healthy and a faulty turbine. Any auditory display which has been derived from such natural sounds has a contextual meaning for the operator. On the other hand, the manager may not understand these sounds equally well but needs more abstract indications of load and risk levels. Of course, parallel indications of different sounds for the same task or problem situation for different user classes in the same workspace have either to complement each other or need to be avoided. The composition of such complementary sounds may be even more difficult in application domains such as operational theaters and intensive care units.

The second dimension of user orientation is related to the human communication aspects which depend directly on the type of interfaces for the human–machine and human–computer interactions. Displays can more or less intuitively be well understood depending on the quality of the interface design. Further, different sensory modalities may be preferred or even needed by different human users depending on their individual capabilities and actual impairments. Today's overvisualization of a huge number of workplaces almost everywhere may or may not be a problem for sighted users. Blind users, however, feel pushed out of bounds by this development more and more dramatically. Auditory displays can be designed as an acceptable means of access to the computerized world for visually impaired and blind users. The challenge to design intelligible auditory displays for these users is so strong that it leads also to improved auditory display solutions for sighted users. One successful example has been shown with transforming graphical displays which are typically used in human–computer interaction into auditory displays for blind users [20]. Pop-up windows, for example, are replaced by whistling sounds where their appearance is indicated with a rising pitch and their disappearance with a descending pitch. Also, graphical diagrams can be presented to blind people by using music alone [21].

The third dimension of user orientation is determined by individual user capabilities. There is always the distinction between the novice, the expert, and the occasional users. These capabilities refer to the level of understanding of the application contexts as well as to the different levels of capacity in dealing with the particular requirements of the interface itself. Training, experience, and motivation may change these levels. In the cases of auditory displays, the practice and the quality of the hearing sense may be influential to the efficiency of the display usage. Even musicality may be important in some advanced auditory displays. Thus, it is of interest to compare the performance of musicians and nonmusicians with the investigation of such displays. Although musicians usually perform better than nonmusicians, the effects may not be huge because most people have an innate capability for understanding music.

Another aspect of the user-oriented classification concerns human errors possibly triggered or avoided by means of auditory displays. Human errors can be classified on different cognitive levels in HMSs, namely, on the three levels of skill-based, rule-based, and knowledge-based behaviors [22]. Errors at the skill-based level consist of slips and lapses, whereas errors or mistakes at the rule-based level may be coarsely divided into the misapplication of good rules and the application of bad rules. Finally, errors or mistakes at the knowledge-based level originate from bounded rationality or an incomplete or inaccurate mental model of the problem space. Errors at the skill level can be minimized by a good design of the HMI; a consistent and structured presentation of the information minimizes the possibility of lapses, and protection against many erroneous actions at the skill level can be built into the interface. Possibilities for providing intelligent support at the rule and the knowledge levels have been suggested with different forms of decision and knowledge support. Thus, the design of auditory displays will primarily strive for minimizing human errors at the skill level. Auditory displays at the rule and the knowledge levels may be only feasible in strong goal-, task-, and context-orientations within a design approach with well-organized user participation. Then, a mapping between sound and thought needs to be brought about.

C. Sound-Oriented Classification

Several sound classifications are possible. Suggestions of sound classification in electronic media are made in [23]. A comprehensive classification for auditory displays is provided in Table 3. It shows that auditory information can be divided into speech and nonspeech sound. Artificial speech outputs have been introduced as auditory displays. The basic aspects of human–computer interaction via voice have been summarized in [24]. Single or a few words and short sentences can be generated from recordings or with speech synthesis. They are mainly applied for indicating special systems states, particularly in failure and emergency situations, and for requesting human control inputs.

Nonspeech sounds may be either natural or artificially created. Both can further be subdivided. Natural sounds are characteristic of many technical system components in real operational environments. An example is the motor noise in a car. However, care has to be taken not to misinterpret natural sounds. As outlined above, accidents have even been caused by mixing up two natural sounds from two different subsystems because of their strong resemblance. On the other hand, natural sounds can also be recorded and, then, can be reused as sound symbols in another context. A slamming door, for example, may illustrate that a work activity or a document has been terminated. In a way, the sound is used as a metaphor. Everyday sounds mapped to computer events have also been called auditory icons [25], [26].

Artificially created sounds can be subdivided into music, sound effects, and sound symbols. Music is mainly generated with traditional music instruments, but increasingly also with electronic means and with computers. In contrast, sound effects and sound symbols are created today solely with computers and synthesizers. The sound effects transmit entertaining and emotional expressions. Sound symbols, on the other hand, are provided with specific meanings and contents. Therefore, they are particularly well suited for human–machine communication. Sound effects and sound symbols may either be produced from artificially created sound patches or by alienation-effect processing of natural sounds, or they may be composed of musical elements.

Sound symbols are also called *earcons* in analogy to icons, which are picture symbols [27]. Guidelines for the creation of earcons have been proposed [28]. Earcons are short expressive audio patterns which are composed of one or a few motifs. A motif is created as a short sequence of tones. The properties of tones, motifs, and earcons can be manipulated by changing musical parameters in a desired way. Designable parameters are rhythm, pitch, timbre (tone or sound color), dynamics and volume, reverberation, and some effect parameters. Further, melody and time duration are important parameters for motifs and earcons. In addition, the parameter harmony as well as, possibly, counterpoint and other compositional design principles may be applicable for multiple earcons. Generally, presenting more than one earcon concurrently may lead to recognition problems [29]. Some more basic sound and musical properties are attack, decay, modulation, consonance and dissonance, etc. Earcons can be used for the information transmission of alarm messages as well as of system information in different application domains. Further, music can be used for debugging software programs and for interacting with computers [30], [31].

Another important characteristic of sound is its spatial orientation. Sounds can have an effect on the human from different directions, including several at the same time. The 3-D presentation of artificially generated sounds is called spatialization. As an example, it can be used in aircraft guidance to artificially improve natural sounds and, thereby, to make their spatial direction as well as their other sound attributes better perceivable [17]. An enhanced spatial sound presentation system has been integrated with an audio emphasis system and a gestural input recognition system for a new user interface [32]. Perspectives for spatial auditory displays have been suggested as a helical keyboard [33]. A cheap way of getting good spatial audio is to use a manekin [34].

All types of auditory information can be characterized by a time-dependent frequency spectrum which may be visualized as a spectrogram. Such diagrams present frequency over time whereby the intensities are indicated as degrees of blackening or color values in such time-frequency planes. This technique is briefly discussed in Section VI-C.

IV. User-Centered Development of Auditory Displays

Participative design is mandatory for achieving better user interface products [35]. This is a strong human factors concern which still needs to be implemented more often and more systematically in industrial design projects. In this respect, HMIs for industrial, transportation, and medical processes are a particularly important and complex subset of all kinds of user interfaces, because the real-time demands of these applications can certainly also be better met with user participation.

The two important characteristics to be achieved in well-designed human–machine interaction are human centeredness and task orientation. Human centeredness is a principle which emphasizes the psychophysiological and cognitive foundations of human behavior in the interaction with any type of machine/application or, more generally, artefact. A cognitive artefact is "an artificial device designed to maintain, display, or operate upon information in order to serve a representational function" with respect to a task [36]. Human centeredness also emphasizes the individual differences between users and hence the importance of user modeling. It also stresses that the views of designers and users may be different and the designer must take this into account.

The second principle, being complementary to the first one of human centeredness, is task orientation. A major handbook of cognitive task design has recently been published [37]. The final objective of any human–machine interaction is the accomplishment of a number of tasks. Thus, looking from a systems engineering perspective,

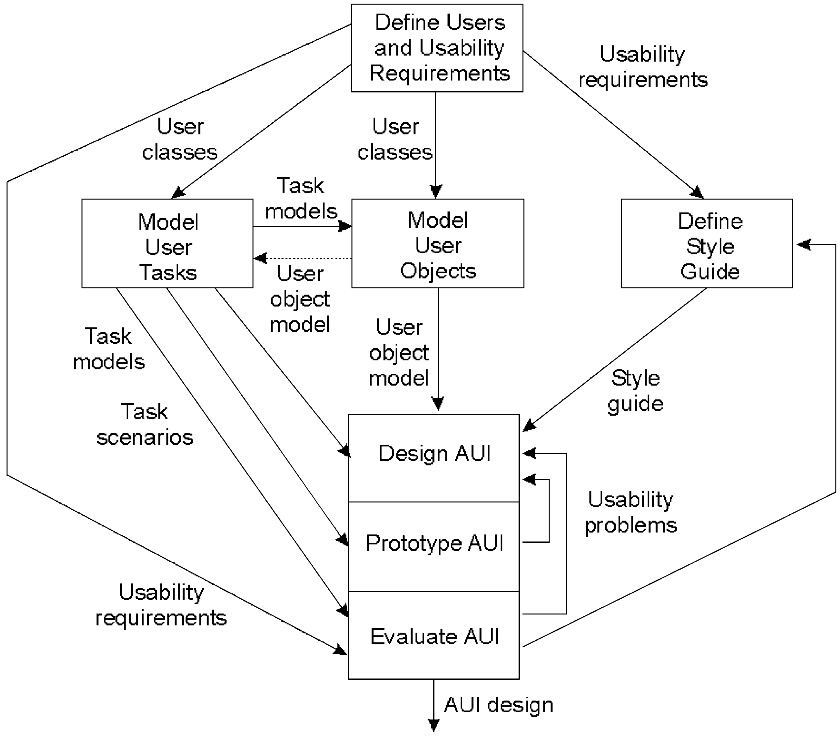

Fig. 2. Cognitive systems life-cycle development of auditory displays (after [39] and [40]).

task orientation is more like a top-down principle for overall purposeful achievements. On the other hand, human centeredness involves more characteristics of a bottom-up principle with respect to available resources and constraints of human users. Human centeredness can also be viewed as the core of user orientation. Both task orientation and user orientation (or human centeredness) can be achieved with the participation of end users in early stages of the design of their respective HMIs. Other forms of user participation during later stages of the design process consider the evaluations of intermediate prototypes and of the final interface product itself.

Sound design and the development of auditory displays, for the usage in technical systems and appliances, have equally well to become more systematic and, particularly, more task and user oriented in the future. Some of the above-mentioned problems in aircraft guidance and in intensive care units, e.g., unclear levels of urgency and low distinctiveness of sounds, can be explained by the fact that user participation during the development process has rarely happened and that the sound design was apparently conducted without sufficient task orientation. Moreover, many sounds used in practical applications are much too simple, e.g., only single tones, without convincing expressiveness, and without satisfying correlation with the intended meaning. If at all, experimental investigations were often carried out with end users only with the already existing sounds which were, then, not changeable anymore.

Cognitive systems engineering methodologies support designers of HMIs in their systematic design and evaluation activities [38]. The objectives are to enhance the efficiency of these activities and to improve the final interface products. The GUI design and evaluation (GUIDE) method is one of these methodologies [39]. It combines systems engineering life-cycle procedures with end-user participation and rapid prototyping for the development of the GUI. Strictly speaking, it is more suitably applicable to human–computer interaction. Nevertheless, many of the GUIDE techniques are also relevant for HMIs in process control of dynamic systems. Further, the design and evaluation method GUIDE strongly emphasizes the presentational issues, however, only with respect to GUIs.

The cognitive analysis and participative design stages of the GUIDE methodology have specific objectives and deliver well-defined outputs [39]. In this paper, they have been redefined with respect to auditory displays as an additional communication means in process control. Thus, the GUIDE methodology has been transformed into the new auditory user interface design and evaluation (AUIDE) method [40].

The AUIDE methodology starts with defining user classes and usability requirements for a specific application domain in the first life-cycle phase; see Fig. 2. Then, task models and task scenarios are produced. The user object model is specified by the designer with the intention to support the development of an interface in which such a user object model may create effective mental models inside the users' mind. Presentational issues are considered by the style guide, which needs to contain particular guidelines for auditory presentation.

End users and other operational personnel already have to be involved intensively in the development process for this important layer of modeling user tasks and user objects as well as the definition of style guides. Complete task situations have mentally to be played through. In particular task scenarios, for example, the relative urgency of different alarm messages in the intensive care unit needs to be clarified. Typical failure and alarm situations may be regarded as user objects in addition to equipment and software components. These objects and their relationships among each other represent the mental model of the later end users.

The style guide lays down the stylistic basic elements and their possible compositional rules, as with GUIs and visual displays. Examples are the establishment of the musical parameters and the sound effects to be used, such as the number or the maximal number of the tones in earcons, the sound color palette, the rhythm styles, and the appeal character.

The stage of designing the auditory user interface (AUI) is based on all the specifications achieved during the life-cycle phases of task modeling, user object modeling, and style guide definition; see Fig. 2. It refers to the initial design of the interface and should provide a solid basis for prototyping. Only during this and the following stages can sounds be composed, generated, and evaluated. These last three life-cycle phases are passed through several times.

Further expert participation of different human user classes needs to be organized during these later design stages, in order to evaluate intermediate prototype designs of the sounds and the auditory displays in the HMI. The aspects of cooperative work between the different user classes have a high priority in these evaluations. The same is true for the final evaluation at the end of the HMI (AUI) design. Intermediate prototype designs and final AUI systems implementation alternate with their corresponding evaluation stages. Thus, several iterations lead to progressively improved versions of prototypes via respective intervening user evaluations and usability testing; see Fig. 2.

The user orientation, as shown in Fig. 2, is a strong requirement for successful sound design and AUI design. The consequences of not obeying this request may be quite severe, as a recent investigation of earcons for mobile phones has demonstrated [41]. Very interesting results were obtained, but the failure of the sound design has also honestly been admitted. Although the investigation was performed by the applications and user-centric technologies laboratory of the manufacturer, the sound design process of the earcons had been accomplished completely without user participation. Obviously, even extensive experimental investigations and usability tests only with the final product of the auditory displays are not enough. This may lead to the possible insight, although too late and similarly as with visual displays, that the development had been bypassing the end users and their requirements. If such products are, nevertheless, introduced into the market, risks with bad numbers of sales and even accidents are likely.

Other examples with cars and medical appliances have shown that most sounds and earcons of these applications are too simple, not unequivocally relatable to intended meanings, insufficient in the mapping of relative urgencies and, in the medical domain, in their numbers much too high. Also here, only large-scale task analyses and the development of task and user–object models under early user participation can lead to success. This user participation should also be directed toward the definition of the style guides and has further iteratively to be continued during the true sound design process. The successes of software industries demonstrate with the development of GUIs that this high effort for usability engineering embedded in the systems life-cycle approach is profitable.

What user participation cannot yield, however, are design ideas themselves. Here, the artistic creativity with sound design is likewise required as with graphics or product design. However, sound designers can only be successful if they understand the systems engineering associations of the application domain equally well in all task situations and in all user views. For instance, it is conceivable for medical appliances to process natural sounds of life functions with sound effects and to transform them into earcons. For cars, the necessity arises to develop also sounds for appeal, in addition to improve sounds for systems information, e.g., about the state of the lighting. Perhaps solid sound effects are imaginable for the slamming of a car door, including hints at nondesired intermediate states such as not completely clicked into place or slammed too roughly [42].

In the following two sections of this paper, the directional orientation and the intent communication, particularly with a mobile robot in a supermarket scenario, characterize an example application domain in more detail. Two user classes are considered, namely, nonmusicians and musicians. The usability requirements include understandability and recallability of auditory sounds and sound tracks. Two successive task scenarios deal with the directional orientation in an abstract open space with eight possible directions and, then, with the communication of robot trajectories and states in a fixed supermarket floor plan. Important user objects are directions and intentions of movements as well as additional states such as *heavy load* and *near obstacle*.

The development of style guides for auditory displays will require major efforts in the future, mainly on the levels of research, sound design, classification, and standardization. For the work in this paper, a heuristic approach based upon knowledge from music theory and sound engineering has been chosen. The systems life-cycle phases of design and prototyping, as shown in Fig. 2, are explained in Section V. The experimental investigation with user evaluations and usability testing is reported in Section VI.

V. Design of Auditory Displays for Mobile Systems

A. Design Objectives and Approach

Auditory displays have been developed by the author for autonomous mobile service robots as an example application domain for all kinds of mobile systems and other dynamic sociotechnical systems [43]–[46]. The expressiveness and physical interaction in human–robot communication has been investigated in [47]. In the design reported here, the rich body of knowledge from music theory as well as from auditory science and sound engineering is exploited. The idea is to combine relevant original noise signals of robot movements with basic musical elements as intelligible auditory symbols (earcons) and, then, to create sound tracks from them. Robots communicate their actual positions, movements, and intentions as well as failures and related warnings by means of nonspeech audio symbolic expressions to a human listener.

Several auditory symbols were invented, as directional sounds and as robot state sounds, by the author of this paper. They were composed as pure musical sounds with different melody and rhythm patterns. Results in the literature "show that high levels of recognition can be achieved by careful use of pitch, rhythm and timbre" [48]. All the sounds were designed and created by the author with a powerful PC and Windows, Logic Audio software, Cool Edit audio editor, a MIDI synthesizer, and a keyboard. The pure musical sounds from the synthesizer were recorded with Yamaha DSP Factory with its audio expansion unit under the Logic Audio software on the PC. Thus, WAV files were produced which needed some minor audio editing.

The equivalent directional sounds of robot noises were derived from DAT recordings of the movement noises of a real service robot Nomad 200 (called "Tom") in the laboratory [43]. The variations of the directional sounds based on robot noises were generated through several steps of audio editing from the recordings of the movements of this service robot. The same melody and rhythm patterns were composed by time and frequency editing as in the pure musical sound cases. The Logic Audio editor was used for frequency transpositions of the pitch levels as requested in the different tones of the directional sounds. The Cool Edit audio editor was more appropriate for cutting and assembling the necessary time slices of each sound element to the desired directional sounds. Some amplifications with fading-in and fading-out effects were performed for achieving a clear separation between the different tones of each sound.

Several experiments were designed and performed by the author within an extensive exploratory study. In the first experiments, human subjects had to learn, understand, and recall auditory symbols with related meanings about spatial orientation and directional movements. In succeeding experiments, intelligible auditory symbols and sound tracks were presented in a supermarket scenario with the simulated environment of mobile service robots. Intended trajectories with directional sounds and robot state sounds for moving-straight, moving-curved, *heavy load*, *waiting*, *near obstacle*, and *low battery* were communicated by the robot to the human. A group of nonmusicians and a control group of musicians participated in all the experiments.

B. Auditory Symbols for Directional Information

For the first experiments, a new set of auditory symbols was designed as directional sounds for eight possible directions of motion of the robot in space [44]. In this investigation, directional sounds do not mean sounds produced by directional sources (sound propagation in space), but sounds associated with specific directions of movements. These directions are the four main directions of left, up, right, and down as well as the intermediate directions of down–left, up–left, up–right, and down–right. This is shown in Fig. 3. Each directional sound consists of three tones.

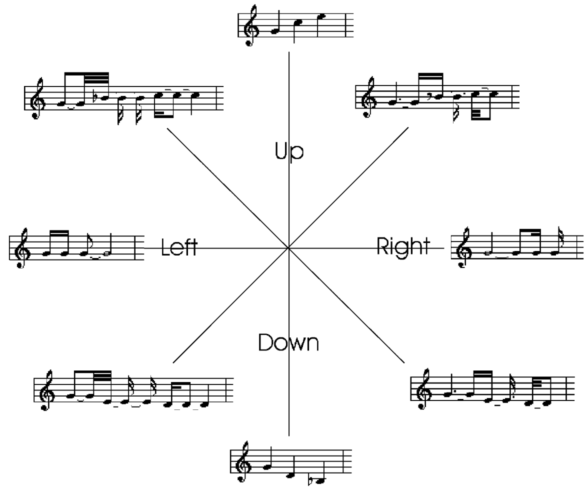

Fig. 3. Auditory symbols for eight directional sounds.

The musical basic elements rhythm and melody are used in the four main directions, independently of each other. The directional sound *up* is represented by a melody upwards, whereas a melody downwards denotes the sound with the meaning *down*. In both cases, each tone is of equal time duration. A rhythm of two short tones followed by one long one, all on the same pitch level, means *left*. Consequently, a rhythm with one long tone followed by two short ones on the same pitch level expresses the direction *right*. Thus, *up–down* has a direct pitch mapping, whereas *left–right* has a nondirect mapping that needs more learning. The musical elements melody and rhythm are combined in the intermediate directions with respective intermediate values of melody span and rhythm.

Each of the eight directions was presented in four variations, i.e., with changed sound color (timbre) or changed tempo. Music instruments and robot noises were used. The four sound presentations were realized by marimba (P1), harpsichord (P2), robot-movement sound (P3), and fast (double) tempo of marimba (P4).

The eight directional sounds are a subset of a more fine-grained resolution of 48 directions around the circuit of a compass card of 360°. This has recently been investigated in a student's project thesis. The experimental setup for the 48 directions is shown in Fig. 4. For the eight directional sounds, only every sixth of all these directions, starting with the zero position for the direction *up*, was included in the experimental set for the research reported here.

C. Sound Design for State Information

Additional sounds for robot states and situations were newly designed. These robot states and situations are *heavy load*, *waiting*, *near obstacle*, and *low battery*. The latter sound was recorded from the original robot's indication of the low battery status. This is a continuous, quite annoying high tone. The other three sounds were played on the MIDI keyboard. The author tried to convey the subjective impression of the meanings of the three sounds [46]. For example, the *heavy load* sound was played with three parallel tubas as one accentuated short time-interval tone followed by one

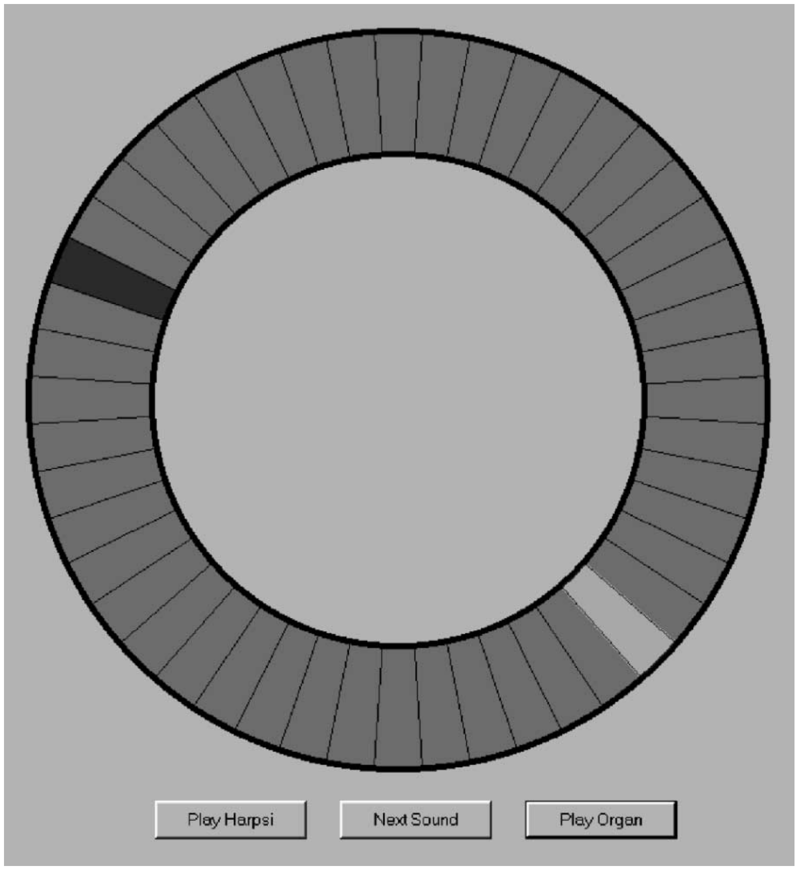

Fig. 4. Compass-card visualization for 48 directional sounds.

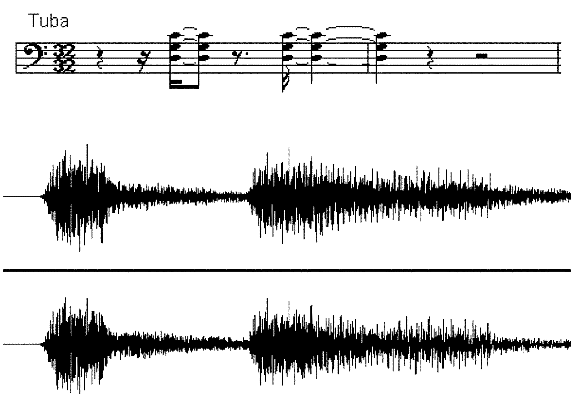

Fig. 5. Score and wave form of the *heavy load* sound.

tone of a longer time duration. Fig. 5 shows the score and the wave form of this *heavy load* sound.

The *waiting* sound and the *near obstacle* sound were played with alt sax and English horn, respectively; their scores and wave forms are presented in Figs. 6 and 7.

D. Design of Sound Tracks for Intent Communication

In the experiments of the supermarket scenario, the auditory perception of intended robot trajectories and of additional robot state sounds was investigated. A simulated supermarket scenario with robot sound tracks was designed by the author in such a way that a mobile service robot can make straight movements and turnings of 45° and 90° [45].

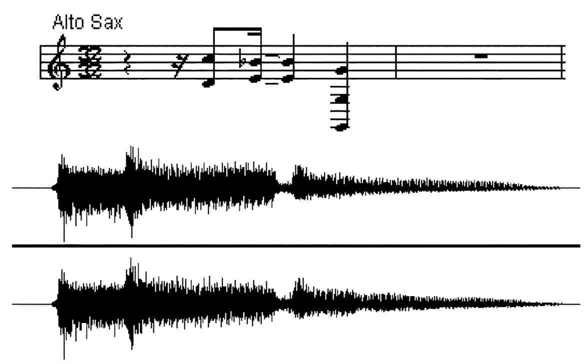

Fig. 6. Score and wave form of the *waiting* sound.

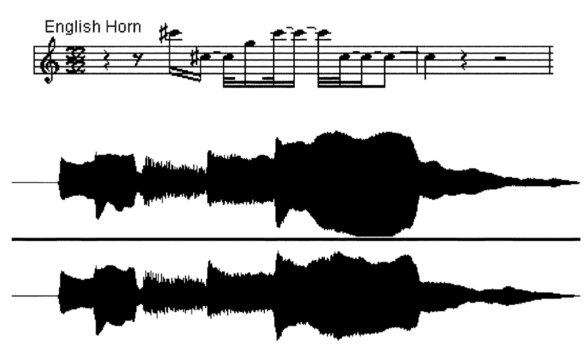

Fig. 7. Score and wave form of the *near obstacle* sound.

The sound tracks for the predictive display of the intended robot trajectories were composed of moving-straight segments and turnings. The predictive auditory display indicates

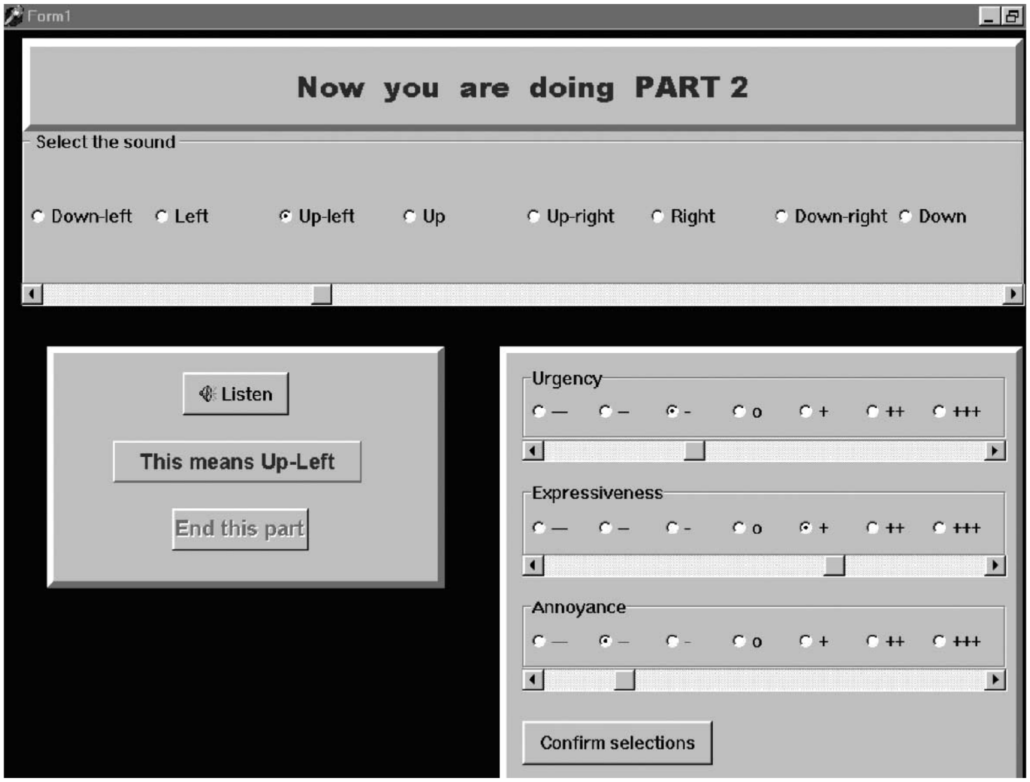

Fig. 8. Selection of eight directional sounds and their subjective evaluation.

a faster-than-real-time presentation. Robot state sounds, as described in the preceding section, are overlayed during some of the segments of the trajectories. The moving-straight segments are represented by the eight directional sounds, as described in Section V-B. The directions left, right, up, and down are particularly used but also a few of the intermediate directions (down–left, up–left, up–right, and down–right) are sometimes possible. Down means downwards on the computer screen and toward the human subject (shown on the lower middle of the screen in Fig. 12). Correspondingly, up means away from the subject.

The turnings (moving-curved sounds) were derived from recordings of the original robot turning sound by transposition. They are always heard with any directional change between any kind of two complete moving-straight sections. If a complete moving-straight section consists of a number of straight segments of the same direction, the appropriate directional sound is repeated correspondingly without any turning sound in between. A segment is defined as the straight connection between two neighboring active decision areas (which are explained in Section VI-B).

VI. Experiments With Auditory Displays for Mobile Systems

A. *Experiments and Results With Directional Sounds*

The intelligibility and the recallability of auditory symbols was investigated in the first experiments. Human subjects had to learn, understand, and recall the auditory symbols of the eight directional sounds with their related meanings, as described in Section V-B. They had to process altogether three test parts at the computer screen and heard the appropriate auditory symbols over loudspeakers. The experimental setup was programmed in Delphi.

In the first part of the experiments, all eight directions in each of the four variations (thus, altogether 32 auditory symbols), as explained in Section V-B, were presented, in random order. For each auditory symbol, its respective meaning was displayed immediately afterwards.

For refreshing the subjects' memory, they were able to select all eight directions again in the second part, each in the four variations, now in any order. After each auditory symbol, the subjects were asked for a subjective evaluation of its sound characteristics regarding urgency, expressiveness, and annoyance; see Fig. 8. For this purpose, they had to make a selection in each case, and had to mark three seven-point scales with respect to their subjective impression, as follows:

from - - - (very small) to + + + (very urgent);
from - - - (very small) to + + + (very expressive); and
from - - - (very small) to + + + (very annoying).

In the third part, the subjects should then show how well they can recognize all eight directions independently of the presented four variations. They received feedback after each of their decisions whether the respective auditory symbol had been recognized correctly.

The experiments with the directional sounds were performed in all their three parts altogether three times, i.e., with two repetitions. A group of eight nonmusicians and a control group of two professional musicians participated in all the experiments.

The results of the experiments with the directional sounds show that the nonmusicians were between 34% and 100%

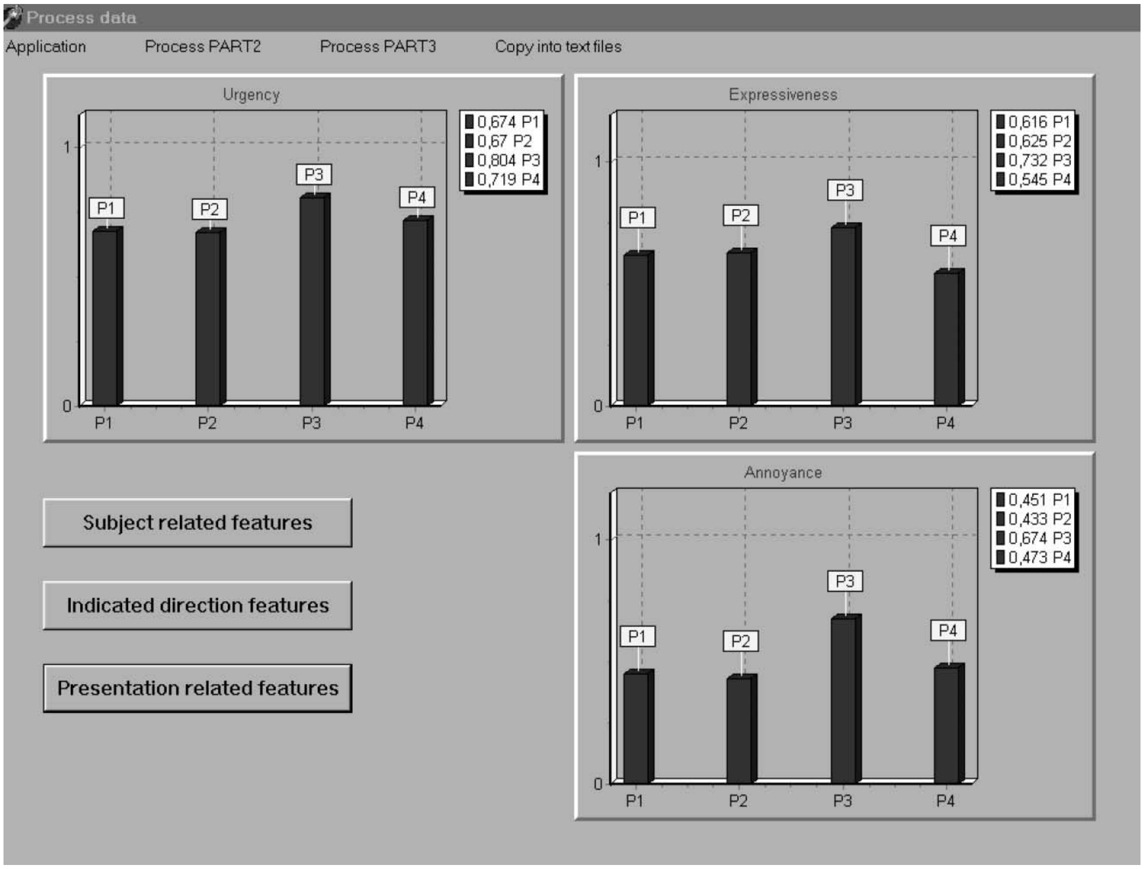

Fig. 9. Presentation-related sound characteristics for nonmusicians.

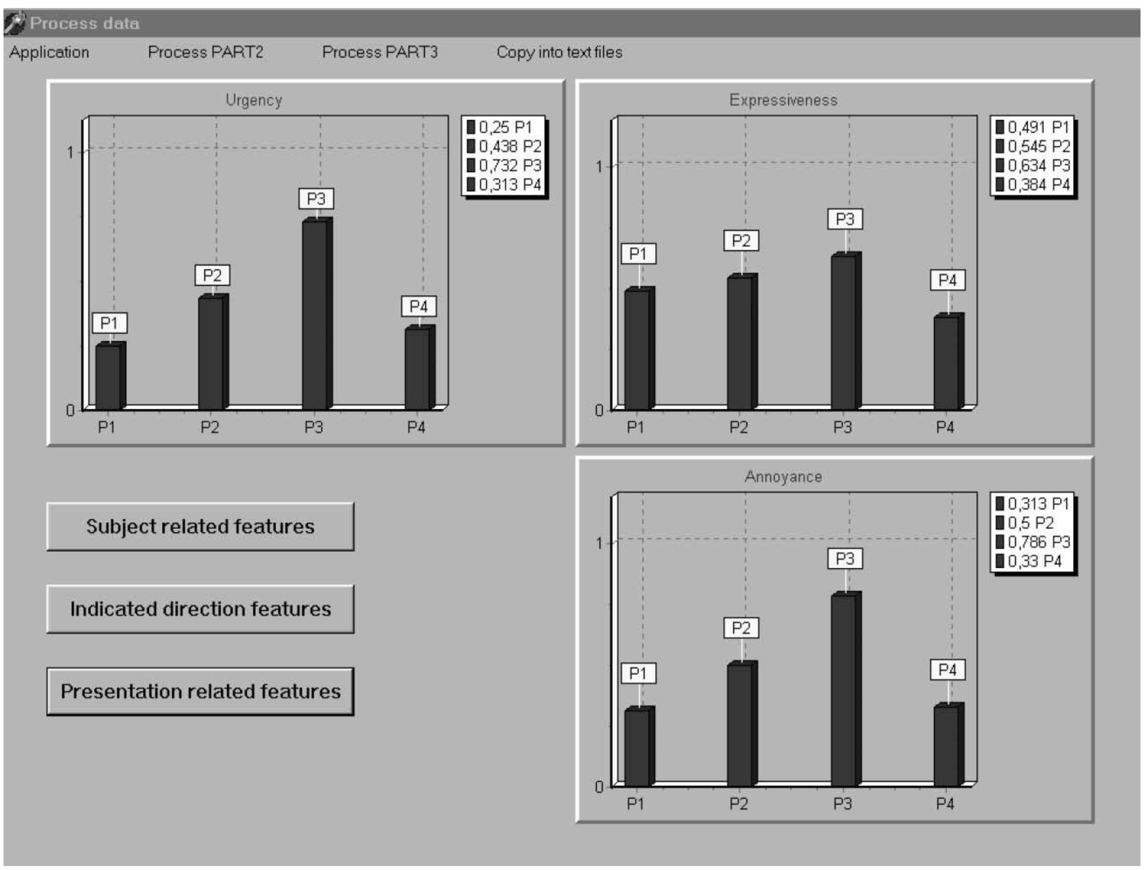

Fig. 10. Presentation-related sound characteristics for musicians.

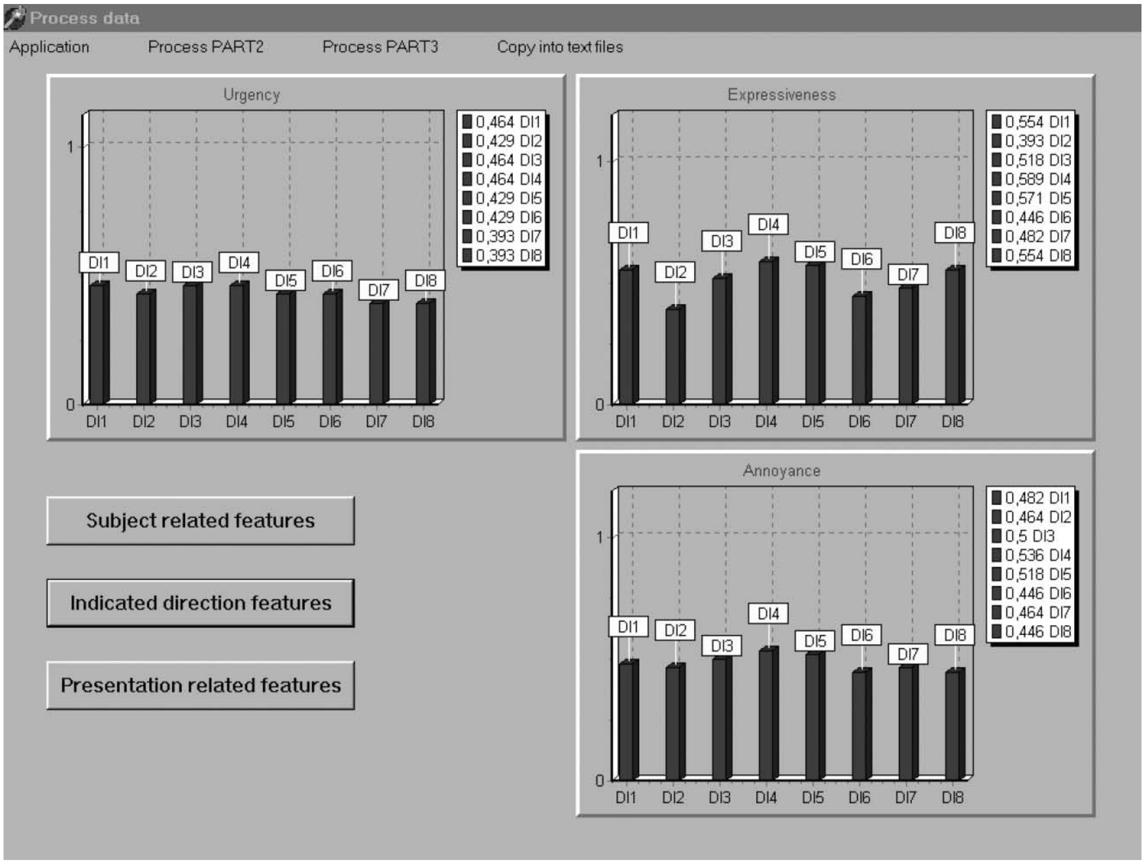

Fig. 11. Indicated-direction-related sound characteristics for musicians.

(one subject) correct, in the second repetition of part 3, whereas the two professional musicians were 97% and 100% correct. The subjects had been informed in the written instructions that they should be able to retain the auditory symbols in their memory because the second experiment with auditory displays in a robot-simulation environment would be based on these.

Subjective evaluations of the sound characteristics urgency, expressiveness, and annoyance were made by the nonmusicians and the musicians. These characteristics were processed with respect to subject-related, indicated-direction-related, and presentation-related features.

During the second and last repetition of the experiments, the group of nonmusicians was divided into two subgroups of four subjects each. One subgroup was formed of the four subjects who showed higher performance during the first experiments, whereas the other subgroup contained the four subjects with lower performance. Fig. 9 shows the presentation-related features for the higher-performance subgroup of the nonmusicians. The presentation of the robot-movement sounds P3 leads to clearly higher values of subjective evaluations for all three characteristics—urgency, expressiveness, and annoyance. All the presentations with the musical instruments are more or less on the same level. Only the faster presentation with the marimba P4 has been assessed slightly higher with respect to urgency and annoyance, whereas the mean value for expressiveness is slightly lower.

The same tendency with higher values for urgency, expressiveness, and annoyance for the robot-movement sounds can be observed for the musicians. Fig. 10, with the mean values of the two musicians, shows that this effect is much more pronounced for the characteristics of urgency and annoyance. These two characteristics seem to be highly correlated over all four presentations. For all three characteristics, the presentation with the harpsichord P2 leads to intermediate values between those for the marimba P1 and the robot-movement sounds P3.

The lowest evaluation of the expressiveness within the indicated-direction-related features has been achieved for the direction left DI2 (short–short–long) by the musicians; see Fig. 11. This effect cannot be found with the nonmusicians. The largest number of mistakes is made with the direction down–left by the nonmusicians. One musician commented in the final interview that all the sounds toward the left with the rhythm short–short–long have been more difficult to be recalled. This is probably due to a bad recognition of the duration of the long tone at the end of the earcon. Nevertheless, he did not make more mistakes with such sounds.

B. Experiments and Results With Sound Tracks

In the second set of experiments, intelligible auditory symbols and sound tracks, as described in Section V-D, were presented in a supermarket scenario with a simulated (Windows and Delphi) environment of a mobile service robot. It

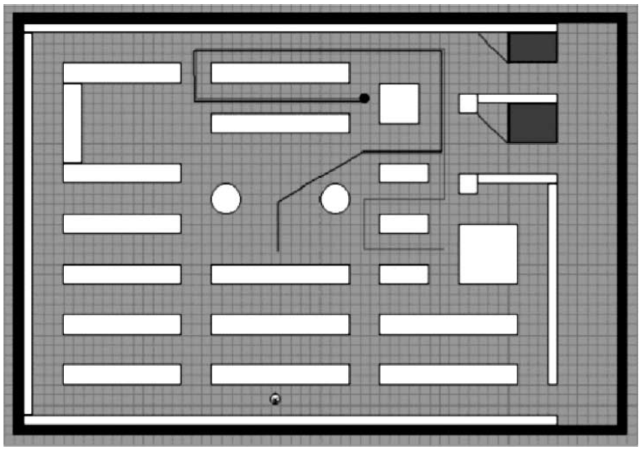

Fig. 12. Replay of intended and perceived robot trajectories in the supermarket scenario.

is assumed that the supermarket is open during seven days a week for 24 h. A mobile service robot for cleaning and for carrying goods will inform the human subject (the customer) with sound symbols of nonspeech auditory predictive displays about the trajectory of its intended movements and about the additional robot states and situations *heavy load*, *waiting*, *near obstacle*, and *low battery*. These additional sounds for the robot situations had to be learned by the human subjects in a training phase at the beginning of the second experiments. The subjects could listen to these sounds in any order as often as they wanted.

A floor plan of the supermarket is visualized on the computer screen. The human subject is shown in the lower middle and the robot in different starting positions, which depend on the investigated trajectory; see Fig. 12.

A matrix of decision areas was constructed. Any intersection between a horizontal and a vertical line together with the respective nearest surrounding of this crossing, in which alternative routes can be chosen (beyond returning the same way), is determined as an active decision area in the visual floor plan of the supermarket; see Fig. 12.

The robot sound tracks actually used in the experiments of the supermarket scenario are the overlays of the sound tracks of the intended robot trajectories and, during some of their segments, of the additional sounds for the robot states and situations *heavy load*, *waiting*, *near obstacle*, and *low battery*. The human subjects were asked to recognize and to understand the intended trajectory of the robot as well as the overlayed additional sounds of the robot situations, from listening to the robot sound track. They had to draw the auditorily perceived trajectory into the visual floor plan of the supermarket on the computer screen and had to mark the perceived additional sounds. The subjects were informed about the correctness of their auditory perception.

Altogether, the subjects performed four subexperiments, each with four different trajectories. The intended trajectory and the perceived trajectory as well as the intended and the perceived additional sounds were recorded. Also, the durations of the training phase and of the drawing of each perceived trajectory were measured. The intended and the perceived trajectories can be compared in the replay mode; see Fig. 12.

In the last two of these subexperiments, the sound tracks of the trajectories were composed of the same sound symbols of the intended trajectories and the overlayed additional sounds for the robot situations. However, the sounds of the real robot movements were now also overlayed. They were presented in real time, whereas the intended trajectories are auditory predictor displays and, thus, faster than real time. This makes the scenario even more difficult for the subjects but it is also more realistic. In a real-world human–robot environment, the real robot movements are also always heard.

The experimental results with the same eight nonmusicians and the two professional musicians showed large differences in their auditory perception. The same three of these ten subjects who performed best in the first experimental study made only very few errors (one musician actually made no errors) with the perception of the robot sound tracks of all trajectories as well as the additional sounds for the robot states and situations.

In summary, the experimental studies showed that the sound symbols for directions and robot states as well as the robot sound tracks are recallable and understandable, at least for more musical people. Positive training effects have been observed with all human subjects. The number of subjects was not large enough for a statistical analysis within this exploratory study. However, the consistency of the achieved individual results indicates that the investigated auditory displays seem to be usable for human–machine communication and interaction.

C. Time-Frequency Analyses of Directional and State Sounds

An attempt has been undertaken in this research to explain subjective evaluations of sound characteristics by means of image patterns in the time-frequency planes (spectrograms) with the wavelet analysis techniques [44]. The subjective evaluations of sound characteristics such as urgency, expressiveness, and annoyance have been considered. Experimental results for selected directional sounds and robot state sounds have been compared with the diagrams which have been computed with two wavelet techniques for time-frequency analyses [49].

Both techniques, the fast wavelet transform and the wavelet packet technique, have been used for finding correlates to the human subjective evaluations of the sound characteristics urgency, expressiveness, and annoyance. The main question is whether the diagrams of the time-frequency analyses can clearly express those differences in the subjective evaluations which have been found in the experimental investigation. The results show some surprising agreements between the experimental findings and the computational results. The comparison between time-frequency analyses and subjective evaluations of sounds seems worth being further pursued. However, it should be enhanced by the additional analysis of suitable high level characteristics from the fields of music cognition and emotion.

VII. Conclusion

The designs of directional sounds, robot state sounds, and robot sound tracks have been accomplished with basic musical elements and with recorded robot noise signals. The experimental studies of this research show that the suggested auditory symbols and sound tracks are feasible means of communication in human–machine interaction. It can be expected that this result will be provable beyond the investigated example application domain. Continuing research considers auditory displays also for other application domains such as car driving, aircraft piloting, and supervision of industrial plants, as well as sound design for medical and domestic appliances.

The need for auditory displays and, thereby, the demand for task- and user-oriented sound design, also with more appropriate style guides, will increase in many application domains in the near future. Such auditory displays can at least partially face the visual overload of the human. The so-called ironies of automation [50], possibly manifested by a loss of vigilance, a loss of situational awareness, or a loss of experience of the human users, may be mitigated by alerting sounds and auditory displays, thus, by music and other sounds with informative contents. Furthermore, multimedia displays can be designed with a much higher versatility on condition of the availability of improved auditory displays. The present-day possibilities and the research activities, which have continuously been growing for about ten years at universities and in industry, bring about good preconditions for innovative product solutions.

Designs of sound for vehicles, for industrial systems, and for appliances have to understand tasks and users, have to secure continual user participation, have to integrate in a systems engineering fashion, and, at the same time, have creatively to compose sound and to master sound engineering.

In addition to all above applications, one can finally imagine that the directional auditory displays suggested in this paper can also be advanced for applications in art and virtual reality scenarios, for example, as a hearing education tool in music or as a notation and instruction tool in dance choreography, e.g., for the musification of movements. This demonstrates again the increasing potential of cross fertilization between application domains in music, other arts, and engineering.

Acknowledgment

The author would like to thank Prof. P. Kopacek, Director, Institute for Handling Devices and Robotics,Vienna University of Technology, Vienna, Austria, who was the author's main host while some of this work was being done, for his kind hospitality. The author would also like to thank the reviewers of this paper for their valuable comments and suggestions.

References

[1] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: The MIT Press, 1992.

[2] G. Johannsen, *Mensch-Maschine-Systeme (Human–Machine Systems) (in German)*. Berlin, Germany: Springer-Verlag, 1993.

[3] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*. New York: Wiley, 2002.

[4] G. Johannsen, "Human supervision and control in engineering and music—foundations and transdisciplinary views," *J. New Music Res. (Special Issue on Human Supervision and Control in Engineering and Music)*, vol. 31, no. 3, pp. 179–190, 2002.

[5] ——, "Cooperative human–machine interfaces for plant-wide control and communication," in *Annual Reviews in Control*, J. J. Gertler, Ed. Oxford, U.K.: Pergamon/Elsevier, 1997, vol. 21, pp. 159–170.

[6] N. A. Stanton and J. Edworthy, Eds., *Human Factors in Auditory Warnings*. Aldershot, U.K.: Ashgate, 1999.

[7] G. Kramer, Ed., *Auditory Display—Sonification, Audification, and Auditory Interfaces*. Reading, MA: Addison-Wesley, 1994.

[8] G. Johannsen, "Toward a new quality of automation in complex man–machine systems," *Automatica*, vol. 28, pp. 355–373, 1992.

[9] J. L. Alty and G. Johannsen, "Knowledge-based dialogue for dynamic systems," *Automatica*, vol. 25, pp. 829–840, 1989.

[10] B.-B. Borys and G. Johannsen, "An experimental multi-media process control room," in *Proc. Annu. Conf. 1997 Human Factors and Ergonomics Soc. Eur. Chapter, Advances in Multimedia and Simulation*, pp. 276–289.

[11] L. Fejes, G. Johannsen, and G. Strätz, "A graphical editor and process visualization system for man–machine interfaces of dynamic systems," *Vis. Comput.*, vol. 10, no. 1, pp. 1–18, 1993.

[12] G. Johannsen, S. Ali, and R. van Paassen, "Intelligent human–machine systems," in *Methods and Applications of Intelligent Control*, S. G. Tzafestas, Ed. Dordrecht, The Netherlands: Kluwer, 1997, pp. 329–356.

[13] J. A. Ballas, "The interpretation of natural sound in the cockpit," in *Human Factors in Auditory Warnings*, N. A. Stanton and J. Edworthy, Eds. Aldershot, U.K.: Ashgate, 1999, pp. 91–112.

[14] P. Ulfvengren, L. Mårtensson, and G. Singer, "Auditory and visual warnings in aircraft," in *Proc. 8th IFAC/IFIP/IFORS/IEA Symp.: Analysis, Design, and Evaluation of Human–Machine Systems 2001*, G. Johannsen, Ed., pp. 53–57.

[15] J. Welch, "Auditory alarms in intensive care," in *Human Factors in Auditory Warnings*, N. A. Stanton and J. Edworthy, Eds. Aldershot, U.K.: Ashgate, 1999, pp. 319–342.

[16] D. Sonnenschein, *Sound Design—The Expressive Power of Music, Voice, and Sound Effects in Cinema*. Studio City, CA: Wiese, 2001.

[17] B.-B. Borys, "Evaluation of auditory displays supporting aircraft approach and landing," in *Proc. 8th IFAC Symp.: Automated Systems Based on Human Skill and Knowledge*, J. Stahre and L. Mårtensson, Eds., 2003, pp. 141–146.

[18] D. D. Woods, "Visual momentum: a concept to improve the cognitive coupling of person and computer," *Int. J. Man–Mach. Stud.*, vol. 21, pp. 229–244, 1984.

[19] K. L. Norman, L. J. Weldon, and B. Shneiderman, "Cognitive layouts of windows and multiple screens for user interfaces," *Int. J. Man–Mach. Stud.*, vol. 25, pp. 229–248, 1986.

[20] E. D. Mynatt, "Transforming graphical interfaces into auditory interfaces for blind users," *Human–Comput. Interact.*, vol. 12, pp. 7–45, 1997.

[21] J. L. Alty and D. I. Rigas, "Communicating graphical information to blind users using music—the role of context," in *Proc. Computer–Human Interaction Conf. (CHI 98)*, vol. 1, C. M. Karat, A. Lund, J. Coutaz, and J. Karat, Eds., pp. 574–581.

[22] J. T. Reason, *Human Error*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[23] J. Pressing, "Some perspectives on performed sound and music in virtual environments," *Presence*, vol. 6, no. 4, pp. 482–503, 1997.

[24] D. O'Shaughnessy, "Interacting with computers by voice: automatic speech recognition and synthesis," *Proc. IEEE (Special Issue on Human–Computer Multimodal Interface)*, vol. 91, pp. 1272–1305, Sept. 2003.

[25] W. W. Gaver, "The SonicFinder: an interface that uses auditory icons," *Human–Comput. Interact.*, vol. 4, pp. 67–94, 1989.

[26] ——, "Using and creating auditory icons," in *Auditory Display—Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Reading, MA: Addison-Wesley, 1994, pp. 417–446.

[27] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: their structure and common design principles," *Human–Comput. Interact.*, vol. 4, pp. 11–44, 1989.

[28] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "Experimentally derived guidelines for the creation of earcons," in *Adjunct Proc. HCI'95*, pp. 155–159.

[29] D. McGookin and S. A. Brewster, "An investigation into the identification of concurrently presented earcons," in *Proc. 2003 Int. Conf. Auditory Display (ICAD03)*, 2003, pp. 42–46.
[30] P. Vickers and J. L. Alty, "Using music to communicate computing information," *Interact. With Comput.*, vol. 14, no. 5, pp. 434–455, 2002.
[31] ——, "Musical program auralization: a structured approach to motif design," *Interact. With Comput.*, vol. 14, no. 5, pp. 456–484, 2002.
[32] M. Cohen and L. F. Ludwig, "Multidimensional audio window management," *Int. J. Man–Mach. Stud.*, vol. 34, no. 3, pp. 319–336, 1991.
[33] J. Herder and M. Cohen, "The helical keyboard: perspectives for spatial auditory displays and visual music," *J. New Music Res. (Special Issue on Human Supervision and Control in Engineering and Music)*, vol. 31, no. 3, pp. 269–281, 2002.
[34] R. N. Mumford and J. L. Alty, "3-D algorithm auralization using the SIMBA toolbox," in *Proc. 18th Eur. Annu. Conf. Human Decision Making and Manual Control*, J. L. Alty, Ed., 2000, pp. 214–225.
[35] G. Johannsen, J. Heuer, and M. Tiemann, "Participative user interface design for industrial processes," in *Proc. Silicon Valley Ergonomics Conf. (ErgoCon'97)*, 1997, pp. 21–28.
[36] D. A. Norman, "Cognitive artifacts," in *Designing Interaction: Psychology of the Human–Computer Interface*, J. M. Caroll, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[37] E. Hollnagel, Ed., *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum, 2003.
[38] J. Rasmussen, A. M. Pejtersen, and L. P. Goodstein, *Cognitive Systems Engineering*. New York: Wiley, 1994.
[39] D. Redmond-Pyle and A. Moore, *Graphical User Interface Design and Evaluation (GUIDE)*. London, U.K.: Prentice-Hall, 1995.
[40] G. Johannsen, "Cognitive systems analysis, design, and experimental investigation of auditory displays for human–machine interfaces," in *Proc. Int. Conf. Cognitive Systems Engineering in Process Control (CSEPC 2000)*, 2000, pp. 150–155.
[41] S. Ronkainen, "Earcons in motion—Defining language for an intelligent mobile device," in *Proc. 2001 Int. Conf. Auditory Display (ICAD01)*, 2001, pp. 126–131.
[42] G. Johannsen, "Sound-Design für Anmutung und Systeminformationen bei Fahrzeugen und Geräten," in *Tagungsband USEWARE 2002: Mensch-Maschine-Kommunikation/Design, Fachtagung der VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik*. Düsseldorf, Germany: VDI-Verlag, 2002, pp. 285–292.
[43] G. Kronreif, R. Probst, and P. Kopacek, "Modular service robots—state of the art and future trends," in *Proc. 8th Int. Conf. Advanced Robotics (ICAR'97)*, 1997, pp. 51–56.
[44] G. Johannsen, "Analysis of audio symbols based on musical and robot-movement sounds using time-frequency methods," in *Proc. Diderot Forum Mathematics and Music—Conf. Computational and Mathematical Methods in Music*, H. G. Feichtinger and M. Dörfler, Eds., 1999, pp. 215–226.
[45] ——, "Auditory displays in human–machine interfaces of mobile robots for nonspeech communication with humans," *J. Intell. Robot. Syst.*, vol. 32, pp. 161–169, 2001.
[46] ——, "Auditory display of directions and states for mobile systems," in *Proc. 8th Int. Conf. Auditory Display (ICAD 2002)*, 2002, pp. 98–103.
[47] A. Camurri, P. Coletta, M. Ricchetti, and G. Volpe, "Expressiveness and physicality in interaction," *J. New Music Res. (Special Issue on Music and Motion)*, vol. 29, no. 3, pp. 187–198, 2000.
[48] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "A detailed investigation into the effectiveness of earcons," in *Auditory Display—Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Reading, MA: Addison-Wesley, 1994, pp. 471–498.
[49] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.
[50] L. Bainbridge, "Ironies of automation," *Automatica (Special Issue on Control Frontiers in Knowledge-Based and Man–Machine Systems)*, vol. 19, pp. 775–779, 1983.

Gunnar Johannsen (Fellow, IEEE) was born in Hamburg, Germany. He received the Dipl.-Ing. degree in communication and information engineering and the Dr.-Ing. degree in flight guidance and manual control from the Technical University of Berlin, Berlin, Germany in 1967 and 1971, respectively. In addition, he studied music for three years within the sound engineering curriculum at the University School of Music, Berlin.

For several years, he studied conducting in Hamburg; Vienna, Austria; and Kassel, Germany. From 1971 to 1982, he was Division Head in the Research Institute for Human Engineering, Bonn, Germany. In 1980, he habilitated (Dr. Habil.) and became Private Docent in the teaching area of human–machine systems of aeronautics and astronautics at the Technical University of Aachen, Aachen, Germany. Since 1982, he has been Professor of systems engineering and human–machine systems with the Institute for Measurement and Automation of the University of Kassel, Kassel, Germany. He has also worked with the University of Illinois, Urbana–Champaign; the Kyoto Institute of Technology, Kyoto, Japan; Kyoto University, Kyoto; and the Technical University of Vienna, Vienna, also supported by the Institute of Electro-Acoustics, Experimental and Applied Music of the University of Music and Performing Arts, Vienna. With the International Federation of Automatic Control (IFAC), he founded and chaired the Working Group on Man–Machine Systems for nine years, and served as Chairman for four of its symposia. He created and was mainly responsible of the International Workshop on Human Supervision and Control in Engineering and Music, with an embedded orchestra concert in which he conducted Takemitsu's "November Steps." His current research interests are in human–machine systems, human-centered design, cognitive systems engineering, graphical, auditory, and multimedia user interfaces, audio and music technologies, gestural control, decision support systems, knowledge engineering, cooperative work, and interrelations between engineering and music.

Prof. Johannsen is a Member of the Association for Computing Machinery (ACM), the Human Factors and Ergonomics Society (HFES), the International Computer Music Association (ICMA), and several German professional societies: VDI/VDE-GMA Gesellschaft Mess- und Automatisierungstechnik (measurement and automation), VDE-ITG Informationstechnische Gesellschaft (information engineering), GI Gesellschaft für Informatik (computer science), GfA Gesellschaft für Arbeitswissenschaft (work science), and DGLR Deutsche Gesellschaft für Luft- und Raumfahrt (aeronautics and astronautics). He received a Japanese–German Research Award, granted by the Japan Society for the Promotion of Science (JSPS).